Título *Revista da Associação Portuguesa de Linguística*, n.º 13

Organização

Adelina Castelo

Alexandra Fiéis

Purificação Silvano

Comissão científica do número 13

Aldina Marques (Universidade do Minho), Alexander Cobbinah (Universidade de São Paulo), Álvaro Iriarte Sanromán (Universidade do Minho), Amália Mendes (Universidade de Lisboa), Ana Espírito Santo (Universidade de Lisboa), Ana Luísa Costa (Instituto Politécnico de Setúbal), Augusto Soares da Silva (Universidade Católica Portuguesa), Carla Almeida (Universidade Aberta), Catarina Castro (Instituto Politécnico de Leiria), Clara Amorim (Instituto Politécnico de Viana do Castelo), Clara Barros (Universidade do Porto), Cristina Flores (Universidade do Minho), Cristina Martins (Universidade de Coimbra), Dina Alves (Instituto Politécnico de Setúbal), Fernanda Pratas (Universidade de Lisboa), Filomena Gonçalves (Universidade de Évora), Idalete Dias (Universidade do Minho), Isabel Almeida Santos (Universidade de Coimbra), Isabel Margarida Duarte (Universidade do Porto), Isabel Pereira (Universidade de Coimbra), Joana Batalha (Universidade de Lisboa), João Costa (Universidade NOVA de Lisboa), João Veloso (Universidade do Porto), Jorge Baptista (Universidade do Algarve), José Brandão de Carvalho (Universidade do Minho), Letícia Almeida (Universidade de Lisboa), Luís Filipe Barbeiro (Instituto Politécnico de Leiria), Maria João Freitas (Universidade de Lisboa), Maria Lobo (Universidade NOVA de Lisboa), Matilde Gonçalves (Universidade NOVA de Lisboa), Nélia Alexandre (Universidade de Lisboa), Paulo Nunes da Silva (Universidade Aberta), Rita Gonçalves (Universidade de Lisboa), Sónia Duarte (Centro de Linguística da Universidade do Porto), Susana Correia (Universidade NOVA de Lisboa).

Revisão Ana Cristina Silva

Data ISSN outubro de 2025
2183-9077

Nota prévia

Reunimos no número 13 da Revista da Associação Portuguesa de Linguística uma seleção de 17 artigos, originalmente apresentados durante o 40.º Encontro Nacional da Associação Portuguesa de Linguística, que teve lugar nos dias 23, 24 e 25 de outubro de 2024, na Universidade dos Açores. Foi um encontro marcante, no qual tivemos o prazer de celebrar os 40 anos da APL, e, pela primeira vez, realizar este evento fora do território continental, contando com uma expressiva participação de sócios e não sócios.

Por ter sido um encontro especial, no 40.º Encontro Nacional da Associação Portuguesa de Linguística, realizaram-se duas conferências plenárias, proferidas por duas ex-presidentes da APL: *Revisitando o acento de palavra em português*, apresentada por Isabel Pereira, da Universidade de Coimbra, e *APL 40 anos: Dar voz à linguística. Criar comunidade*, a cargo de Inês Duarte, da Universidade de Lisboa.

Para além das conferências plenárias, houve ainda 73 comunicações orais e 9 pósteres.

A diversidade de temas e de abordagens teórico-metodológicas presente nos artigos deste volume reflete, de forma particularmente clara, a vitalidade e a riqueza da investigação em Linguística que se desenvolve sobre a língua portuguesa. Entre as áreas contempladas encontram-se, a título de exemplo, a Sociolinguística, a Aquisição e o Ensino de Língua Materna e não Materna, a Fonologia, a Linguística Computacional, a Linguística Educacional, a Pragmática e a Análise do Discurso. Todos os artigos foram submetidos a avaliação científica por pares, tendo cada um sido revisto por, pelo menos, dois especialistas da respetiva área.

Queremos manifestar o nosso mais sincero reconhecimento pela colaboração fundamental dos revisores dos artigos agora publicados, assim como pelo contributo decisivo da comissão científica do Encontro, responsável pela avaliação dos resumos recebidos. Agradecemos igualmente a todos os participantes, conferencistas convidadas, oradores e público, cuja presença e empenho foram determinantes para três dias de partilha linguística, e debates intensos e enriquecedores. Por fim, deixamos um agradecimento muito especial às colegas da Comissão de Organização Local do 40.º Encontro Nacional da APL, Ana Teresa Alves e Helena Montenegro, pelo seu trabalho incansável e dedicação exemplar.

As editoras
Adelina Castelo, Alexandra Fiéis e Purificação Silvano

Alternâncias linguísticas na comunidade portuguesa residente em Andorra: um estudo sociolinguístico

Language alternations in the Portuguese community living in Andorra: a sociolinguistic study

Joana Aguiar[1], *Carolina Bastida Serra*[2]

[1] Centro de Investigação Transdisciplinar em Educação e Desenvolvimento (CITeD) – Instituto Politécnico de Bragança
[2] Grup de Recerca en Llengües –Universitat d'Andorra

Abstract

According to the Department of Statistics of Andorra (2025), there are 8510 Portuguese people living in Andorra, which represents 9.7% of the population. The objective of this paper is to characterise the language alternations present in the oral speech of informants from the Portuguese community in Andorra. More specifically, based on Poplack's (2020) proposal, our aim is to verify how and in what contexts language alternations (Portuguese-Catalan/Portuguese-Spanish) are present in the oral discourse of speakers who have Portuguese as their native language and as a heritage language. In order to study the influence of Catalan and Spanish on the oral discourse of Portuguese residents in Andorra, we interviewed 18 native speakers of Portuguese living in Andorra for more than four years and 9 Portuguese heritage language speakers. The results show that the social variables *schooling* and *years of residence in Andorra* are determining factors in explaining the frequency of occurrence of linguistic alternations and their characterisation. The role of other variables, such as language use in work contexts, in the patterns of alternation displayed by speakers is also discussed.

***Keywords*:** Portuguese community in Andorra, grammatical convergence, code-switching, lexical borrowing, language attitudes.

Resumo

De acordo com o Departament d'Estadística d'Andorra (2025), residem, em Andorra, 8510 portugueses, o que constitui 9,7% da população. O objetivo deste trabalho é caracterizar as alternâncias linguísticas presentes no discurso de informantes da comunidade portuguesa em Andorra. Mais especificamente, a partir da proposta de Poplack (2020), pretende-se verificar de que forma, e em que contextos, as alternâncias linguísticas (português-catalão/português-espanhol) estão presentes no discurso oral dos falantes que têm o português como língua materna (PLM) e como língua de herança (PLH). De forma a estudar a influência do catalão e do espanhol no discurso oral dos portugueses residentes em Andorra, foram entrevistados 18 falantes de português língua materna, residentes há mais de quatro anos em Andorra, e 9 falantes de português língua de herança. Os resultados mostram que, no caso dos falantes de PLM, a variável *escolaridade* e o tempo de residência em Andorra são fatores determinantes para explicar a frequência de ocorrência das alternâncias linguísticas e a sua caracterização. É ainda discutido o papel de outras variáveis, como o uso das línguas em contexto de trabalho nos padrões de alternância exibidos pelos falantes.

Palavras-chave: comunidade portuguesa em Andorra, convergência gramatical, code-switching, empréstimo lexical, atitudes linguísticas.

1. Introdução

Este trabalho apresenta o resultado de um projeto de investigação, realizado no âmbito da Cátedra Camões da Universitat d'Andorra (AD03 - UdA-2023/2024), sobre a língua portuguesa em Andorra e as atitudes linguísticas da comunidade portuguesa. A recolha de amostras de fala foi realizada através de entrevistas informais semiestruturadas a falantes de português língua materna (PLM) e de português língua de herança (PLH). Os objetivos gerais deste projeto são: i) contribuir para o estudo da diversidade linguística em Andorra, com enfoque na comunidade portuguesa; ii) identificar, caracterizar e analisar as instâncias de alternância linguística; iii) identificar eventuais processos de variação e mudança; iv) analisar as atitudes linguísticas em relação ao português e ao catalão. Neste artigo, debruçar-nos-emos sobre as instâncias de alternância linguística no discurso oral da comunidade portuguesa e afloraremos as atitudes linguísticas em relação ao português e ao catalão.

Este artigo está organizado da seguinte forma: na secção 2, é feita uma descrição da evolução sociodemográfica da comunidade portuguesa em Andorra; na secção 3, é feita uma breve descrição das línguas em contacto e dos seus usos sociais. É, também, feita referência aos sistemas educativos vigentes em Andorra e às implicações que estes têm na manutenção do multilinguismo. A secção 4 é dedicada à caracterização das alternâncias linguísticas segundo a proposta de Poplack (2020) e na secção 5 é explanada a metodologia adotada. As secções 6 e 7 são dedicadas à apresentação e discussão dos resultados, respetivamente.

2. A comunidade portuguesa em Andorra

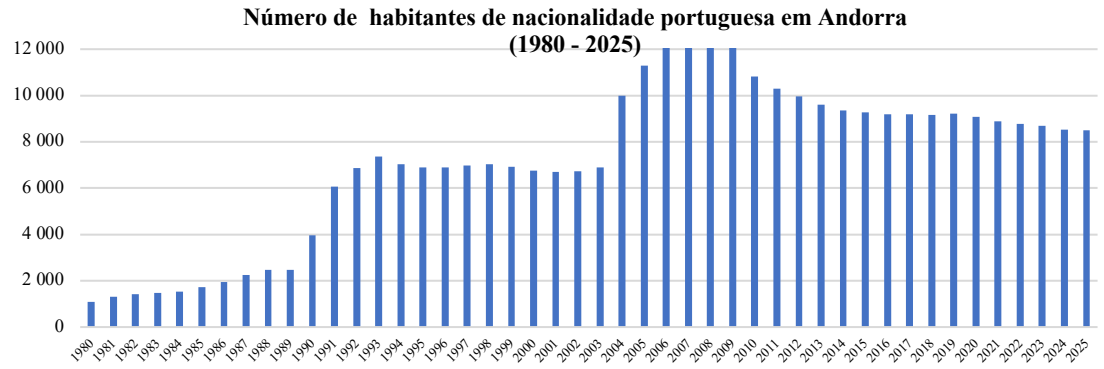
De acordo com informação do Departament d'Estadística do Govern d'Andorra, residem, em Andorra, 8510 portugueses, o que constitui 9,7% da população[1]. A maioria da população portuguesa residente em território andorrano concentra-se nas paróquias de Encamp, Andorra La Vella e Sant Julià de Lòria (Departament d'Estadística d'Andorra, 2025).

Apesar de o estabelecimento de portugueses no Principado ter tido início em XX (Carvalho, 2007), só a partir da década de 80 do século XX é que a presença da comunidade portuguesa começou a ser expressiva. Nesse sentido, de acordo com a tabela relativa à evolução do número de portugueses em Andorra (Figura 1), podemos identificar três principais movimentos migratórios: (i) o início dos anos 80, (ii) o início dos anos 90 e (iii) o início do século XXI (entre 2003 e 2006).

[1] Dados de fevereiro 2025 (Departament d'Estadística d'Andorra, 2025).

Figura 1

Comunidade Portuguesa em Andorra em Número de Habitantes de 1980 a 2025



Nota. Fontes: Carvalho (2007); Departament d'Estadística d'Andorra (2025); Matias (2008); Observatório da Emigração (2010). Elaboração própria.

De uma forma geral, a comunidade portuguesa é relativamente homogénea no que diz respeito à região de origem. A maioria é proveniente do Alto Minho e Trás-os-Montes (Matias, 2008). A escolha de Andorra como país de acolhimento resulta de um processo de "migração em cadeia" (Matias, 2008), motivado por familiares ou amigos já estabelecidos em Andorra.

Na década de 80, os imigrantes portugueses em Andorra eram provenientes de zonas rurais (Carvalho, 2007) e trabalhavam principalmente em trabalhos pouco qualificados, nomeadamente na construção e em serviços domésticos ou hotelaria. A comunidade portuguesa a residir em Andorra nessa época é descrita por Comas e Pujadas (1999) como a menos qualificada, razão pela qual ocupava a posição mais baixa na escala social: "As the newest and the least qualified group they occupy the lowest position on the social scale: the men mainly work as day labourers in agriculture and construction; the women are mainly shop assistants or cleaners" (Comas & Pujadas, 1999, p. 260).

A partir do início do século XXI, a caracterização sociodemográfica dos portugueses que emigram para Andorra sofre alterações: são mais jovens e provenientes de centros urbanos (Matias, 2008). Em termos de escolaridade, terminaram o ensino secundário ou um curso superior em Portugal, facto que facilitou a sua integração na sociedade andorrana (Carvalho, 2007).

3. Línguas em contacto

Andorra é um território pequeno, com uma população reduzida, onde, devido aos sucessivos fluxos migratórios e sazonais, confluem comunidades e línguas distintas. A partir da segunda metade do século XX, assiste-se a um crescimento de migrantes, provenientes principalmente de Espanha e de Portugal. Em resultado das sucessivas vagas migratórias, o perfil linguístico da população adulta residente em Andorra é diverso (Govern d'Andorra, 2023): 44,1% da população adulta é falante nativo de catalão, 40,3% de espanhol e 13,5% de português.

A língua oficial no Principado de Andorra é o catalão. Apesar de o espanhol ser comummente usado, o catalão é a língua das instituições e dos serviços públicos. Não obstante os esforços para a promoção da aprendizagem e do uso do catalão, nomeadamente a disponibilização de cursos de língua catalã gratuitos, a publicação da *Llei d'ordenació de l'ús de la llengua oficial* (Govern d'Andorra, 2000) e a recente aprovação da

Llei 6/2024, *de la llengua própria i oficial*[2] (Govern d'Andorra, 2024), há fatores que comprometem o uso da língua oficial. Por um lado, podemos identificar dois fatores demográficos relevantes, como explanaremos de seguida. Por outro lado, ao contrário do que observamos noutros países, como a Alemanha, por exemplo, o uso de outras línguas, como o espanhol ou o português, em detrimento da língua oficial, não é discriminatório (Ballarín, 2006). A proximidade geográfica de Andorra com a Espanha e a França permitiu a coexistência do catalão com o francês e, principalmente, do catalão com o espanhol.

Relativamente aos fatores demográficos, 54,9% dos residentes em Andorra não têm nacionalidade andorrana. Adicionalmente, verifica-se uma grande flutuação de residentes devido ao trabalho sazonal (Bastida & Nicolau, 2012) na denominada "temporada da neve". Esta necessidade de mão de obra é maioritariamente preenchida por falantes nativos de espanhol, provenientes da América do Sul, principalmente da Argentina (Matias, 2008). Concorre, ainda, para o uso frequente do espanhol o facto de 42,6% dos habitantes de Andorra terem nacionalidade espanhola ou serem provenientes de países onde se fala espanhol (Departament d'Estadística d'Andorra, 2025). Para além das comunidades provenientes de Espanha e Portugal, residem em Andorra 3.920 franceses (4,5%) e 14.801 cidadãos de outras nacionalidades (16,9%), provenientes maioritariamente de países da América do Sul. O contacto entre línguas, entendido como a situação em que os membros de uma comunidade de fala acedem regularmente a duas ou mais línguas no decorrer das suas interações linguísticas (Poplack, 2020, p. 47), é evidente. O país caracteriza-se, assim, pelo seu multilinguismo individual e social (Hawkey & Horner, 2022). Por esta razão, no entender de Comas e Pujadas (1999), o uso do espanhol em Andorra não tem, para o falante, um valor identitário ou simbólico, mas um caráter instrumental: "[Spanish] has an instrumental rather than a symbolic or identifying value" (Comas & Pujadas, 1999, p. 260). Neste sentido, é possível afirmar que o espanhol se assemelha a uma língua franca (Hawkey & Horner, 2022), utilizada não só em trocas comerciais por agentes de turismo e turistas, mas também pela maioria dos residentes não andorranos. O espanhol é, também, comummente utilizado por portugueses, principalmente por aqueles que (ainda) não aprenderam catalão. A este propósito, convém salientar que a maioria dos portugueses residentes em Andorra provém de regiões raianas em Portugal (Matias, 2008). Como tal, o contacto com o espanhol ou o galego era frequente antes de residirem em Andorra. Muitos não aprenderam formalmente o espanhol e falam uma mistura de português com espanhol, um linguajar informalmente denominado de *portunhol*, como refere o informante 14M3SE[3] (1):

> (1) Mas, claro, vir aqui e começar a trabalhar e tudo isso foi uma aventura. E eu, talvez dos poucos que depois nas entrevistas vai ver, que primeiro adaptei-me à Andorra, cursos de catalão, participar em associações andorranas, ou seja, integração total, praticamente total. **O que eu acho é que muita gente acomodou-se ao castelhano, e eu vou dizer um exagero, é o portunhol, muitas pessoas com quem falaste encontraste isso, gente que mistura o castelhano com o português e alguma palavra em catalão,** se é que se lembra, mas muito pouco. [14M3SE]

O inquérito ao conhecimento e usos linguísticos de 2023 (Govern d'Andorra, 2023) revela que o catalão é atualmente a língua materna da maioria da população residente em Andorra. O mesmo inquérito revela que 63,7% dos inquiridos referem que o catalão é a língua que habitualmente usam para comunicar com outras pessoas. Nos setores da construção e do comércio, há uma maior presença de espanhóis e portugueses, o que permite que parte da população empregada nestes segmentos profissionais comunique principalmente na sua língua materna (português ou espanhol) (Margarit et al., 2010). Adicionalmente, o português é a língua de comunicação em contextos familiares em que pelo menos um dos progenitores é português. Contribuem, ainda, para o uso do português a existência de associações culturais e recreativas, com uma presença ativa na sociedade

[2] A *Llei 6/2024* prevê a exigência de um conhecimento mínimo de catalão para pedidos e renovações da autorização de residência, assim como a garantia de um serviço público em catalão.
[3] Para mais informações relativamente ao código alfanumérico adotado para codificação dos informantes, veja-se a secção 5.2.

andorrana, como o Grupo de Folclore Casa de Portugal e a Associação de Portugueses do Alto Minho. São frequentes as exposições e os eventos culturais e etnográficos alusivos à cultura e tradições portuguesas (o dia de S. Martinho, o cantar das janeiras, o mercado tradicional, entre outros). Estas manifestações culturais promovem o contacto entre falantes e a perpetuação da cultura portuguesa e da identificação coletiva da comunidade portuguesa em Andorra.

Relativamente ao conhecimento explícito das línguas, o inquérito ao conhecimento e usos linguísticos de 2023 (Govern d'Andorra, 2023) indica que, apesar de o conhecimento explícito do catalão ter aumentado, a maioria dos entrevistados considera ter mais competências linguísticas a espanhol do que a catalão. Os jovens residentes em Andorra, entre os 16 e os 20 anos, são predominantemente bilingues (espanhol e catalão), havendo, no entanto, casos de monolinguismo e plurilinguismo (Margarit et al., 2010). Os autores referem, ainda, que, apesar de, em contexto familiar, predominar o monolinguismo, é comum o uso de duas línguas para comunicar na escola ou com os amigos. O inglês é a língua que os adolescentes dominam menos, o que se deve à sua introdução tardia no currículo escolar (Margarit et al., 2010).

O multilinguismo de Andorra está patente no sistema educativo público andorrano, que compreende três tipologias de ensino, estreitamente associadas à língua adotada para veicular o conhecimento: o andorrano, o francês e o espanhol[4]. Apesar de existirem acordos específicos com Espanha e com França, todos os sistemas são regidos pela *Llei qualificada d'educació* (de 3 setembro de 1993) e pela *Llei d'ordenament del sistema educatiu andorrà* (de 09 de junho de 1994). A escolha do sistema de ensino condiciona, por isso, não só a exposição formal ao catalão, mas também os contextos de uso da língua (Bastida & Nicolau, 2012). No sistema espanhol, o ensino é feito em espanhol. No sistema andorrano, a lecionação é feita em catalão. As escolas do sistema francês são financiadas pelo governo francês, seguem o *curriculum* elaborado pelo Ministério da Educação francês e o ensino é feito em francês. No sistema francês, a disciplina de português está disponível como optativa, desde o ensino primário até ao *lycée,* e é lecionada após as restantes atividades letivas ou durante o intervalo letivo para almoço.

A opção por um sistema em detrimento dos demais prende-se principalmente com fatores familiares e geográficos (Matias, 2008). No caso dos filhos de portugueses que frequentam ou frequentaram a escola andorrana, Matias (2008) considera que fatores como uma maior identificação cultural e o desejo de uma maior integração na sociedade andorrana poderão ser relevantes para explicar esta escolha. O trabalho de Hawkey (2024) aponta, ainda, como fator determinante para a escolha do sistema educativo a política linguística familiar. Segundo o autor, as famílias que privilegiam a aprendizagem e uso do português matriculam os filhos no sistema francês.

4. A alternância linguística

A alternância de códigos linguísticos diz respeito ao uso de duas ou mais línguas, pelo mesmo falante, no mesmo ato comunicativo. A alternância linguística corresponde a uma estratégia comunicativa, com motivações diversas e cuja extensão é variável, do ponto de vista sintático. Concorrem para as alternâncias entre duas línguas, entre outros, os seguintes aspetos (Poplack, 1980):

- a idade e o nível de escolaridade;
- a forma e o *timing* de aquisição das línguas envolvidas: idade de aquisição, período de exposição à(s) língua(s), contacto com a(s) comunidade(s) de fala e tipologia das redes sociais às quais os intervenientes pertencem e com as quais se relacionam;
- a identidade étnica, a profissão e o local de trabalho.

Do ponto de vista linguístico, as alternâncias linguísticas permitem que o locutor molde a sua interação de acordo com o contexto, o tópico e o interlocutor e possibilitam o preenchimento de possíveis lacunas

[4] Em Andorra, existem, também, escolas privadas (por exemplo, o *British College of Andorra* e a *Agora International School*) em que o ensino é feito em inglês.

gramaticais ou de vocabulário (Flores, 2004). É também comum o uso de outra língua "quando esta dispõe de componentes mais económicas para expressar um dado conteúdo verbal" (Flores, 2004). O uso de outra língua permite, ainda, enfatizar ou clarificar o que é dito (Androutsopoulos, 2005) e substituir expressões vernáculas e calão.

Neste trabalho, é adotada a proposta de Poplack (2020) para análise das alternâncias linguísticas. Baseada nos pressupostos metodológicos da linguística variacionista (Labov, 1966, 1972), a autora propõe uma metodologia de análise das *manifestações linguísticas decorrentes do contacto linguístico* (Poplack, 2020, p. 47). Assim, dependendo da sua extensão e das suas propriedades morfossintáticas exibidas, é possível distinguir as seguintes estratégias: (i) o empréstimo lexical (*lexical borrowing*), (ii) o code-switching (*multi-word intrasentential code-switching*) e (iii) a convergência gramatical (*grammatical convergence*), que explanaremos de seguida.

O empréstimo lexical consiste na incorporação de palavras da língua doadora (L_D), o espanhol ou o catalão, no discurso da língua recipiente (L_R), o português. Prevalece a morfologia e a sintaxe da L_R.

Considerem-se os exemplos (2) e (3):

(2) Para mim é uma das línguas mais feias do mundo. Prefiro o **chino**. [24F3SU]

(3) [16F2SE] Nesta ponte, estes são os túneis novos que fizeram que ligam, digamos, Encamp a Massana. Quando estavam a fazer, aí **sim que** houve um grande acidente, morreram vários portugueses.
[Entrevistadora] Mas caíram pedras, foi?
[16F2SE] Não, derramou-se. Estavam a pôr **hormigón** e tudo isso, algo não estava bem feito [16F2SE]

Nos exemplos (2) e (3), os termos do espanhol *chino*[5] e *hormigón*[6] são inseridos no interior de uma frase em português, antecedida de outra, também produzida em português. Não há alteração da morfologia nem da sintaxe do português. Em "aí **sim que** houve um grande acidente" (exemplo 3) podemos identificar o uso da locução adverbial do catalão e do espanhol *sí que*, usada coloquialmente para designar *certamente, de verdade*[7].

O termo *code-switching* tem sido utilizado por diferentes autores para designar, *lato sensu*, a justaposição, dentro da mesma interação discursiva, de elementos pertencentes a dois (sub)sistemas gramaticais distintos (Gumperz, 1982). De acordo com Poplack (2020), o termo *code-switching* designa a alternância de línguas ou variedades em contacto dentro da mesma frase. Normalmente, a extensão é superior à palavra (no original, *multi-word intrasentential code-switching*). No *code-switching*, nenhuma das línguas é tida como a recipiente, uma vez que se verifica a manutenção da morfologia e da sintaxe das línguas envolvidas, como ilustram os exemplos (4) e (5).

(4) Claro que **estás bien, has descansado un rato.** [6F2SE]

(5) **Y porque hay un chico** que está de manhã e ele faz o turno de manhã, mas precisavam alguém para fazer o turno de tarde. [28F1SE]

Note-se que, ainda que alternâncias de itens lexicais isolados sejam tendencialmente etiquetadas como *empréstimo lexical*, a diferença substancial entre *empréstimo lexical* e *code-switching* não é a extensão da alternância, mas a identidade da gramática subjacente a estas duas estratégias (Poplack, 2020).

O termo *convergência gramatical* designa as alternâncias linguísticas em que se observa a alteração da gramática da L_R por influência da L_D. Nesta tipologia, incluem-se propriedades ou processos que diferem, quer

[5] Em português, *chinês*.
[6] Em português, *betão armado*.
[7] Real Academia Española (s.d.).

pela presença quer pela ausência, dos observados na variedade não exposta ao contacto e que emergiram em resultado do contacto linguístico com a L_D (Poplack, 2020, p. 47). Observem-se os seguintes exemplos:

> (6) Fui para Barcelona, comecei a estudar tradução e interpretação, com base no francês e no russo, e depois de um tempo descobri que não era aquilo que eu gostava e que queria fazer, com o qual depois mudei **a** Humanidades. [21M2SU]
>
> (7) Os apartamentos estão caríssimos. (...) E depois, **te fica** muito pouco para comer, ou para te vestires, ou para.... [24F3SU]

No exemplo (6), o sintagma preposicional *a Humanidades* é introduzido pela preposição *a.* Em português, o verbo *mudar,* nos contextos em que designa alteração da área científica ou de curso superior, seleciona a preposição *para*, ao passo que, em espanhol e catalão, a preposição selecionada pelos verbos *cambiar* e *canviar*, respetivamente, é *a: cambiar a Humanidades/ canviar a Humanitats*[8].

No exemplo (7), verificamos a anteposição do clítico (**te** fica). Em português europeu, na ausência de atratores de próclise ou estruturas sintáticas que permitam ou exijam a próclise, o clítico ocorre em ênclise, ou seja, a estrutura seria: *E depois, fica-te muito pouco para comer*. Ao invés do que ocorre em português europeu, em espanhol e catalão, é privilegiada a próclise. Estes dois exemplos ilustram, assim, a influência do espanhol e do catalão (as línguas *doadoras*) no português (a língua *recipiente*).

5. Metodologia

Nesta secção, é descrita a metodologia adotada para a recolha e tratamento de dados linguísticos. É, ainda, apresentado o perfil sociolinguístico dos informantes entrevistados.

5.1 A entrevista

De forma a verificarmos a influência do catalão e do espanhol no discurso oral dos portugueses residentes em Andorra, foram recolhidas amostras de fala, em formato de entrevista informal semiestruturada. As entrevistas foram conduzidas por uma falante nativa de português europeu. Apesar de ter sido desenhado um protocolo de entrevista semiestruturada, de forma a uniformizar, sempre que possível, os temas em torno dos quais se desenvolvia a entrevista, pretendeu-se que o discurso fosse o mais espontâneo e informal possível, tendo o curso da interação sido adaptado ao entrevistado. Procurou-se, assim, desviar a atenção do informante para outros aspetos que não o processo comunicativo, de forma a facilitar a produção espontânea de formas vernáculas utilizadas no dia a dia (Labov, 2001; Tagliamonte, 2012). A análise da variação linguística a partir do modelo variacionista pode ajudar a esclarecer alguns aspetos relacionados não só com o funcionamento da linguagem, mas também com a forma como a linguagem está representada mentalmente e socialmente (Henry, 2002). Os temas principais foram a história de vida do entrevistado e as vivências em Andorra, assim como o uso do português em contextos familiares, sociais e laborais. As entrevistas foram realizadas em locais escolhidos pelos entrevistados, minimizando o impacto do momento da entrevista na vida quotidiana do informante. Frequentemente, as entrevistas decorreram em esplanadas, bancos de jardim, convívios associativos ou em deslocações de ou para o local de trabalho. Todos os informantes assinaram uma declaração de consentimento autorizando o uso da entrevista para fins científicos. No caso dos informantes menores, a declaração de consentimento foi assinada pelo encarregado de educação ou tutor legal.

Para a seleção de informantes e posterior entrevista foi determinante o estabelecimento de uma rede de contactos, quer institucionais (o Consulado Honorário de Portugal em Andorra, o Instituto Camões, a

[8] De referir que também os verbos *pasar* (espanhol) e *passar* (catalão) selecionam a preposição *a* quando designam mudança de área científica ou percurso académico: *pasarse a Humanidades* (espanhol) / *passar-se a Humanitats* (catalão).

Universitat d'Andorra) quer informais (a Associação de Portugueses do Alto Minho e o Grupo de Folclore Casa de Portugal). Os amigos de infância e familiares de antigos alunos foram, também, determinantes para o contacto com a comunidade portuguesa em Andorra. As entrevistas foram realizadas entre fevereiro e junho de 2024, durante duas deslocações ao Principado de Andorra.

5.2 **Os informantes**

As 27 entrevistas foram realizadas seguindo os pressupostos da Sociolinguística Variacionista. A amostra está estratificada de acordo com as seguintes variáveis sociais: escolaridade (até 12 anos de escolaridade/ ensino superior), idade (entre 16-25, 26-44 e >44) e género. O tempo de permanência em Andorra, expresso em anos, foi, também, considerado para efeitos de tratamento estatístico e posterior análise quantitativa. Para além desses dados, foram, ainda, registados os seguintes elementos: o concelho e o distrito de origem dos informantes, a atividade profissional exercida e os contextos de uso das línguas.

Tabela 1

Número de Informantes Entrevistados por Variável Social (Escolaridade, Intervalo Etário, Género)

Escolaridade	**Até ao 12.º ano**						**Licenciatura**			
Idade	**16–25**		**26–44**		**>44**		**26–44**		**>44**	
Género	F	M	F	M	F	M	F	M	F	M
Informantes PLM	--	--	3	3	3	3	3	--	3	--
Informantes PLH	3	3	--	--	--	--	--	3	--	--

Os informantes entrevistados podem ser agrupados em duas categorias. Por um lado, temos 18 falantes nascidos em Portugal, cuja língua materna é o português, e que emigraram para Andorra depois dos 16 anos. Quando se lhes perguntou se haviam aprendido catalão em contextos de ensino formal, 9 falantes de português língua materna responderam que sim e 9 informantes responderam que não. Os falantes que aprenderam catalão trabalham em serviços públicos e serviços administrativos ou técnicos em empresas privadas. Quatro são do sexo masculino e cinco do sexo feminino. Os falantes que não aprenderam catalão trabalham na construção civil, no atendimento ao público em estabelecimentos comerciais e em serviços de limpeza. Sete são do sexo feminino e dois do sexo masculino. Destes, só uma informante tem formação superior.

Para além dos falantes de PLM, foram entrevistados 9 informantes nascidos em Andorra, filhos de portugueses. Consideram-se falantes de herança os que crescem em contexto de migração e adquirem, a par da língua dominante, a língua de origem da sua família (a língua de herança) desde a nascença, normalmente em contexto familiar e informal (Rinke et al., 2019).

Os falantes PLH que entrevistámos são informantes trilingues (catalão, espanhol e português), têm entre os 16 e os 25 anos, do sexo feminino (3 informantes) e masculino (3 informantes). Foram, ainda, entrevistados informantes do sexo masculino (3 informantes), com um curso superior, entre os 26 e os 44 anos. Esta bipartição da amostra resultou da impossibilidade de entrevistar falantes de PLM com menos de 26 anos. Note-se que o objetivo inicial do projeto era entrevistar somente falantes de português língua materna. No entanto, um dos constrangimentos sentidos, ao longo da recolha, foi o contacto com informantes com as características sociolinguísticas desejadas, pelo que se optou por entrevistar falantes de português língua de herança e considerar esta variável na análise dos dados. Independentemente de terem o português como língua materna ou de herança, não foi possível identificar e entrevistar informantes do sexo masculino com um curso superior e com mais de 44 anos.

5.3 **Transcrição, codificação e análise**

As entrevistas foram transcritas preservando as instâncias de alternância linguística e eventuais regionalismos presentes no discurso espontâneo dos informantes. Para garantir o anonimato das entrevistas, a cada informante foi atribuído um código alfanumérico. Considere-se o informante 4F3SE. O primeiro dígito corresponde ao número da entrevista. A letra F refere-se ao género do informante (feminino). O número seguinte diz respeito ao intervalo etário (1 para os informantes entre 16 e 25 anos, 2 para os informantes entre 26 e 44 anos e 3 para os informantes com mais de 44 anos). O código termina com a indicação do grau de escolaridade: SE indica informantes com um grau de escolaridade até ao 12º ano e SU indica que o informante é detentor de um curso superior.

Após a transcrição, as entrevistas foram compiladas na ferramenta *Sketch Engine* (Kilgarriff et al., 2014[9]). Esta ferramenta de análise de *corpus* permite extrair informação metalinguística e obter dados de frequência. Possibilita, ainda, buscas por termos, categoria gramatical e contexto sintático, entre outras funcionalidades. Em paralelo, foram identificadas e codificadas as alternâncias linguísticas de acordo com as variáveis sociais e linguísticas consideradas relevantes para o presente estudo. A análise quantitativa foi realizada no software IBM SPSS Statistics. Considerando que os dois grupos de informantes (PLM e PLH) não partilham as mesmas características sociodemográficas, os resultados não são diretamente comparáveis e, como tal, a análise estatística foi realizada em separado.

As variáveis sociais consideradas pertinentes foram: o género, a idade, a escolaridade, os anos em Andorra, a idade do informante quando emigrou para Andorra, a língua de trabalho e a aprendizagem do catalão em contextos formais de ensino. No caso dos informantes nascidos em Andorra, considerou-se, ainda, o sistema educativo que frequentou ou frequenta em Andorra.

As variáveis linguísticas consideradas foram: a tipologia de alternância (Poplack, 2020), a extensão da alternância (item lexical, sintagma, oração, expressão fixa), a categoria gramatical (aplicável apenas aos itens lexicais isolados) e a língua doadora envolvida na alternância.

Após codificação, os dados foram revistos por uma falante nativa de espanhol e catalão.

6. **Resultados**

Nesta secção, são apresentados os resultados da análise às alternâncias linguísticas. É, ainda, feita uma breve reflexão relativamente às perceções linguísticas.

Considerando que a amostra é constituída por informantes nascidos em Portugal e que têm o português como língua materna (PLM) e por informantes nascidos em Andorra, filhos de pais portugueses, que têm o português como língua de herança (PLH), os dados não podem ser analisados conjuntamente. Assim, os resultados serão apresentados fazendo referência aos dados recolhidos em entrevistas realizadas a falantes de português língua materna, que, por conveniência, etiquetamos de subcorpus I, e aos dados recolhidos em entrevistas realizadas a falantes PLH, aos quais nos referimos como subcorpus II.

O subcorpus I (falantes PLM) é constituído por 18 entrevistas. Foram identificadas 532 ocorrências de alternâncias linguísticas. Não há diferenças significativas a observar na distribuição das alternâncias por género. No que diz respeito ao subcorpus II (falantes PLH), foram recolhidas 9 entrevistas e identificadas 464 ocorrências de alternância linguística.

No que diz respeito às línguas das alternâncias (cf. Tabela 2), verificamos que o padrão de ocorrências das alternâncias de acordo com a língua é semelhante no discurso de falantes do PLM e do PLH, ainda que haja ligeiras diferenças percentuais. Assim, no caso dos falantes de PLM, 44,5% das alternâncias envolvem itens lexicais comuns ao catalão e ao espanhol, 38,9% das alternâncias são feitas com o espanhol, 10,9% com o catalão, 4,9% com o inglês e apenas 0,8% com o francês. No discurso dos falantes de PLH, 45,9% das alternâncias são feitas com o espanhol e 43,5% com termos lexicais ou expressões comuns ao espanhol e o

[9] Disponível em http://www.sketchengine.eu

catalão. Apenas 8,4% das alternâncias ocorrem com termos exclusivos do catalão. Uma minoria das alternâncias envolve o inglês (2,2%). Não se registam alternâncias com o francês neste grupo de falantes.

Tabela 2

Número e Percentagem de Ocorrência das Alternâncias, de Acordo com a Língua e o Subcorpus *em Análise*

	Subcorpus I - PLM		**Subcorpus II - PLH**	
Línguas de alternância	**N.**	**%**	**N.**	**%**
Espanhol	237	44,5%	202	43,5%
Espanhol + Catalão	207	38,9%	213	45,9%
Catalão	58	10,9%	39	8,4%
Inglês	26	4,9%	10	2,2%
Francês	4	0,8%	0	0
Total	**532**	**100**	**464**	**100**

No que diz respeito às alternâncias em inglês, trata-se de itens lexicais cujo uso está convencionado e, em alguns casos, dicionarizado. Observem-se os seguintes exemplos:

(8) Sim, também hay un tema que também é os ***youtubers*** [25M3SE]

(9) Porque a segurança social dá-te o dinheiro do ***ticket*** do gasol. [6F2SE]

(10) Lá está, se eu estivesse a estudar fora, uma coisa que eu não queria, para fazer o meu ***máster***, ia ser muito pesado e ia desistir. [20M1SE]

(11) Vivi num ***camping,*** num ***bungalow.*** [4F3SE]

A integração destes termos do inglês no discurso é ilustrativa do que De Decker e Vandekerckhove (2012) apelidam de alternâncias "não intencionais", na medida em que este uso não resulta de uma escolha deliberada, pois não existe uma tradução do termo em português, espanhol ou catalão ou o seu uso não é produtivo, como se verifica com *youtubers* ou *bungalow*. Outros termos em inglês, como *máster* ou *camping*, estão dicionarizados em espanhol e em catalão[10].

No caso das quatro alternâncias com o francês, verifica-se que se registam no discurso de uma informante que relata um episódio ocorrido em França (exemplo (12)) aquando do registo de nascimento do filho mais velho. Por esta razão, estes exemplos foram excluídos da análise de resultados.

(12) Escreveram-nos uma direção num papel. Incorreta. Lá fomos nós para a França com aquela direção. Rua para aqui, rua para acolá. Amostravamos o papel... à ***droite***, à ***gauche*** e à ***droite*** e à ***gauche***. E para aqui.., E lá andávamos... E Deus... eu acho que é Deus... nos orientou lá. E registramo-lo lá. [4F3SE]

De igual modo, considerando as especificidades da inclusão dos termos em inglês no discurso, aliadas à percentagem residual de ocorrência, os exemplos de alternâncias linguísticas envolvendo o inglês foram excluídos da análise que doravante se apresenta.

[10] O uso de itens lexicais provenientes do inglês no espanhol e no catalão não é raro. A este propósito, veja-se Lorenzo Criado (1996) e Freixa Aymerich (2019).

Atendendo apenas às alternâncias com o espanhol e o catalão (Figura 2 e

Figura **3**), existe uma preponderância de empréstimos lexicais (85,5% nos falantes de PLM e 70,7% nos falantes de PLH). Apenas uma minoria das alternâncias envolve a inserção de constituintes frásicos ou oracionais (5,3% no discurso dos falantes de PLM e 7,3% no discurso dos falantes de PLH). Um dado relevante é a percentagem de ocorrência de convergência gramatical. Esta tipologia de alternância representa 22% do total de alternâncias no discurso dos falantes de PLH e apenas 9,2% no discurso dos falantes de PLM.

Figura 2

Tipologia de Alternância Linguística no Discurso dos Falantes PLM

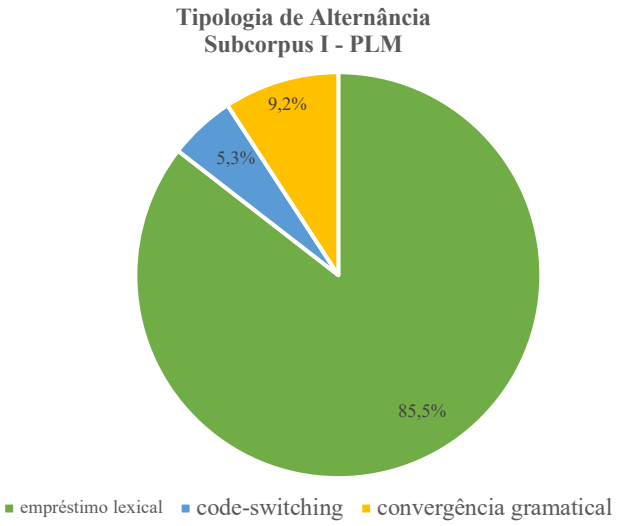
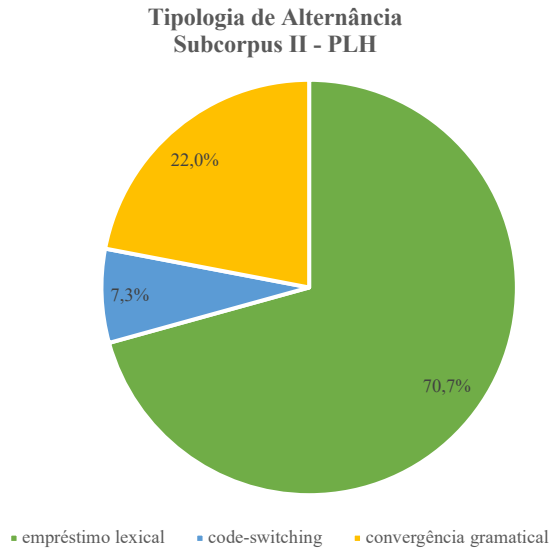


Figura 3

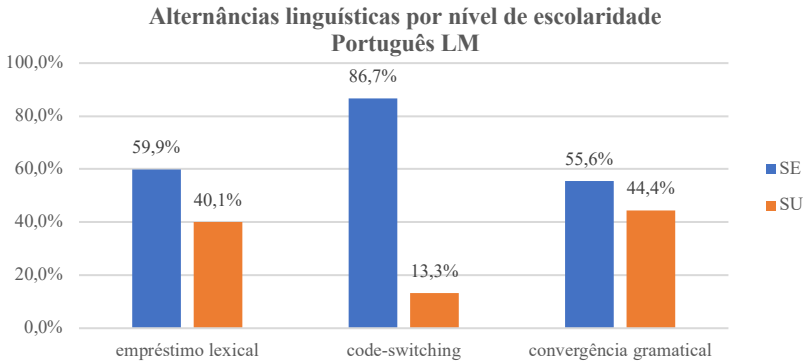
Tipologia de Alternância Linguística no Discurso dos Falantes PLH



Ainda no que diz respeito à tipologia de alternância linguística no discurso dos falantes de PLM, analisámos a sua distribuição por nível de escolaridade (falantes com escolaridade até ao 12º ano (SE) vs. falantes com um curso superior (SU)). Considerando a distribuição das entrevistas pela variável *nível de escolaridade*, as amostras foram equilibradas. Verificamos que, em termos percentuais, as alternâncias linguísticas são mais frequentes nos falantes com escolaridade até ao 12º ano, independentemente da tipologia de alternância (cf. Figura 4). Adicionalmente, é nestes falantes que a percentagem de *code-switching* é mais frequente (86,7% vs. 13,3%).

Figura 4

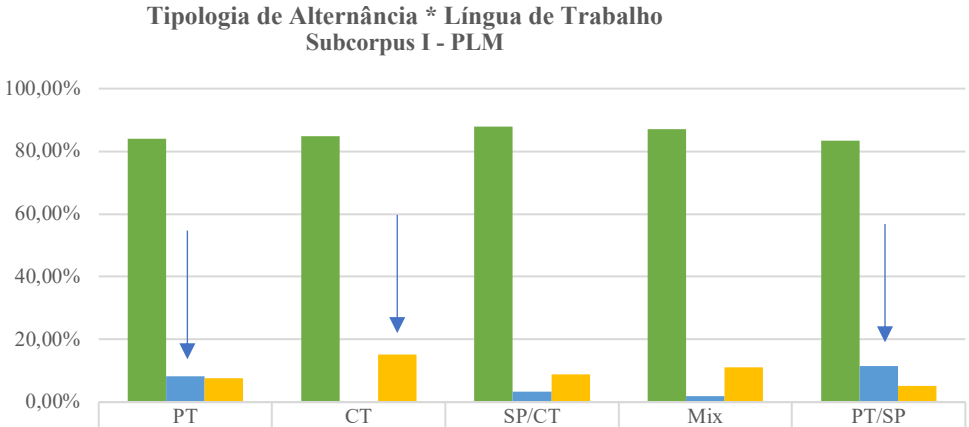
Tipologia de Alternâncias Linguísticas por Nível de Escolaridade (SE – até ao 12.º Ano, SU – com um Curso Superior) no Discurso de Falantes de Português LM



Outro aspeto analisado foi a influência da língua de trabalho. A Figura 5 ilustra a distribuição das três tipologias de alternância linguística de acordo com a(s) língua(s) que os informantes falam em contexto laboral (PT – português, CT – catalão, SP/CT – espanhol e catalão, Mix – português, espanhol, catalão e/ou francês, PT/SP- português e espanhol).

Figura 5

Tipologia de Alternâncias Linguísticas de Acordo com a Língua de Trabalho (Falantes de PLM).



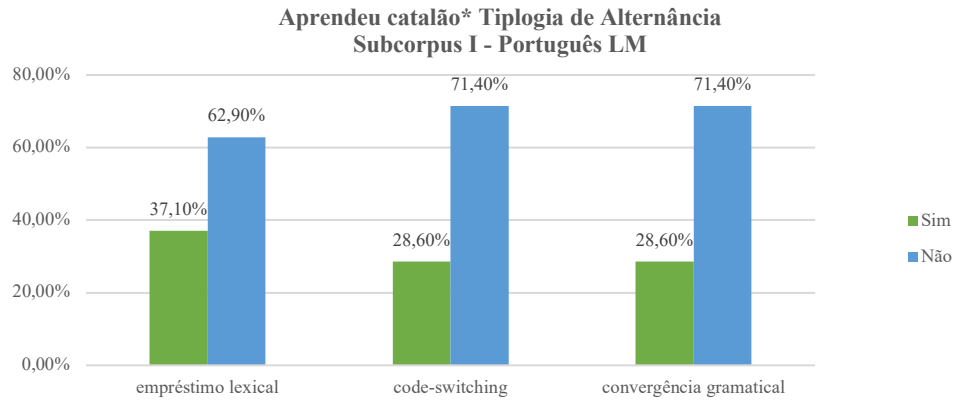
	PT	CT	SP/CT	Mix	PT/SP
empréstimo lexical	84,10%	84,80%	87,80%	87,00%	83,30%
code-switching	8,30%	0,00%	3,30%	1,90%	11,50%
convergência gramatical	7,60%	15,20%	8,90%	11,10%	5,10%

Independentemente da língua de trabalho, as instâncias de empréstimo lexical são a tipologia de alternância mais frequente. Este dado é decorrente da frequência de ocorrência desta tipologia de alternância e a língua de trabalho dos informantes não é relevante para explicar a sua preponderância. Há, no entanto, aspetos que merecem referência. A percentagem de ocorrência de *code-switching* é maior no discurso dos informantes que usam o português como língua de trabalho (8,3% no discurso dos informantes que usam português como língua de trabalho e 11,5% no discurso dos informantes que usam tanto o português como o espanhol como língua de trabalho). Outro dado relevante é a percentagem de ocorrência de instâncias de convergência gramatical no discurso dos informantes que falam exclusivamente catalão no seu local de trabalho (15,2%). De referir, ainda, que a variável *língua de trabalho* está correlacionada com a profissão exercida e com a aprendizagem do catalão. Uma vez que o catalão é a língua usada nos serviços públicos, incluindo os de saúde, os informantes que exercem funções públicas ou trabalham no hospital ou em outros serviços de saúde são falantes de catalão.

A Figura 6 diz respeito à distribuição das alternâncias linguísticas de acordo com a aprendizagem formal do catalão (considerando apenas os falantes de PLM).

Figura 6

Tipologia de Alternâncias Linguísticas de Acordo com a Aprendizagem Formal do Catalão (Falantes de PLM)



Os dados revelam que os falantes que não aprenderam o catalão em cursos de línguas exibem uma maior percentagem de alternâncias linguísticas independentemente da tipologia. Os dois grupos em análise (os que aprenderam catalão e os que não aprenderam) exibem, no entanto, o mesmo padrão no que diz respeito à distribuição das alternâncias, não sendo possível, do ponto de vista estatístico, afirmar que a aprendizagem formal do catalão influencia a tipologia de alternância presente no discurso ($\chi2(2)= 2.102$, $p= 0.35$)[11].

Considerando as características da amostra (secção 5.2) e a correlação entre a profissão exercida e o domínio da língua catalã, não é possível afirmar categoricamente se os resultados obtidos são influenciados pela aprendizagem formal do catalão ou pela profissão exercida.

Analisámos, ainda, a influência da variável "anos em Andorra". A

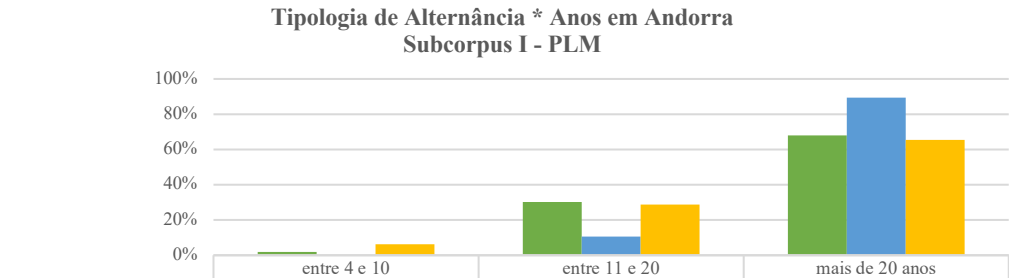
[11] De forma a aferir se a distribuição da tipologia de alternâncias pode ser associada às características sociodemográficas dos informantes considerados, usámos o teste do Qui-Quadrado. Para efeitos de clareza na apresentação dos resultados, é referido o valor de prova (p-value). Quando o valor de prova é superior a 5% (>0,05), as variáveis são independentes.

Figura **7** mostra a distribuição das tipologias de alternância de acordo com três intervalos de tempo: entre 4 e 10 anos de permanência em Andorra, entre 11 e 20 anos, e mais de 20 anos, independentemente do grau de escolaridade[12]. Os dados mostram que o tempo de residência em Andorra é um fator relevante para explicar a presença de alternâncias linguísticas no discurso, principalmente no que diz respeito ao aumento de ocorrências do tipo *code-switching.*

[12] O cruzamento das variáveis *tempo em Andorra* e *escolaridade* não permite extrair conclusões robustas do ponto de vista estatístico, uma vez que a maioria dos informantes que residem há mais tempo em Andorra não possui curso superior. Esta característica da amostra decorre das especificidades dos diferentes fluxos migratórios que caracterizam a comunidade portuguesa em Andorra (cf. secção 2).

Figura 7

Tipologia de Alternâncias Linguísticas de Acordo com os Anos de Residência em Andorra (Falantes de PLM)



	entre 4 e 10	entre 11 e 20	mais de 20 anos
empréstimo lexical	1,80%	30,30%	67,90%
code-switching	0,00%	10,70%	89,30%
convergência gramatical	6,10%	28,60%	65,30%

No que diz respeito às variáveis linguísticas, as categorias gramaticais mais frequentes nos empréstimos lexicais são: nomes (13) (14), (17), verbos (15), interjeições (16), adjetivos (17) e advérbios (18).

(13) É, azeite, **almendra**. É uma **salsa** especial mesmo para **calçots**[13]. [1M2SE]

(14) Ou cantamos alguma coisa em **català**, se por exemplo for o padre que começa a cantar e a gente segue. [2F3SE]

(15) Agora, se **parlas** o catalão perfeito, se falas o catalão perfeito, sim, é melhor. [23M2SU]

(16) **Bueno**, é um país pequeno.

(17) O mal dos espanhóis é que dobravam as **películas**, ou seja, não eram **subtituladas**. [14M3SE]

(18) Ao ser pequeno, nos conhecemos **casi** todos. [28F1SE]

Quando analisamos as instâncias de empréstimo lexical nos dados recolhidos, verificamos que, independentemente do subcorpus em análise, é possível traçar uma tendência no que diz respeito à categoria gramatical alvo de alternância (cf. **Figura 8**). Verifica-se que os nomes ocorrem mais frequentemente como empréstimo lexical (49,7% no corpus I – PLM e 39,3% no subcorpus II – PLH), seguindo-se a categoria verbo (15,5% no corpus I – PLM e 20,1% no subcorpus II – PLH). A categoria "Outros" compreende locuções adverbiais (19) (20) e expressões (21).

(19) Às vezes misturo um bocadinho. Acho que uma palavra de cada idioma, depende de como...**ao melhor**[14] acabas por misturar e eu digo, ah, mas já nem sei o que é que eu estou a falar. [16F2SE]

[13] *Calçots* são cebolas brancas, tradicionalmente comidas assadas. No entender de Blas Arroyo (2021), as alternâncias linguísticas envolvendo termos para designar comida ou bebida, expressões idiomáticas ou fórmulas de saudação, correspondem a *tag-switches*. Funcionam, no entender do autor, a uma estratégia de acomodação, solidariedade ou sinal de identidade coletiva, podendo inclusive ocorrer no discurso de monolíngues com uma competência passiva na língua ou variedade dominante.

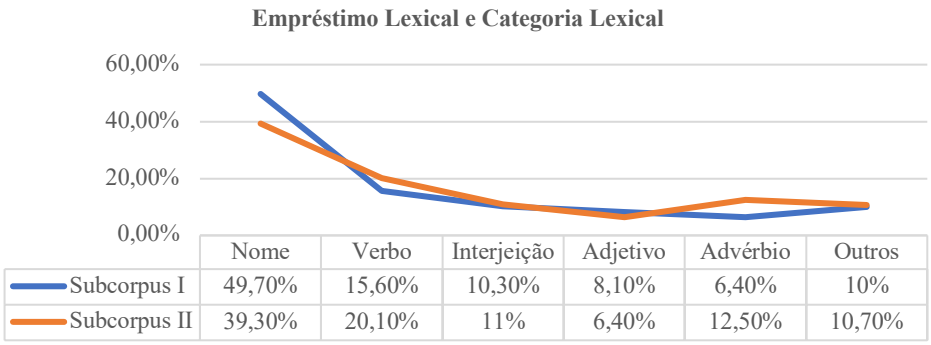
[14] Do espanhol *a lo mejor*. Locução adverbial coloquial com o significado de *talvez* (Real Academia Española, s.d.)

(20) **Hoje por hoje**[15] eu já vejo muitas mulheres que estão [a trabalhar] em fábricas, enfermeiras. Vejo que [Portugal] evoluiu muito, está bastante mais avançado do que quando eu estava lá. Quando eu estava lá era o campo, campo e campo. [2F3SE]

(21) Aí tive [estive] a ir a trabalhar 20 anos! **E nada**[16]... e é assim [4F3SE]

Figura 8

Empréstimo Lexical: Percentagem de Ocorrência das Categorias Lexicais nos Dois Subcorpora



	Nome	Verbo	Interjeição	Adjetivo	Advérbio	Outros
Subcorpus I	49,70%	15,60%	10,30%	8,10%	6,40%	10%
Subcorpus II	39,30%	20,10%	11%	6,40%	12,50%	10,70%

No que diz respeito à convergência gramatical, os exemplos mais frequentes no *corpus* em análise compreendem:

i. a próclise em contextos típicos de ênclise (10 ocorrências no subcorpus I e 25 no subcorpus II):

(22) Eu **me identifico** mais como *andorrà*. [28F1SE]

(23) E depois, **te fica** muito pouco para comer [24F3SU]

ii. processos que envolvem o sintagma preposicional (9 ocorrências no subcorpus I e 34 no subcorpus II), nomeadamente a seleção da preposição (24) e a determinação do nome no interior do sintagma preposicional (25):

(24) O único susto, estava a vir de treinar futebol e caiu um javali do monte **ao** meio da estrada. [8M1SE]

(25) A minha primeira viagem a Portugal, fi-la eu a última vez que fui, **em Páscoa,** que aliás conduzi a metade da viagem, 5 horitas. [20M1SE]

iii. a concordância de género (5 ocorrências - subcorpus II)

[15] Do espanhol *hoy por hoy*. Locução adverbial usada para indicar que algo é ou acontece de uma certa forma agora, mas que pode mudar mais tarde (Real Academia Española, .d.).
[16] Concorreu para a classificação de empréstimo lexical a alta frequência de ocorrência da expressão *y nada* em espanhol, a par de *nada* e *pues nada* (Enghels & Tanghe, 2019), principalmente no discurso dos falantes mais jovens, assim como a pronúncia do informante ao produzir este segmento do discurso.

(26) eu quando vou lá em Portugal, eu tenho a sensação de que a gente reparam para mim como aquele estrangeiro que chegou lá e, pronto, eu faço **aquelas pontapés** no dicionário e tudo. [26M2SU]

(27) [Café com gelo] É **uma costume muito espanhola,** de quando vemos o primeiro raio do sol, o que significa verão ou calor, toma-se assim. [20M1SE]

No exemplo (24), o informante 8M1SE faz referência a uma situação em que um javali caiu (das montanhas) para o meio da estrada. Neste exemplo, verifica-se a influência do catalão, língua que, neste contexto, seleciona *al* (contração da preposição *a* e do artigo *el*): "va caure **al** mig del carrer". No exemplo (25) está em causa a determinação do nome, através da ausência do determinante artigo. Neste caso, o uso da preposição simples *em* em vez da preposição contraída *na* é o resultado da influência do espanhol, língua em que o nome de festividades religiosas como o *Natal* ou a *Páscoa* é antecedido da preposição simples *em,* em vez da forma contraída como se verifica em português.

Por fim, os exemplos (26) e (27) ilustram a influência da L_D no que diz respeito à atribuição de um valor de género indevido. Nestes casos, o falante, por influência do catalão e/ou espanhol, altera o género dos elementos que concordam com o nome. No exemplo (26), o nome masculino *pontapés* concorda com o determinante demonstrativo feminino *aquelas*, resultado da influência dos nomes femininos *patadas* (em espanhol) e *puntadas* (catalão). No exemplo (27), por influência de *costumbre* (nome feminino) do espanhol, o nome masculino *costume* concorda com o determinante artigo feminino *uma* e com o adjetivo feminino *espanhola*.

Para além da análise das alternâncias linguísticas, este trabalho aflora, ainda, as perceções da comunidade portuguesa em Andorra em relação ao uso do português e ao uso das línguas dominantes: o catalão e o espanhol. Os relatos recolhidos reiteram o facto de a comunidade portuguesa em Andorra ter orgulho em falar em português e nos costumes portugueses. Os imigrantes portugueses em Andorra vêm a Portugal várias vezes por ano, normalmente durante as férias escolares, ouvem estações de rádio portuguesas, veem televisão em português e sentem orgulho no facto de os filhos saberem falar português, como relatam os informantes 1M2SE (28) e 16F2SE (29):

(28) Mas connosco em casa [os filhos falam] sempre português, porque [o catalão] foi uma das coisas que eu nunca entendi muito bem. Eu também não falo muito bem espanhol. [1M2SE]

(29) Em casa sempre, em casa sempre falamos [português]. De feito, os meus filhos vão a Portugal e todo mundo me diz, ai os teus miúdos, que lindos estão, porque eles falam português e eles entendem-nos e nós entendemos bem eles. Porque há muitos, por exemplo, eu tenho primos da França, que eles não falam mais que francês para os miúdos. [16F2SE]

Os falantes de PLM consideram vital o uso da língua portuguesa no contexto familiar. A língua portuguesa é, assim, tida como um bem simbólico **(**Bourdieu, 1991)**,** um legado imaterial que pais transmitem aos filhos. O facto de os filhos falarem português é motivo de orgulho, apesar de eles próprios terem consciência de que já misturam o português com o espanhol e o catalão (exemplos (30), (31) e (32)).

(30) Por exemplo, há palavras, que eu sou sincera, e há palavras em português que eu agora já não sei o significado delas. [2F3SE]

(31) Os meus filhos falam português, sempre falaram português. Também já... mas nem eu nem eles falamos um português já correto, correto. Também não há... em Portugal também...o minhoto também é diferente. [4F3SE]

(32) Às vezes misturo um bocadinho. Acho que uma palavra de cada idioma, depende de como... ao melhor, acabas por misturar. E eu digo, ah, mas já nem sei o que é que eu estou a falar. Não sei se estou a falar de uma coisa, se estou a falar de outra. [16F2SE]

No caso dos falantes PLH que já concluíram os seus estudos em Andorra ou em Espanha e que já constituíram família em Andorra, a língua portuguesa funciona como uma ligação simbólica às origens (33).

(33) vamos todos os anos para Portugal, porque eu acho que é importante que os meus filhos tenham aquela ligação, não é? (...) A língua portuguesa lá vai, está no ar, não é todos os dias, mas, às vezes, procuro para eles, para que eles tenham aquela ligação com Portugal, com o português, não é sempre, mas procuro. [26M2SU]

Os portugueses de 2ª e 3ª gerações adquiriram o português como língua de herança (e não como L2[17]). O *input* foi maioritariamente oral e informal, pelo que alguns informantes, como o informante 26M2SU, relatam não terem "domínio absoluto" da língua (34):

(34) vou falar consigo em português, faz-me ilusão, mas não é uma língua que eu tenha a sensação de... tenha a perceção de domínio absoluto. [26M2SU]

Apesar de todos os informantes PLH terem frequentado aulas de português, oferecidas pelo Instituto Camões como disciplina extracurricular, tipicamente até aos Níveis Comuns de Referência B1 ou B2, gostariam de ter podido aprofundar mais o conhecimento linguístico da língua e o domínio da produção escrita, principalmente o registo formal (35).

(35) [Informante - 30F1SE] – Eu, quando era mais pequena, sim, que fui a aulas de português e se não me engano, tenho o B2. É muito importante pôr um nível mais alto, porque parece que somos já... tipo, filhos de portugueses, o A1 não nos serve. Eu gostei muito [das aulas de português], mas também acho que não aprendi grande coisa, porque, tipo, sou uma pessoa que fala bastante português, que tem toda a sua família portuguesa, então falo muito português. Mas se fossem, tipo, níveis mais altos, sim, que iria aprender. Já não o vocabulário, porque o vocabulário praticamente temos muito. Sino... tipo, a forma de falar, se calhar, ou de referir-se a pessoas mais importantes.
[JA] O português mais formal.
[Informante - 30F1SE] Sim.

7. Discussão

A análise das três tipologias de alternâncias (empréstimo lexical, *code-switching* e convergência gramatical) permite concluir que falantes de português LM e LH exibem padrões de alternância linguística semelhantes, apesar de a frequência de ocorrência ser diferente. O cruzamento da frequência de ocorrência das alternâncias linguísticas e das variáveis sociais permite enquadrar os resultados obtidos. Mais especificamente, nos falantes de PLM, o tempo de residência em Andorra e a aprendizagem formal do catalão são fatores relevantes para explicar os padrões de alternância observados. O nível de escolaridade é, também, uma variável determinante, pois falantes com um curso superior exibem uma percentagem menor de alternâncias no discurso. De referir que, considerando as características da comunidade portuguesa em Andorra, as variáveis sociais *anos*

[17] Veja-se Rinke et al. (2019) a propósito das implicações que as diferentes fontes de *input* têm no desenvolvimento da gramática mental do falante.

de residência em Andorra, nível de escolaridade e *profissão* estão muitas vezes interligadas, sendo necessário alargar a amostra para obter resultados significativos do ponto de vista estatístico. Assim, os informantes que residem há mais tempo em Andorra têm tipicamente menos anos de escolaridade e trabalham na construção civil, restauração ou em serviços de limpezas. Estes informantes comunicam normalmente em português ou "portunhol" e não aprenderam o catalão em contextos formais, como refere o informante [1M2SE]:

> (36) A base é português. Nós aqui em Andorra, verdadeiramente, se não quisermos falar outra língua durante o ano, não precisamos. E então não há necessidade mesmo de aprender. (...) Que deveríamos, deveríamos. Eu é uma coisa que tenho assente, não faço, mas deveríamos. Porque estamos num país que não é o nosso, devíamos falar a língua do país. [1M2SE]

É, também, no discurso dos informantes a residir há mais anos em Andorra que identificamos uma maior percentagem de alternâncias lexicais. Este resultado reitera as afirmações impressionistas decorrentes de um trabalho sociológico realizado na comunidade portuguesa em Andorra (Matias, 2008). Segundo a autora, "Estas variações linguísticas, na introdução de palavras numa língua diversa daquela com que se produz o discurso, estão presentes sobretudo nos emigrantes com mais anos de residência ou com menor nível de estudos" (Matias 2008, p. 62). Por outro lado, os informantes com um curso superior trabalham em serviços públicos ou empresas privadas e aprenderam catalão (apenas uma informante diz não ter aprendido catalão em contextos formais). Estes informantes são críticos relativamente às gerações anteriores e ao facto de não dominarem a língua do país de acolhimento.

No que diz respeito aos *empréstimos lexicais* e às categorias lexicais envolvidas, o padrão observado tanto no discurso dos falantes de PLM como nos falantes de PLH foi já identificado noutros trabalhos sobre alternâncias linguísticas. Veja-se, por exemplo, Torres Cacoullos e Travis (2020, p. 259), segundo os quais "single-word insertions are disproportionately nouns". Distribuição semelhante é observada em Poplack (2018), num trabalho sobre os empréstimos lexicais do inglês no discurso em francês canadiano: nomes (64%), verbos (14%), interjeições e expressões fixas (12%), adjetivos (8%). Ainda no que diz respeito aos empréstimos lexicais, verificou-se que, nos falantes de PLH, os empréstimos lexicais exclusivos do catalão ocorrem com mais frequência nas produções dos falantes que frequentaram a escola andorrana (49,1%) do que nas produções dos informantes que frequentaram a escola espanhola (10,3%) ou a escola francesa (36,8%).

Em relação à categoria convergência gramatical, como referimos na secção anterior, os processos mais frequentes nas entrevistas recolhidas são: (i) a ênclise em contextos de próclise, (ii) a seleção da preposição e a determinação do nome no interior do sintagma preposicional e (iii) a atribuição de um valor de género indevido e a consequente ausência de concordância de género. Estes resultados encontram eco no trabalho de Scetti (2016) sobre a comunidade portuguesa em Montreal. De acordo com o autor, é possível identificar sete "pontos de mudança" no português falado em Montreal, resultado do contacto com o francês e o inglês, entre os quais o uso da próclise em contextos tipicamente de ênclise e ausência de concordância de género em construções N_{fem} + Adjetivo. Neste contexto, o adjetivo surge frequentemente na forma masculina. De acordo com o autor, estes processos são mais frequentes em falantes mais jovens, que, no trabalho de Scetti (2016) correspondem a falantes de PLH).

No que diz respeito aos clíticos pronominais, como é sabido, em português europeu, a sua colocação depende principalmente de fatores sintáticos (Martins 2015, 2016), que sintetizamos:

i. A polaridade da frase: em frases afirmativas, observa-se a ênclise; em frases de polaridade negativa, o clítico pronominal é atraído para posição pré-verbal;

ii. A ocorrência, antes do verbo, de quantificadores, sintagmas QU- e alguns advérbios (*até, só, lá enfático, também, sempre, já, talvez*);

iii. A oposição oração principal / oração subordinada finita: em estruturas subordinadas, o clítico ocupa a posição pré-verbal;

Assim, na ausência de atratores de próclise ou de estruturas sintáticas que permitam ou exijam a próclise, o clítico pronominal ocupa, em português europeu contemporâneo, a posição enclítica[18]. Em espanhol e catalão, a posição do clítico é condicionada pela morfologia verbal: a ênclise ocorre com o infinitivo, o gerúndio e o imperativo, ao passo que a próclise ocorre geralmente com as formas finitas (Martins, 2016).

Como refere Martins (2015, p. 84), "o conceito de proclisador/palavra proclisadora aplica-se ao português europeu (e ao galego), mas não às outras línguas românicas, porque nas orações principais afirmativas ocorre normalmente a ênclise e só em casos particulares a próclise (concretamente quando uma palavra proclisadora precede o verbo)". Considerando os dados apresentados, assim como os trabalhos de Scetti (2016), podemos afirmar que, no que diz respeito à ordem, os clíticos pronominais parecem ser elementos instáveis e permeáveis a processos de interferência estrutural decorrentes do contacto de línguas. Ademais, é sabido que, do ponto de vista diacrónico, são elementos sujeitos a processos de variação[19] (Martins, 1994 e seguintes).

Considerando os dados extraídos das entrevistas recolhidas, verifica-se que a próclise em contextos típicos da ênclise ocorre principalmente em falantes PLH (25 ocorrências). As dez ocorrências identificadas no discurso de falantes PLM são produzidas por quatro falantes (três do sexo feminino e um do sexo masculino) a residir em Andorra há mais de 13 anos. Nenhum aprendeu catalão formalmente. Segundo Poplack, os anos de contacto linguístico, assim como o grau de integração do falante na comunidade dominante, são fatores que poderão exponenciar a ocorrência de alternâncias linguísticas e a "interferência estrutural". Como enfatiza a autora, "the longer and more intense the contact, and the fewer and more marginal the speakers, the greater the likelihood of language mixing and structural interference" (Poplack, 2020, p. 49).

Relativamente a processos envolvendo o sintagma preposicional, o trabalho de Rinke et al. (2024) mostra que a seleção da preposição (simples ou contraída) pode ser desafiante para falantes de PLH. O mesmo trabalho faz referência à dificuldade que os falantes bilingues apresentam na atribuição de género (Montrul & Ionin, 2010). Não obstante a complexidade linguística e a opacidade semântica de algumas das estruturas que envolvem preposições (Rinke et al., 2024), quer para falantes bilingues quer para falantes monolingues, os exemplos aqui apresentados ilustram a influência da gramática das línguas dominantes (o catalão e o espanhol) no português.

Apesar de a escassez de dados não permitir extrair conclusões robustas relativamente ao efeito do ensino formal da língua catalã na ocorrência de instâncias de alternância linguística, nomeadamente as do tipo convergência gramatical, este é um aspeto que será aprofundado em publicações futuras.

Outro aspeto aflorado neste trabalho foram as perceções da comunidade portuguesa em Andorra em relação ao uso do português e do catalão. De uma forma geral, os falantes de PLM estão cientes da importância do uso do catalão e do prestígio social que este pode proporcionar. O domínio do catalão é, na realidade, um fator que determina o acesso a oportunidades económicas e sociais. A este propósito, Hawkey e Horner (2022, p. 212) realçam que as políticas monolinguísticas oficiais colidem com a prática multilinguística da sociedade andorrana: "these official de jure state monolingual policies stand in contradiction with widely practiced de facto societal multilingualism.". A *Llei d'ordenació de l'ús de la llengua oficial* (Govern d'Andorra, 2000) determina que o catalão seja a língua da comunicação oficial, assim como dos serviços públicos, incluindo os de saúde. O acesso a empregos públicos e aos quadros técnicos superiores depende, assim, do domínio da língua,

[18] Ao longo da história da língua portuguesa, a colocação dos clíticos pronominais sofreu processos de variação e mudança. A este propósito, veja-se Martins (1994, 2011 e seguintes).

[19] A este propósito, Martins (2011, 2015) propõe a existência, no português antigo, de duas variedades linguísticas: "o português antigo originou dois dialetos diferentes no que diz respeito à colocação dos pronomes clíticos: um em que a ênclise nunca deixou de ser dominante nos contextos de variação e que representa a linha de continuidade entre o português antigo e o português contemporâneo, e outro que representa uma linha evolutiva a que poderíamos chamar «pan-ibérica» e que veio a perder-se, embora pareça ter correspondido à variedade de prestígio no português quinhentista e seiscentista" (Martins, 2015, p. 84).

o que não significa que o catalão seja utilizado em exclusivo, como ilustram os resultados do relatório *Coneixements i usos lingüistics de la población d'Andorra* (Govern d'Andorra, 2023). De acordo com os testemunhos recolhidos, o espanhol substitui frequentemente o catalão nos contextos formais e administrativos. Mesmo em contextos sociais e situações laborais em que o catalão é exigido, o espanhol ou o português é frequente, como referem os informantes 15F3SU e 11M2SE:

> (37) Quando estou de cara ao usuário, como sou enfermeira, obrigam-nos a falar em catalão, que é a língua oficial do país. Como trabalho em saúde laboral, há muita gente de muitos sítios e de muitas origens. E quando começo a ver alguém, por exemplo, português, que está ali a patinar um bocado e que não... pois... troco e falo em português. [15F3SU]
>
> (38) Isto [falar espanhol] está tão mentalizado que mesmo aqueles mais ferrenhos, aqueles mesmo do catalão a sério, veem que se tu respondes em castelhano, eles começam a responder em castelhano. [11M2SE]

Relativamente às perceções dos informantes PLH entrevistados no que diz respeito ao uso e aprendizagem formal do português, o relato da Informante 30F1SE (exemplo 35) é paradigmático e espelha uma das conclusões do trabalho de Rinke et al. (2019). Segundo as autoras, o conhecimento fonológico e lexical de falantes de PLH é semelhante ao dos falantes de PLM. No entanto, falantes de PLH apresentam dificuldades em alguns processos morfológicos e sintáticos. O ensino formal a falantes de PLH deve, assim, incidir sobre processos linguísticos complexos, que ocorrem principalmente no discurso formal e que são, tipicamente, adquiridos mais tarde.

8. Conclusão

Este trabalho constituiu uma primeira abordagem ao mosaico linguístico constituído pela comunidade portuguesa em Andorra, centrada principalmente nas alternâncias linguísticas e na inevitável influência social que as línguas dominantes, o espanhol e o catalão, exercem. A adoção da proposta de Poplack (2020) permitiu observar padrões de alternância relevantes. Por um lado, falantes de PLM e falantes de PLH não exibem padrões de alternâncias distintos: as instâncias de empréstimos linguísticos constituem a maioria das alternâncias nos dois grupos de falantes, seguidas das instâncias de convergência gramatical. Por outro lado, verificou-se que, no discurso dos falantes de PLH, a percentagem de convergência gramatical é superior. No caso dos falantes PLM, de acordo com os dados recolhidos, o tempo de residência em Andorra e o menor grau de escolaridade são fatores que potenciam as instâncias de convergência gramatical. Por fim, a ocorrência dos processos identificados como convergência gramatical, nomeadamente a posição do clítico pronominal, a atribuição de um valor de género indevido e a consequente ausência de concordância de género e a seleção da preposição, não é aleatória, o que nos leva a colocar a hipótese de haver elementos mais permeáveis à influência exercida pelo contacto linguístico, entre outros fatores.

Espera-se que os resultados aqui apresentados tenham uma capacidade multiplicadora, na medida em que poderão constituir o ponto de partida para análises futuras no campo dos estudos sociolinguísticos e da política da língua, assim como para estudos sobre o português língua de herança e o português falado pela comunidade portuguesa no estrangeiro.

Agradecimentos

Agradecemos as observações e sugestões de melhoria propostas pelos revisores anónimos. Qualquer erro ou imprecisão remanescente é de inteira responsabilidade das autoras.

Um agradecimento especial aos informantes, sem os quais este trabalho não seria possível.

Financiamento

Investigação realizada no âmbito da Cátedra Camões (Universitat d'Andorra). A autora Joana Aguiar reconhece a bolsa de investigação recebida da Universidade de Andorra, com a referência AD03 - UdA-2023/2024. Este trabalho foi apoiado pela FCT - Fundação para a Ciência e Tecnologia no âmbito do Centro de Investigação Transdisciplinar em Educação e Desenvolvimento (https://doi.org/10.54499/UIDB/05777/2020).

Referências

Androutsopoulos, J. (2005). Research on youth language. In U. Ammon, N. Dittmar, K. J. Mattheier, & P. Trudgill (Eds.), *Sociolinguistics: An international handbook of the science of language and society* (pp. 1496–1505). De Gruyter.

Ballarín, J. (2006). Identitats culturals i usos lingüístics a Andorra. *Treballs de Sociolingüística Catalana*, *19*, 183–220.

Bastida, C., & Nicolau, M. (2012). Un projecte de prospectiva sociolingüística: El model sistèmic de l'evolució de l'ús del català a Andorra. *Llengua i Ús, 51*, 100–112.

Blas Arroyo, J. (2021). Desenlaces del contacto de lenguas en la Comunidad Valenciana: Entre el cambio de código y el préstamo léxico. In M. Sanz Gil (Ed.), *Estrategias lingüísticas para la sociedad multilingüe* (pp. 13–24). Ediciones Octaedro, S.L.

Bourdieu, P. (1991). *Language and symbolic power* (G. Raymond & M. Adamson, Trans.; J. B. Thompson, Eds.). Harvard University Press.

Carvalho, J. (2007). *Portugueses em Andorra: Uma visão global*. Andbanc.

Comas, D., & Pujadas, J. (1999). Living in/on frontier: Migration, identities and citizenship in Andorra. *Social Anthropology, 7*(3), 253–264.

De Decker, B., & Vandekerckhove, R. (2012). English in Flemish adolescents' computer-mediated discourse: A corpus-based study. *English WorldWide, 33*(3), 321–351.

Departament d'Estadística d'Andorra. (2025). *Nota de premsa A001. Estimacions de població. Març del 2025 i A003. Estadística dels censos parroquials.* https://sig.govern.ad/SIGDDE.Public/Files/Documents/Notes_premsa_noticies/A001_20250313_A.pdf

Enghels, R., & Tanghe, S. (2019). On the interplay between historical pragmatics and sociolinguistics: The case of the Spanish pragmatic marker nada and its recent grammaticalization process. ONOMAZEIN, *44*, 132–165. https://doi.org/10.7764/onomazein.44.07

Flores, C. (2004). «Lá está ele a schmatzen!»: Particularidades do discurso de bilingues luso-alemães. *Diacrítica, 19*(1), 89–108.

Freixa Aymerich, J. (2019). Augment constant de manlleus en la premsa catalana: Una conseqüència de la mundialització? *Treballs de Sociolingüística Catalana*, *29*, 43–64. https://doi.org/10.2436/20.2504.01.154

Govern d'Andorra. (2000). "Llei d'ordenació de l'ús de la llengua oficial". *Butlletí Oficial del Principat d'Andorra*, *12*(2), 66–72. https://www.bopa.ad/Documents/Detall?doc=1B83E

Govern d'Andorra. (2023). *Coneixements i usos lingüística de la población d'Andorra. Situació actual evolució (1995–2022).* Govern d'Andorra. https://www.ari.ad/images/pdf/llengua.pdf

Govern d'Andorra. (2024). "Llei 6/2024, del 25 d'abril, de la llengua pròpia i oficial". *Butlletí Oficial del Principat d'Andorra, 59*. https://www.bopa.ad/Documents/Detall?doc=CGL_2024_05_21_09_12_55

Gumperz, J. (1982). *Discourse strategies*. Cambridge University Press.

Hawkey, J. (2024). "Aquestes arrels que queden sempre et criden una mica": Actituds i ideologies envers la llengua portuguesa a Andorra. *Treballs de Sociolingüística Catalana, 34*, 51–66.

Hawkey, J., & Horner, K. (2022). Officiality and strategic ambiguity in language policy: Exploring migrant experiences in Andorra and Luxembourg. *Language Policy,* 21(2), 195–215. https://doi.org/10.1007/s10993-021-09602-3

Henry, A. (2002). Variation and syntactic theory. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *Handbook of language variation and change* (pp. 267–282). Blackwell.
Kilgarriff, A., Baisa, V., Busta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*, 7–36.
Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.
Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
Labov, W. (2001). *Principles of linguistic change: External factors*. Blackwell.
Lorenzo Criado, E. (1996). *Anglicismos hispânicos*. Gredos.
Margarit, E., Micó, J., & Monné, A. (2010). La interrelació dels joves amb les llengües i els grups lingüístics presents al Principat d'Andorra. In O. Guasch & M. Milian (Eds.), *L'educació lingüística i literària en entorns multilingües: Recerca per a nous contextos* (pp. 331–346). Universitat Autònoma de Barcelona - Servei de Publicacions.
Martins, A. M. (1994). *Clíticos na história do português*. [Dissertação de Doutoramento, Universidade de Lisboa].
Martins, A. M. (2011). Clíticos na história do português à luz do teatro vicentino. *Estudos de Lingüística Galega, 3,* 83–109.
Martins, A. M. (2015). Variação sintática no português quinhentista: a colocação dos pronomes clíticos. *Estudos De Lingüística Galega*, *7,* 83–94. https://doi.org/10.15304/elg.7.2373
Martins, A. M. (2016). A colocação dos pronomes clíticos em sincronia e diacronia. In A. M. Martins & E. Carrilho (Eds.), *Manual de linguística portuguesa* (pp. 401–430). De Gruyter.
Matias, M. (2008). *A que sabe um Sumol a 2.200 metros de altitude?*. Pagès Editora.
Montrul, S., & Ionin, T. (2010). Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism: Language and Cognition, 13*(4), 449–473. https://doi.org/10.1017/S1366728910000040
Observatório da Emigração. (2010). *Portugueses em Andorra.* https://observatorioemigracao.pt/np4/1667.html
Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics, 18*, 582–617.
Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar.* Oxford University Press.
Poplack, S. (2020). A variationist perspective on language contact. In E. Adamou & Y. Matras (Eds.), *Routledge handbook of language contact* (pp. 46–62). Routledge.
Real Academia Española. (s.d.). *a lo mejor.* In *Diccionario de la lengua española*. Acedido a 23 abril 2025, em https://dle.rae.es/a%20lo%20mejor
Real Academia Española. (s.d.). *hoy por hoy*. In *Diccionario de la lengua española*. Acedido a 23 abril 2025, em https://dle.rae.es/hoy%20por%20hoy
Real Academia Española. (s.d.). *sí que*. In *Diccionario de la lengua española*. Acedido a 23 abril 2025, em https://dle.rae.es/sí%20que
Rinke, E., Flores, C., & Santos, A. L. (2019). Heritage languages at school: Implications of linguistic research on bilingualism for heritage language teaching. In C. Gabriel, J. Grünke, & S. Thiele (Eds.), *Romanische Sprachen in ihrer Vielfalt: Brückenschläge zwischen linguistischer Theoriebildung und Fremdsprachenunterricht* (Romanische Sprachen und ihre Didaktik) (pp. 211–232). Ibidem-Verlag.
Rinke, E., Flores, C., & Torregrossa, J. (2024). How different types of complexity can account for difficult structures in bilingual and monolingual language acquisition. In M. Polinsky & M. T. Putnam (Eds.), *Formal approaches to complexity in heritage languages* (pp. 43–71). Language Science Press. https://doi.org/10.5281/zenodo.12090434

Scetti, F. (2016). The Portuguese language in the particular context of the "Portuguese community" of Montreal. In R. Muhr (Ed.), *Pluricentric languages and nondominant varieties worldwide: The pluricentricity of Portuguese and Spanish* (pp. 261–274). Peter Lang.

Tagliamonte, S. (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Wiley-Blackwell.

Torres Cacoullos, R., & Travis, C. (2020). Code-switching and bilinguals' grammars. In E. Adamou, & Y. Matras (Eds.), *The Routledge handbook of language contact* (pp. 252–274). Routledge. https://doi.org/10.4324/9781351109154

“O que são sons?” – Dados sobre a competência metafonológica de futuros professores

“What are sounds?” – Data on the metaphonological competence of future teachers

Clara Amorim[1], *Adelina Castelo*[2]

[1] Instituto Politécnico de Viana do Castelo; inED, Centro de Investigação e Inovação em Educação & Universidade do Porto, Centro de Linguística (CLUP)

[2] Universidade Aberta, Departamento de Humanidades e Laboratório de Educação a Distância e E-Learning & Universidade de Lisboa, Centro de Linguística (CLUL)

Abstract

This study examines the performance of undergraduate and master’s students enrolled in teacher education programs on tasks assessing phonological awareness, with the goal of identifying gaps in initial teacher training in the field of Phonology. The sample comprises 98 students who completed an online questionnaire containing items on syllabic and phonemic awareness, phoneme–grapheme correspondence, and explicit phonological knowledge. The results indicate substantial limitations in metaphonological knowledge, particularly in the identification of word stress, syllables, sounds, and rhymes, as well as in the explicit explanation of phonological phenomena. Responses also suggest an influence of orthography, with master’s students performing slightly better overall. These findings highlight an urgent need to reinforce explicit instruction in Phonology within teacher education programs.

Keywords: phonological awareness, initial teacher education, phonology, European Portuguese.

Resumo

Este estudo investiga os níveis de desempenho de estudantes em cursos de formação de professores em tarefas de consciência fonológica, com o objetivo de identificar lacunas na formação inicial de professores no domínio da Fonologia. A amostra inclui 98 estudantes de licenciatura e mestrado, que responderam a um questionário *online* com questões sobre consciência silábica, segmental, relação fonema-grafema e conhecimento fonológico explícito. Os resultados evidenciam limitações significativas no conhecimento metafonológico, nomeadamente na identificação de acento de palavra, sílabas, sons e rimas, bem como na explicação de fenómenos fonológicos, observando-se uma influência da ortografia nas respostas dos participantes e um desempenho ligeiramente superior entre os estudantes de mestrado. Estes resultados revelam a necessidade urgente de reforçar o ensino explícito da Fonologia na formação de professores.

Palavras-chave: consciência fonológica, formação inicial de professores, fonologia, português europeu.

1. Introdução

A consciência fonológica tem sido amplamente reconhecida como uma competência central para o sucesso na aprendizagem da leitura e da escrita, sobretudo em línguas com sistemas de escrita alfabéticos, como o português, que apresenta uma correspondência, nem sempre biunívoca, entre grafemas e fonemas (Freitas et al., 2007; Hulme et al., 2005). Esta competência implica a capacidade de refletir sobre a estrutura sonora da língua, incluindo a identificação e manipulação de unidades fonológicas, como sílabas e fonemas, sendo a sua importância, em particular a que se refere à consciência segmental (ou fonémica), especialmente crítica nos

primeiros anos de escolaridade, quando os alunos estão a iniciar o processo de aquisição do código escrito (Liberman et al., 1974; Morais et al., 1979).

Apesar do consenso científico sobre a relevância da consciência fonológica, a sua promoção sistemática nas salas de aula continua a enfrentar dificuldades. Um dos principais obstáculos relaciona-se com a formação inicial e contínua dos professores, que nem sempre inclui de forma suficiente o estudo da consciência fonológica e de outros processos relacionados com a leitura e a escrita, e a sua relação com a aprendizagem do modo escrito (Giménez et al., 2023; Jaskolski & Moyle, 2023). Também em Portugal, estudos anteriores têm revelado lacunas significativas nos conhecimentos e práticas dos educadores e professores relativamente à promoção da consciência fonológica (Amorim & Castelo, 2025; Leite et al., 2022; Lopes et al., 2014). A ausência de preparação específica nesta área compromete, assim, a capacidade dos (futuros) professores para identificarem dificuldades nos seus alunos na iniciação à leitura e à escrita e para implementarem estratégias pedagógicas eficazes.

O presente estudo parte deste cenário para investigar o conhecimento metafonológico de estudantes do primeiro ano de licenciatura em Educação Básica e de mestrados que conferem habilitação profissional para a docência. Especificamente, pretende-se descrever o desempenho dos participantes em diferentes tarefas de consciência fonológica. Este objetivo central sustenta as perguntas de investigação formuladas: (i) que tarefas de consciência fonológica colocam maiores dificuldades aos participantes? e (ii) que padrões de resposta e justificações surgem com maior frequência nas tarefas propostas? A partir da análise desses resultados, procura-se identificar áreas que devem ser reforçadas na formação de professores para uma promoção mais eficaz da consciência fonológica em sala de aula.

Na secção seguinte, será enquadrado o conceito de consciência fonológica e discutida a sua relevância na aquisição da leitura e da escrita, bem como na formação de professores. A secção 3 descreve a metodologia adotada no estudo, apresentando-se, em 4, os principais resultados obtidos, cuja discussão se faz na secção 5. O artigo encerra com as conclusões e implicações para a formação docente, na secção 6.

2. A consciência fonológica na iniciação à leitura e à escrita

A consciência fonológica é um conceito fundamental na investigação sobre aquisição fonológica e sobre os processos de aprendizagem da leitura e da escrita em línguas alfabéticas, inserindo-se no domínio mais amplo da consciência linguística. De acordo com Duarte (2008), a consciência linguística é a capacidade de refletir, sistematizar e distanciar-se do uso espontâneo da língua. A consciência fonológica, por sua vez, é a capacidade de refletir sobre a dimensão sonora da língua, identificando e manipulando unidades fonológicas, como palavras, sílabas (consciência silábica), elementos dentro da sílaba (consciência intrassilábica) e fonemas ou segmentos (consciência fonémica ou segmental) (Freitas et al., 2007).

A consciência segmental – definida como a capacidade de segmentar e manipular os segmentos da língua – destaca-se por apresentar uma inter-relação muito forte com a aprendizagem da leitura e da escrita. Por um lado, esta competência tem-se revelado um forte preditor de sucesso na aprendizagem da leitura e da escrita, em particular em sistemas alfabéticos (Hulme et al., 2005; Leite, 2022); por outro, o seu desenvolvimento e consolidação dependem necessariamente da aprendizagem da leitura e da escrita (Alegria & Morais, 1979; Liberman et al., 1974; Morais, 2009). Com efeito, estudos clássicos conduzidos por Morais e colaboradores demonstraram como a perceção e a manipulação de segmentos não emergem de forma robusta sem a mediação do código escrito, o que reforça a necessidade de instrução explícita destas competências em contextos educativos (Morais et al., 1979). Esta inter-relação entre conhecimento fonológico e código escrito tem sido também destacada em diversos outros estudos no âmbito do português, que demonstram tanto a influência do conhecimento ortográfico sobre o plano fonológico como a possibilidade de os desvios ortográficos dos aprendentes refletirem as suas representações fonológicas (Costa et al., 2021; Miranda & Matzenauer, 2010; Rodrigues & Lourenço-Gomes, 2016, 2018; Santos, 2013; Santos et al., 2014; Veloso, 2003).

Para além das relações entre fonemas e grafemas, também as propriedades internas aos segmentos são relevantes para a consciência segmental e a sua relação com a decifração. Por exemplo, Alves (2013) demonstrou que o modo de articulação exerce um papel facilitador na segmentação fonémica em português europeu, situando fricativas e oclusivas nos extremos de uma escala de desempenho. De acordo com a autora, as propriedades das fricativas – em particular, o traço [+contínuo] – proporcionam representações acústicas mais robustas e consistentes do que as oclusivas, favorecendo a sua consciencialização e facilitando a sua identificação como unidade isolada. Esta conclusão coloca em causa a ideia tradicional de que as oclusivas deverão ser as primeiras consoantes a trabalhar em sala de aula.

Sendo a aprendizagem da leitura e da escrita fundamental para o desenvolvimento da consciência segmental, é fundamental conhecer as especificidades do sistema de escrita vigente. No caso do português, o sistema ortográfico, apesar de ser alfabético, é caracterizado por um grau intermédio de opacidade, pois a correspondência entre fonemas e grafemas nem sempre é direta, apresentando diversas irregularidades e variações contextuais (Veloso, 2005, 2022). Entre os aspetos que conferem opacidade, o autor refere a representação das vogais /a, ɛ, e, ɔ, o/ em posição átona, que foneticamente são elevadas e recuadas, reduzindo-se a [ɐ, ɨ, u]; a grafia de marcas etimológicas, como o <h>; casos de relações multívocas grafema-fonema, tanto foneticamente motivadas (e.g., <c> antes de <a, o, u> vs. <e, i>) como imotivadas (e.g., [s] antes de <e, i> grafado com <s> ou <c>) e verdadeiras irregularidades, como os diversos valores fonéticos imprevisíveis do grafema <x>. Esta complexidade do sistema ortográfico exige um domínio sólido da estrutura fonológica da língua, bem como estratégias pedagógicas que assegurem uma abordagem explícita e sistemática da relação entre som e grafia (Castelo, 2012; Freitas et al., 2007; Veloso, 2005, 2022). Assim, a aprendizagem da leitura e da escrita beneficia, desde os primeiros anos de escolaridade, de práticas que incluam atividades de segmentação fonémica, discriminação auditiva e associação fonema-grafema, entre outras.

Para que essas práticas sejam eficazes, é imprescindível que os professores possuam um conhecimento fonológico robusto. Esse conhecimento não se limita ao uso intuitivo da língua (conhecimento implícito), mas envolve também a capacidade de refletir sobre o funcionamento do sistema linguístico, identificando e nomeando unidades linguísticas, caracterizando-as e explicando como se combinam (Duarte, 2008). Quando este conhecimento metalinguístico se aplica ao nível fonológico, surge a competência metafonológica, que permite aos professores monitorizar e intervir conscientemente nos processos de aprendizagem da leitura e da escrita, reconhecendo e interpretando desvios em produções dos alunos, para planear e aplicar práticas pedagógicas fundamentadas e ajustadas (Carroll et al., 2012; Leite, 2022; Leite et al., 2022; Lopes, 2022; Veloso, 2022). Tal conhecimento não surge espontaneamente, sendo adquirido por via de formação específica e contínua. No entanto, investigação recente tem evidenciado falhas significativas nesta área: muitos professores não dominam este tipo de conhecimento, o que compromete a implementação de práticas pedagógicas baseadas nesta evidência científica (Amorim & Castelo, 2025; Leite et al., 2022). Esta lacuna é referida também na formação de professores de outros países, como Estados Unidos da América e Brasil (Treiman, 2022; Vale, 2022). Este défice de formação sobre a estrutura fonológica da língua é frequentemente atribuído à ausência de uma abordagem sistemática desses conteúdos nos currículos das licenciaturas e mestrados em educação. Como referem Leite e colegas (2022), os programas de formação inicial tendem a negligenciar a dimensão fonológica da linguagem, apesar de os documentos curriculares oficiais, como as *Orientações Curriculares para a Educação Pré-Escolar* (Silva et al., 2016) e as *Aprendizagens Essenciais* (Direção-Geral da Educação, 2018), salientarem a importância da consciência fonológica no percurso escolar.

Assim, compreender o estado atual dos conhecimentos metafonológicos entre futuros professores é essencial para repensar a formação inicial e para garantir que os educadores estejam devidamente preparados para promover aprendizagens linguísticas eficazes junto dos seus alunos.

3. Metodologia

O presente estudo segue uma abordagem de investigação descritiva com componentes quantitativas e qualitativas. Participaram na investigação 98 estudantes, 64 do 1.º ano da licenciatura em Educação Básica e 34 do 1.º ano de cursos de mestrado em Educação Pré-Escolar e Ensino do 1.º Ciclo co Ensino Básico (CEB) ou em Ensino do 1.º e 2.º CEB. Todos eram falantes nativos de dialetos setentrionais do português europeu, sendo a maioria (93,9%) do género feminino.

Em termos de formação prévia, 58% dos estudantes indicaram ter estudado fonética ou fonologia durante o ensino básico ou secundário. Entre os estudantes do mestrado, 73,5% afirmaram ter abordado estes conteúdos durante a licenciatura. Essa informação permite analisar potenciais diferenças de desempenho associadas ao grau de experiência e ao contacto anterior com a área da fonologia.

As respostas foram recolhidas através de um questionário *online*, aplicado no início do ano letivo, na primeira aula de uma unidade curricular comum a todos os cursos (cf. questionário apresentado no Anexo 1).

O questionário continha 35 questões, oito das quais se destinavam à caracterização dos participantes. As restantes 27 questões foram organizadas em categorias, incidindo sobre tarefas de: (i) consciência silábica e intrassilábica; (ii) consciência segmental; (iii) relação fonema-grafema e (iv) conhecimento fonológico explícito. Para cada tipo, procurou-se diversificar os itens de acordo com variáveis linguísticas como estrutura silábica, transparência ortográfica, presença de ditongos nasais com ou sem til, distinção entre vogal e glide, e posição da sílaba tónica. Algumas questões implicavam apenas a identificação de fenómenos fonológicos, enquanto outras exigiam uma justificação ou explicação, permitindo observar a capacidade de reflexão metalinguística dos participantes. A Tabela 1 sintetiza as variáveis manipuladas por tarefa / bloco de tarefas.

Tabela 1

Síntese das Tarefas de Conhecimento Metafonológico Presentes no Questionário

Tipo de conhecimento metafonológico	Tarefas	Propriedade(s) linguística(s) manipulada(s)	Questões exemplificativas
Consciência silábica	Divisão silábica	Encontros vocálicos; transparência/opacidade ortográfica	Quantas sílabas tem a palavra *esmiuçar*?
Consciência do acento de palavra	Identificação de sílaba tónica	Posição acentual; vogal nasal; transparência/opacidade ortográfica	Qual é a sílaba tónica da palavra *bênção*?
Consciência da rima	Identificação de rimas	Constituintes silábicos transparência/opacidade ortográfica	Indique a palavra que rima com *órgão*.
Consciência segmental	Identificação de número de sons	Transparência/opacidade ortográfica	Quantos sons tem a palavra *sangue*? Considerando a pronúncia do português europeu padrão, assinale a(s) palavra(s) que conté(ê)m mais letras do que sons.
Relação fonema-grafema	Correspondência fonema-grafema	Transparência/opacidade ortográfica	Considerando a pronúncia do português europeu padrão, que som representa a letra a negrito na palavra *mu**s**go?*
Conhecimento fonológico explícito	Distinção vogal/glide, vozeado/não vozeado, vogal oral/nasal	Vozeamento; nasalidade; vogal/glide; transparência/opacidade ortográfica	Qual das seguintes palavras termina na sequência consoante + vogal + semivogal? Considerando a pronúncia do português europeu padrão, que sons mudam na primeira sílaba das palavras *vaca* e *faca*?

Os dados recolhidos foram tratados estatisticamente para identificar níveis de desempenho e padrões de acerto e desvio em cada tipo de tarefa. Paralelamente, as respostas abertas e justificações fornecidas pelos participantes foram analisadas qualitativamente, permitindo interpretar não apenas os erros, mas também os raciocínios subjacentes, nomeadamente no que se refere à influência da ortografia e ao tipo de conhecimento mobilizado (implícito ou explícito).

Esta abordagem mista permitiu uma análise mais rica e contextualizada dos dados, articulando os resultados quantitativos com as inferências qualitativas extraídas das justificações dos estudantes.

4. Resultados

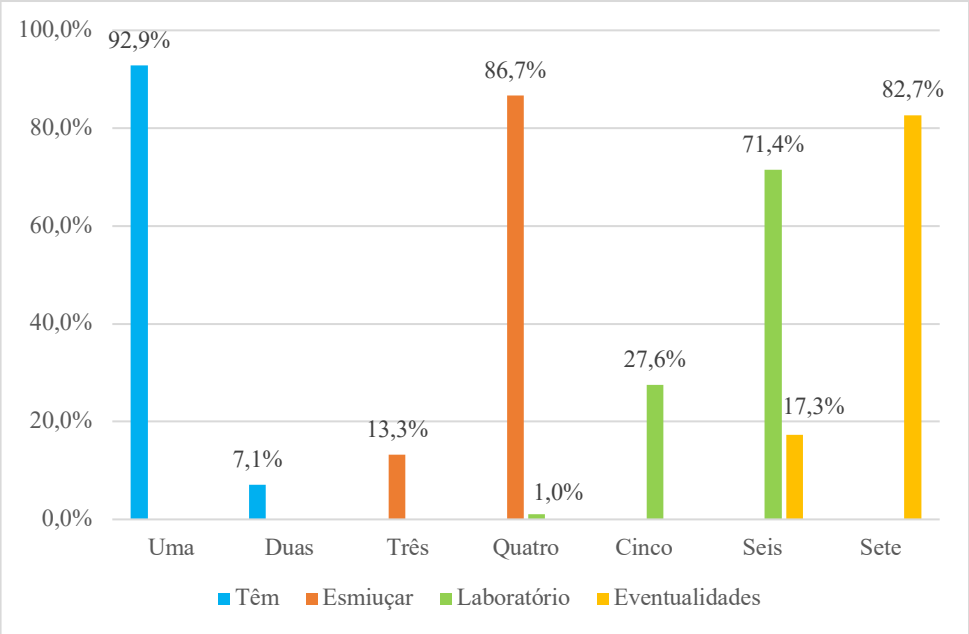
Nesta secção, apresentam-se os resultados obtidos pelos dois grupos de estudantes nas diferentes tarefas do questionário. De modo a facilitar a leitura, a análise segue a matriz apresentada na Metodologia (cf. Tabela 1).

4.1. Consciência silábica

A Figura 1 inclui a percentagem de cada resposta dada pelos estudantes (dos dois grupos, em conjunto) numa tarefa de contagem de sílabas de quatro palavras diferentes.

Figura 1

Respostas na Contagem de Sílabas, nos Dois Grupos de studantes



Apesar de apresentar duas sílabas, quase todos os estudantes atribuem uma única sílaba à palavra *têm* (92,9% das respostas). Quanto a *esmiuçar* e a *eventualidades*, mais de 80% dos respondentes consideram que as palavras apresentam quatro (86,7% das respostas) e sete sílabas (82,7%), respetivamente. No caso da palavra *laboratório*, também se verifica um padrão de resposta predominante (71,4% das respostas identificam seis sílabas), mas não tão acentuado como os padrões para as restantes palavras.

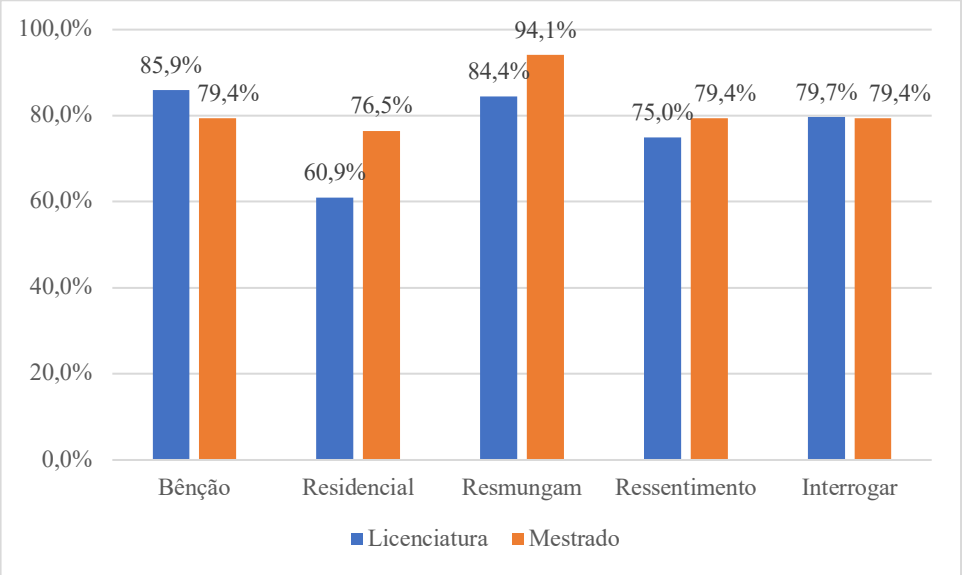
Analisadas as respostas por grupo de estudantes, verifica-se que as diferenças não são significativas, registando-se as maiores diferenças na identificação do número de sílabas das palavras *têm* e *laboratório*. Relativamente à primeira, a percentagem de estudantes de mestrado que considera que a palavra possui apenas uma sílaba (97,1%) é superior à de estudantes de licenciatura (90,6%). Relativamente a *laboratório*, há mais estudantes de mestrado que consideram que a palavra tem cinco sílabas (35,3% vs. 23,4%).

4.2. **Consciência do acento de palavra**

Na Figura 2, regista-se a percentagem de respostas corretas na identificação da sílaba tónica de cinco palavras, distinguindo os resultados em função do grupo de estudantes.

Figura 2

Respostas Corretas na Identificação da Sílaba Tónica, por Grupo de Estudantes



Comparando os resultados dos dois grupos de estudantes, verifica-se que, também aqui, as diferenças são reduzidas: os alunos de mestrado mostram um desempenho ligeiramente superior em três palavras (*residencial, resmungam, ressentimento*) e inferior numa outra (*bênção*).

As percentagens de respostas corretas situam-se entre os 60,9% e os 94,1%, ocorrendo maioritariamente por volta dos 70% e dos 80%. A palavra com melhores resultados é *resmungam* (84,4%–94,1%), a única em que se registam resultados próximos de efeitos de teto, enquanto *residencial* apresenta a menor taxa de sucesso (60,9%–76,5%).

A Tabela 2 apresenta o número de respostas para cada sílaba sobre o total de repostas incorretas por palavra e por grupo de estudantes.

Tabela 2

Respostas Incorretas na Identificação da Sílaba Tónica, por Grupo de Estudantes

	Sílaba	Licenciatura	Mestrado
Residencial	*si*	3/25	1/8
	den	15/25	6/8
	ci	7/25	1/8
Resmungam	*res*	3/10	2/2
	gam	7/10	0
Ressentimento	*ssen*	5/16	3/7
	to	5/16	3/7
Interrogar	*in*	8/13	3/7
	te	2/13	1/7
	rro	3/13	3/7

Verifica-se que, em cada palavra, a sílaba incorreta mais escolhida é, *grosso modo*, a mesma nos dois grupos. Na palavra que registou mais respostas incorretas (*residencial*), a sílaba <den> foi selecionada mais frequentemente (60,0% dos 25 estudantes de licenciatura com respostas incorretas para esta palavra vs. 75,0% dos 8 estudantes de mestrado). As restantes sílabas foram indicadas por apenas um número residual de estudantes de mestrado, registando-se, porém, 7 respostas para a sílaba <ci> entre as respostas erradas dos estudantes de licenciatura (28,0%). Na palavra *resmungam*, apenas dois estudantes de mestrado identificaram incorretamente a sílaba tónica e ambos selecionaram a sílaba <res>, enquanto a maioria dos estudantes de licenciatura que completaram a tarefa sem sucesso (7/10) optou pela sílaba <gam>. Já na palavra *interrogar*, 62,5% das 13 respostas erradas entre os estudantes de licenciatura concentraram-se na sílaba <in>, enquanto a maioria de respostas erradas entre o outro grupo de estudantes está dividida entre essa sílaba (3/7) e <rro> (3/7).

4.3. **Consciência da rima**

Na Tabela 3 encontram-se as percentagens de respostas dadas numa tarefa de identificação de rimas para as palavras *mãe* e *órgão* (consciência da rima), em função do grupo de estudantes. Note-se que foram consideradas corretas apenas as rimas soantes, isto é, aquelas em que a correspondência dos sons é completa a partir da vogal tónica.

Tabela 3

Respostas na Identificação de Palavras que Rimam com Mãe e Órgão, por Grupo de Estudantes

		Licenciatura	Mestrado
Mãe	*Porém*	28,1%	26,5%
	Cem	29,7%	35,3%
	Pães	65,6%	52,9%
	Sai	4,7%	8,8%
	Não responde	1,6%	2,9%
Órgão	*Outorgam*	18,8%	35,3%
	Órfão	54,7%	58,8%
	Berbigão	7,8%	2,9%
	Orégão	18,8%	2,9%

Nota. As respostas corretas são registadas a azul. A soma das percentagens para cada palavra é superior a 100% porque os participantes podiam selecionar mais do que uma resposta.

Mais uma vez se registam tendências de resposta bastante semelhantes nos dois grupos, com os estudantes de mestrado a apresentar resultados ligeiramente superiores. Verifica-se apenas uma diferença maior no caso de duas respostas para a rima de *órgão*: a percentagem de respostas (corretas) na identificação de *outorgam* é superior entre os alunos de mestrado (35,3% vs. 18,8% entre os alunos de licenciatura) e a de respostas que selecionam (incorretamente) *orégão* é inferior nos mesmos alunos (2,9% vs. 18,8% entre os alunos de licenciatura).

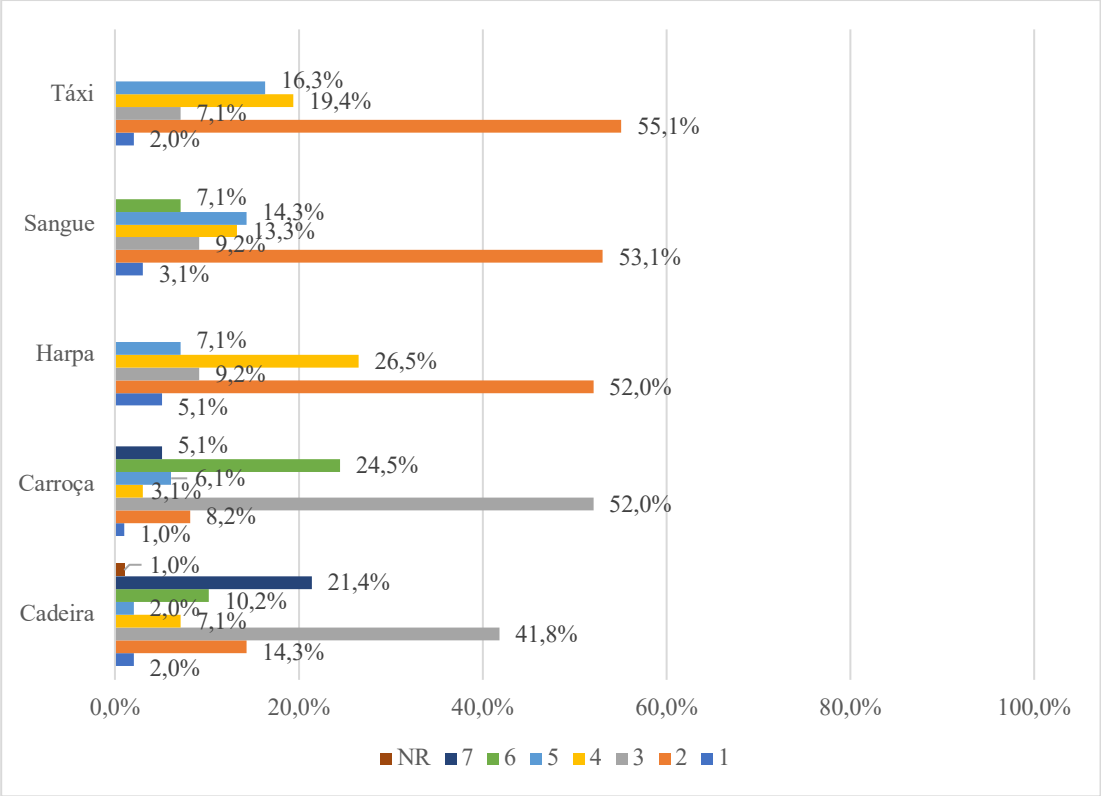
Observando as respostas no conjunto, constata-se que a percentagem de identificações corretas (*porém, cem*) se situa entre os 26,5% e 35,3% para a palavra *mãe*, enquanto a de identificação incorreta de *pães* é bastante superior, entre os 52,9% e os 65,6%. Também no caso de *órgão*, a proporção de respostas corretas é bastante inferior (*outorgam*: 18,8%–35,3%) à de respostas incorretas (nomeadamente, *órfão*: 54,7%–58,8%).

4.4. Consciência segmental

A Figura 3 inclui as respostas na contagem de sons, dadas pelo conjunto dos dois grupos de participantes, para cada uma de cinco palavras.

Figura 3

Respostas na Contagem de Sons, nos Dois Grupos de Estudantes



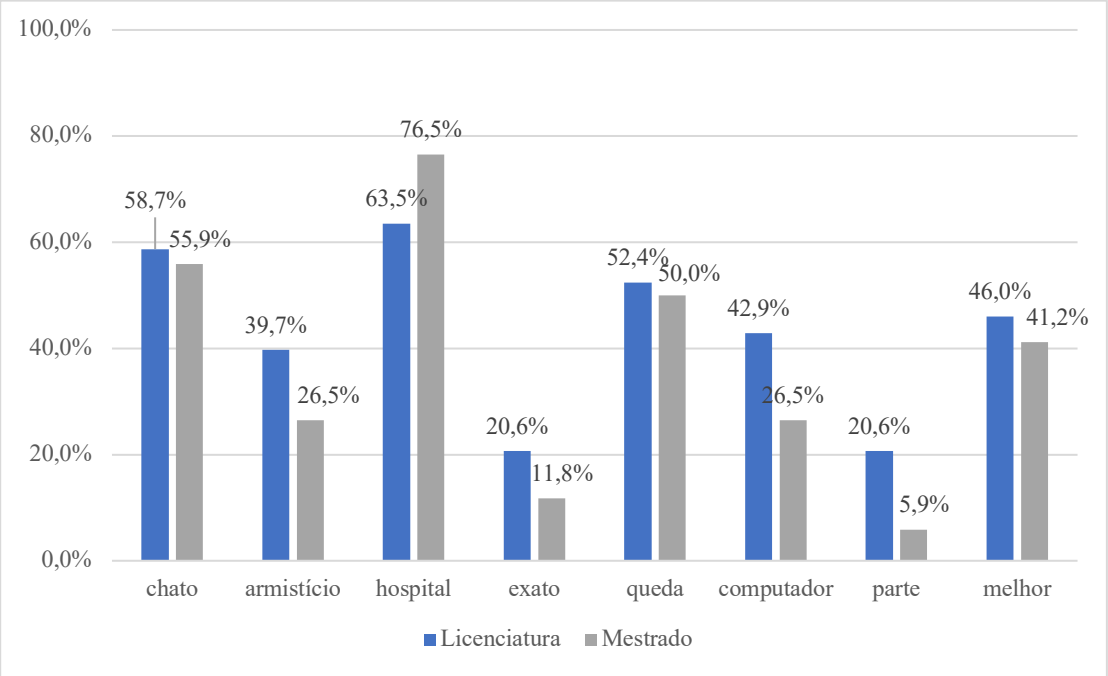
Para a questão "Quantos sons tem a palavra *X*?", que se segue a perguntas como "Quantas sílabas tem a palavra *X*?", registam-se bastantes respostas diferentes, mas a maioria indica um número de sons muito inferior ao correto. Por exemplo, para as palavras *cadeira* e *carroça*, a resposta mais frequente é três sons (*cadeira*:

41,8%; *carroça*: 52,0%) e, para as restantes, é dois sons (*harpa*: 52,0%; *sangue*: 53,1%; *táxi*: 55,1%). O número de respostas corretas é muito inferior (7 sons em *cadeira*: 21,4%; 6 sons em *carroça*: 24,5%; 4 sons em *harpa*: 26,5%; 4 sons em *sangue*: 13,3%; 5 sons em *táxi*: 16,3%).

A Figura 4 mostra a percentagem de respostas na identificação de palavras com mais letras do que sons, em função do grupo de estudantes.

Figura 4

Respostas na Identificação de Palavras com Mais Letras do que Sons, por Grupo de Estudantes



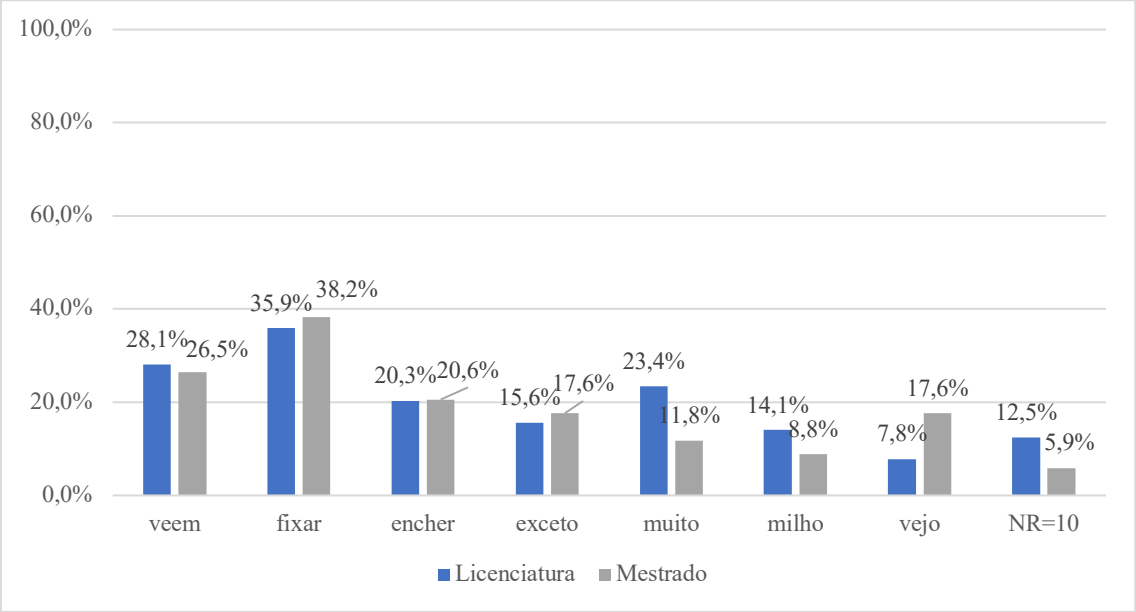
As palavras que devem ser selecionadas, por terem um maior número de letras do que de sons, são *chato*, *hospital*, *queda*, *computador* e *melhor*. A que obtém uma maior percentagem de respostas corretas é *hospital* (63,5%–76,5% de identificações), seguindo-se *chato* (55,9%–58,7%), *queda* (50,0%–52,4%), *melhor* (41,2%–46,0%), e registando-se o menor número de identificações corretas em *computador* (26,5%–42,9%). Quanto às palavras que não devem ser selecionadas (por terem o mesmo número de letras e de sons), ainda se constatam várias identificações incorretas, sobretudo para *armistício* (26,5%–39,7%).

São de destacar, mais uma vez, as mesmas tendências de resposta nos dois grupos de participantes, registando-se diferenças superiores a 10% apenas em algumas palavras: *armistício*, *parte* e *hospital*, com melhores resultados nos estudantes de mestrado, e *computador*, com melhor nível de desempenho nos estudantes de licenciatura.

Na Figura 5, regista-se a percentagem de respostas na identificação de palavras com mais sons do que letras, de acordo com o grupo de participantes.

Figura 5

Respostas na Identificação de Palavras com Mais Sons do que Letras, por Grupo de Estudantes



Deveriam ser selecionadas as palavras que têm mais sons do que letras: obrigatoriamente *fixar* e eventualmente *exceto* e *vejo* (que podem ser produzidas como [ɐjʃˈsɛtu] e [ˈvɐjʒu], respetivamente). De facto, a palavra *fixar* apresenta o maior número de identificações, embora estas ainda sejam relativamente baixas (35,9%–38,2%). O número de identificações para as palavras *exceto* e *vejo* é bastante inferior (15,6%–17,6% e 7,8%–17,6%, respetivamente), o que pode ser relacionado com o facto de as palavras apresentarem diferentes produções possíveis (com e sem maior número de sons do que de letras). Quanto às restantes palavras, o número de identificações incorretas, assinalando-as como tendo maior número de sons do que letras, oscila entre 8,8%–14,1% para *milho* e 26,5%-28,1% para *veem*.

Comparando os resultados dos dois grupos de participantes, verifica-se que as tendências de resposta são novamente semelhantes, embora sejam um pouco melhores entre os alunos de mestrado: mais identificações corretas (cf. *fixar*, *exceto* e *vejo*) e menos identificações incorretas (cf. *muito* e *milho*).

4.5. **Correspondência fonema-grafema**

A Tabela 4 mostra as percentagens de respostas mais frequentes numa tarefa de identificação de sons (segmentos específicos) representados por diferentes grafemas em diversas palavras (*e.g.* Considerando a pronúncia do português europeu padrão, que som representa a letra a negrito na palavra *mu***s**go?), em função do grupo de estudantes.

Tabela 4

Respostas mais Frequentes na Identificação de Sons (Segmentos Específicos), por Grupo de Estudantes

		Licenciatura	Mestrado
Leite	[ɐ] de *uma*	11,1%	5,9%
	[e] de *vê*	52,4%	64,7%
	[ɛ] de p**é**	20,6%	14,7%
Paz	[ʃ] de *ch**á***	42,9%	52,9%
	[z] de ***Z**é*	25,4%	8,8%
Pasta	[ʃ] de *ch**á***	69,4%	78,1%
	[s] de ***s**ol*	14,5%	9,4%
Musgo	[ʒ] de *j**á***	24,0%	41,2%
	[ʃ] de *ch**á***	24,0%	38,2%
	[s] de ***s**ol*	15,9%	17,6%

Nota. Os segmentos específicos e as respostas corretas são registados a azul.

Comparando os resultados dos dois grupos de participantes na tarefa sobre a correspondência fonema-grafema em causa, que solicita que se considere a pronúncia do PE padrão, constatam-se novamente as mesmas tendências nas respostas de alunos dos dois níveis académicos e percentagens de respostas corretas quase sempre superiores nos estudantes de mestrado (e.g. 52,9% de identificações corretas de [ʃ] em *paz* e 41,2% de [ʒ] em *musgo* vs. 42,9% e 24,0%, respetivamente, de identificações corretas entre os alunos de licenciatura).

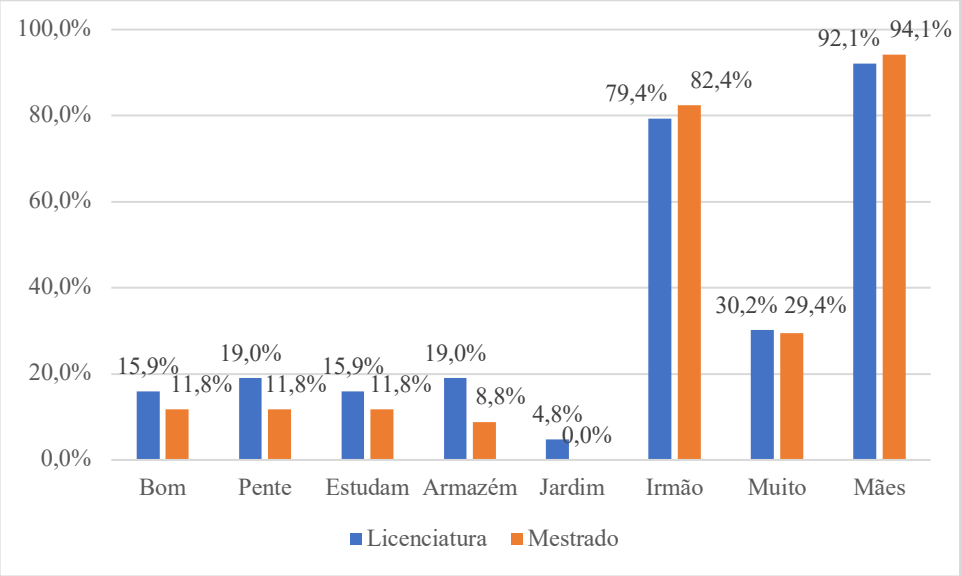
Analisando o conjunto de respostas, é de salientar que o segmento mais frequentemente identificado é o [ʃ] de *pasta* (69,4–78,1%), seguido pelo [ʃ] em *paz* (42,9%-–52,9%). O grau de dificuldade na identificação segmental é muito mais elevado para o [ʒ] em *musgo* (24,0%–41,2%) e, sobretudo, para a vogal [ɐ] em *leite* (5,9%–11,1%). Neste último caso, a maioria das respostas consiste em [e] (52,4%–64,7%), correspondendo à produção típica da vogal na zona dialetal dos participantes. Não obstante, poderá ter havido também uma associação do símbolo fonético ([e]) ao grafema (<e>).

4.6. **Conhecimento fonológico explícito**

Na Figura 6, regista-se a percentagem de escolha de diversas palavras como tendo ditongos nasais, considerando o conjunto de respostas dos dois grupos de participantes.

Figura 6

Respostas na Identificação de Ditongos Nasais, por Grupo de Estudantes

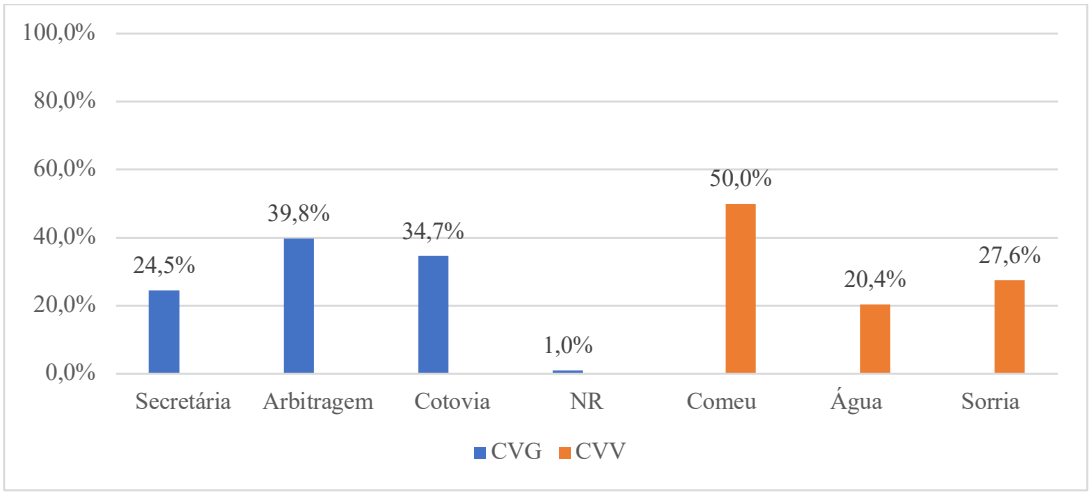


Registam-se efeitos de teto apenas para a palavra *mães*, com um total de 91,8% de respostas que identificam esta palavra como tendo um ditongo nasal (92,1% dos estudantes de licenciatura e 94,1% dos de mestrado). A percentagem de respostas corretas é ainda elevada para *irmão* (79,4% vs. 82,4%) e bastante mais baixa para *muito* (30,2% vs. 29,4), *armazém* (19,0% vs. 8,8%) e *estudam* (15,9% vs. 11,8%). As palavras que contêm apenas uma vogal nasal são erradamente selecionadas por alguns participantes: *pente* (19,0% vs. 11,8%); *bom* (15,9% vs. 11,8%); *jardim* (4,8% vs. 0%). De uma forma geral, observa-se um desempenho ligeiramente superior nos estudantes de mestrado em todos os itens, à exceção das palavras *estudam* e *armazém*.

A Figura 7 apresenta a percentagem de respostas, dadas pelo conjunto dos dois grupos de participantes, na identificação da palavra que terminava na sequência CVG (e.g., *arbitragem*) e na que terminava na sequência CVV (e.g., *sorria*).

Figura 7

Respostas na Identificação de Palavras Terminadas nas Sequências CVG e CVV, nos Dois Grupos de Estudantes



Os resultados nesta tarefa mostram o desempenho dos participantes (em conjunto) em termos de identificação de encontros vocálicos, nomeadamente, a distinção entre hiatos e diferentes tipos de ditongos. Devendo selecionar a palavra com o formato *consoante* + *vogal* + *semivogal* (CVG), a resposta mais frequente é a escolha correta de *arbitra**gem*** (39,8%), mas muitos participantes optam por *coto**via*** (34,7%) e outros por *secretá**ria*** (24,5%). Para o formato *consoante* + *vogal* + *vogal* (CVV), a resposta correta ***sorria*** é selecionada em apenas 27,6% dos casos, predominando a escolha incorreta de ***comeu***, com ditongo decrescente (50,0%), e registando-se ainda várias escolhas de ***água***, com ditongo crescente (20,4%).

Avançando agora para outra tarefa de conhecimento explícito, na Tabela 5 encontram-se exemplos de respostas, dadas por membros de um dos dois grupos de participantes, às seguintes questões: "Considerando a pronúncia do português europeu padrão, que sons mudam na primeira sílaba das palavras ***vaca*** e ***faca***? Porque é que esses sons são diferentes?".

Tabela 5

Exemplos de Respostas às Perguntas sobre os Sons que Mudam na 1.ª Sílaba de vaca / faca, nos Dois Grupos de Estudantes

Código do participante	Resposta
03	Os sons são diferentes, porque a letra inicial é diferente.
10	A palavra vaca inicia-se com a letra v, já a letra faca inicia-se com a letra f.
13	VA e FA produzem sons diferentes por se tratarem de consoantes diferentes.
17	Os sons são diferentes pois não começam pela mesma letra.
65	/v/ e /f/. O que muda nas palavras vaca e faca são as consoantes com que se inicia a palavra, sendo que uma das consoantes é fricativa (f) e a outra é oral (v).
66	vaca tem um som mais aberto, V e faca têm um som mais fechado, F.
77	O som [v] e o som [f].

Quanto à primeira questão, apenas duas respostas identificam claramente os sons que mudam (cf. respostas de 65 e 77), enquanto as restantes parecem confundir os sons com as letras (e.g. resposta de 66) e/ou as sílabas (e.g. resposta de 13).

Em termos de justificação das diferenças entre os sons, a maioria das respostas fundamenta-as no uso de letras diversas (e.g. respostas de 03, 10, 17). Uma das respostas não tenta justificar a diferença (cf. resposta de 77) e duas revelam uma tentativa de explicitação de propriedades orais, que não é, contudo, bem-sucedida (cf. respostas de 65 e 66). Os resultados nesta tarefa são semelhantes nos dois grupos de respondentes.

Para terminar a apresentação dos resultados, a Tabela 6 integra exemplos de respostas dadas, por elementos de algum dos dois grupos de estudantes, para explicar os sons que mudam na primeira sílaba das palavras *fita* vs. *finta*.

Tabela 6

Exemplos de Respostas às Perguntas sobre os Sons que Mudam na 1.ª Sílaba de fita / finta, nos Dois Grupos de Estudantes

Código do participante	Resposta
07	A adição do som "n" em finta altera o significado da palavra.
08	Os sons são diferentes devido ao "n" na palavra finta que produz um som nasal na sua pronúncia.
09	O som muda na primeira sílaba, pois na palavra "finta", é acrescentado um "n", fazendo as palavras soar de forma diferente.
10	Estas palavras têm o som diferente na medida em que na primeira sílaba tem o som n
44	os sons que mudam são "fi" e "fin" e são diferentes porque "fin" termina com um "n".
45	Porque "finta" tem um fonema adicional que "fita" não tem.
70	Os sons i e in.

Relativamente à identificação dos sons que mudam na primeira sílaba, apenas uma resposta pode ser considerada correta mesmo não recorrendo às convenções gráficas próprias da representação de sons (cf. resposta 70). Mais uma vez, várias respostas recorrem às letras (e.g. resposta de 09) e/ou às sílabas (e.g. resposta de 44).

Quanto à justificação da diferença, são de destacar uma resposta parcialmente correta, com a referência à nasalidade (cf. resposta 08), uma resposta que aborda propriedades orais das palavras (cf. resposta 09) e uma resposta que menciona informações lexicais (cf. resposta 07). Mais uma vez se destacam tendências semelhantes nas respostas de estudantes de licenciatura e de mestrado.

5. Discussão dos resultados

A discussão dos resultados organiza-se em torno das duas perguntas de investigação que nortearam este estudo: (i) que tarefas de consciência fonológica colocam maiores dificuldades aos participantes? e (ii) que padrões de resposta e justificações surgem com maior frequência nas tarefas propostas?

Relativamente à primeira questão, os resultados obtidos revelaram melhor desempenho nas tarefas de identificação de rimas e nas que envolvem a unidade sílaba. Estas tarefas estão associadas à consciência fonológica holística (Castelo, 2012), que emerge naturalmente por volta dos três anos e se vai desenvolvendo ao longo da idade pré-escolar e escolar (Alves, 2013). Este nível de consciência fonológica manifesta-se na capacidade de prestar atenção e manipular "as propriedades mais salientes e gerais da língua" (Castelo, 2023, p. 3), sem que se faça uma análise propriamente dita dessas propriedades. Apesar de serem tarefas mais

acessíveis, não se verificou um efeito de teto em nenhuma, dado que surgiram dificuldades motivadas por interferências da ortografia. Por exemplo, na tarefa de contagem de sílabas da palavra *têm*, a maioria dos participantes não contabilizou corretamente as duas sílabas, pelo facto de a ortografia da palavra não refletir claramente a sua estrutura fonológica, levando a que fosse interpretada como monossilábica (Figura 1). De igual modo, na identificação de rimas, mais de metade dos participantes selecionou a opção gráfica *pães* como rimando com *mãe*, embora a correspondência fonológica não fosse válida (Tabela 2). Nestes e noutros casos, observa-se que as dificuldades surgem frequentemente em palavras cuja ortografia não revela inequivocamente a estrutura sonora, o que evidencia o papel da interferência ortográfica.

Registou-se maior dificuldade nas tarefas associadas à consciência fonológica analítica (cf. Castelo, 2012), que tem de ser estimulada de forma explícita e que se manifesta na capacidade de análise e manipulação de propriedades linguísticas mais específicas e menos salientes, como os segmentos (consciência segmental). Estas dificuldades foram evidentes em tarefas como a que envolveu a contagem de segmentos (Figura 4), na qual a maioria dos participantes indicou um número de sons significativamente inferior ao correto em todas as palavras testadas (e.g., para *cadeira*, resposta modal = 3, resposta correta = 7). Os resultados mostraram também dificuldades nas tarefas que exigem conhecimento fonológico explícito, como a explicação das diferenças entre pares mínimos. Na identificação de ditongos nasais, também se registaram muitas respostas incorretas, sobretudo em palavras como *muito* ou *armazém*: respetivamente, apenas 30,2% e 19,0% dos estudantes identificaram corretamente a presença do ditongo (Figura 3). Por fim, na distinção entre vogal e glide, apenas 39,8% reconheceram corretamente *arbitragem* como palavra com a sequência final CVG e apenas 27,6% identificaram *sorria* como terminando em CVV, prevalecendo escolhas baseadas em padrões gráficos.

Os resultados obtidos coincidem com os de Castelo (2012), que analisou o desempenho de alunos do ensino secundário e do primeiro ano do ensino superior em diferentes cursos. O paralelismo é particularmente preocupante se considerarmos que, no presente estudo, os participantes são estudantes de formação inicial para a docência, o que reforça a urgência de uma intervenção formativa mais sólida e contínua neste domínio.

Passamos agora a analisar os resultados obtidos em função da questão (ii), acerca dos padrões de resposta encontrados e respetivas justificações.

Na consciência de rimas, os participantes selecionaram maioritariamente palavras terminadas na mesma sequência de grafemas, apesar de não rimarem, demonstrando fragilidades acrescidas em palavras com ditongos nasais que não são marcados graficamente com til (e.g. identificação de *outorgam* como palavra que rima com *órgão*). Isto sugere não apenas desconhecimento das convenções ortográficas (neste exemplo, ignora-se o que pode ser um caso de "relações multívocas grafema-fonema contextualmente motivadas", na categorização usada por Veloso (2005) – a sequência *-am* em sílaba final átona representa o ditongo nasal [ɐ̃w̃], grafado com *-ão* quando em sílaba tónica), mas também uma fraca perceção auditiva da semelhança sonora entre as palavras, revelando uma sobreposição do plano ortográfico sobre o fonológico.

Na contagem de sílabas, como referido anteriormente, registou-se maior dificuldade na palavra *têm* (Figura 1). Nas restantes três palavras apresentadas nesta tarefa de consciência silábica, que integram potenciais ditongos crescentes, dois em posição pretónica (*esm**iu**çar* e *event**ua**lidades*) e um em posição pós-tónica (*laboratór**io***), não se verificaram dificuldades, mas apenas uma percentagem reduzida de participantes teve em consideração o ditongo crescente na identificação do número de sílabas das palavras que continham encontros vocálicos, baseando a resposta na produção fonética habitual das palavras. Este resultado confirma que há a ativação de diferentes vias concorrentes no processamento de tarefas de consciência fonológica, como sugerido por Freitas et al. (2015): o acesso à representação fonológica da palavra, por um lado, e o processamento do sinal acústico, por outro. De todas, a palavra *laboratório* é a que apresenta um número superior de respostas com identificação do ditongo, o que pode dever-se ao facto de, ao contrário das outras, a sequência de vogais ocorrer em posição pós-tónica. Efetivamente, esta diferença no contexto acentual já tinha sido relatada por Cabeleira e Correia (2004), num estudo que avaliava a perceção de sequências vocálicas por alunos do 5.º e do 8.º ano. As autoras relatam que as sequências vocálicas são percecionadas como hiatos, por ambos os grupos, quando ocorrem em posição pretónica, ao contrário do que acontece em posição pós-tónica, contexto em que a

maioria dos participantes do 8.º ano perceciona um ditongo crescente. No entanto, a identificação do ditongo nesta palavra pode, também, ter sido condicionada pelas regras ortográficas, em particular, pela necessidade de evitar letras isoladas, na translineação. Em síntese, na resolução da tarefa, os participantes parecem ter recorrido ao conceito de sílaba ortográfica, que é o mais convocado ao longo da escolaridade obrigatória, pela importância que tem na translineação.

Relativamente ao acento de palavra, apesar de a maioria dos participantes identificar corretamente a sílaba tónica em palavras mais simples, observou-se pior desempenho na palavra *residencial* (Figura 2). Entre os participantes que não responderam corretamente, 69,7% indicaram a sílaba <den>. Este desvio poderá explicar-se pela interferência lexical de *residência*, palavra com acento gráfico.

Outra hipótese a considerar é o peso da sílaba com vogal nasal[1]. Com efeito, também na palavra *interrogar*, a sílaba incorreta mais identificada (55% das respostas erradas) foi a que contém a vogal nasal, apesar de se situar fora da "janela das três sílabas". Esta regra do português impede que o acento recue além da antepenúltima sílaba, sendo consideradas agramaticais palavras com acentuação fora dessa janela trissilábica[2] (Pereira, 2020). Um estudo percetivo recente sobre o acento silábico do português revela que a sílaba com vogal nasal (CVN) é, logo depois da sílaba com núcleo ramificado (CVG), aquela que mais atrai o acento: numa tarefa de juízo de aceitabilidade com diferentes tipos de rima complexa, as pseudopalavras proparoxítonas cuja penúltima sílaba era do tipo CVN foram as que registaram a maior taxa de rejeição, logo depois das que continham uma sílaba CVG (Im, 2025). Também num estudo sobre o acento secundário rítmico no PE, Castelo (2004) identificou como as vogais com maior "saliência percetiva" (incluindo as nasais) tendem a atrair os acentos secundários rítmicos, de posição variável.

Num outro trabalho sobre a perceção do acento em português por alunos do 3.º ciclo do ensino básico e do ensino secundário, Araújo (2004) apontou a qualidade vocálica como um fator determinante na identificação do acento. A autora concluiu que a identificação da sílaba tónica pode ser influenciada por vários fatores relacionados com a qualidade das vogais, em particular, a manutenção de traços vocálicos específicos em contexto átono. Com efeito, a proeminência conferida a estas vogais, pelo facto de não alterarem a sua qualidade em contexto átono (por exemplo, mantendo o traço [+bx] ou [+alt]), é um fator que pode levar a identificações incorretas, o que pode explicar os resultados encontrados na palavra *ressentimento*: analisadas as respostas erradas, encontra-se o mesmo número de respostas (7/25) na sílaba <ssen> e na sílaba <ti>. Assim, estas características do conhecimento fonológico implícito dos participantes podem estar a ter um papel relevante no desempenho nas tarefas de consciência fonológica, como se os mesmos mobilizassem o conhecimento fonológico intuitivo de forma incorreta por dificuldade em consciencializá-lo.

As tarefas de consciência segmental, por sua vez, expuseram de forma clara a dificuldade em segmentar palavras em fonemas. Por exemplo, apenas uma minoria dos participantes foi capaz de identificar corretamente os sete fonemas que compõem a palavra *cadeira* (Figura 3). A dificuldade em dissociar grafemas e fonemas foi confirmada por outras tarefas, como a que envolvia o som correspondente ao grafema <e> na palavra *leite* (Tabela 4), demonstrando a persistência de estratégias baseadas na ortografia e não no som, o que dificulta o reconhecimento de unidades fonológicas elementares.

No que respeita ao conhecimento fonológico explícito, as justificações dos participantes basearam-se frequentemente em critérios ortográficos ou lexicais, refletindo uma fraca consciência metalinguística. Este tipo de resposta está em consonância com os resultados de Castelo (2012), que observou fenómenos semelhantes entre alunos do ensino secundário e do 1.º ano de cursos do ensino superior.

Igualmente de destacar são as confusões frequentes entre sílabas e segmentos sonoros. Por exemplo, na tarefa de contagem dos segmentos, muitos participantes deram como resposta um número que corresponde ao número de sílabas e, na tarefa de explicitação das diferenças entre duas palavras, por vezes, os participantes

[1] Em análises fonológicas influentes do PE, considera-se que as vogais nasais ocorrem apenas no nível fonético, correspondendo, fonologicamente, a uma vogal e a um autossegmento nasal (e.g., Mateus & Andrade, 2000).

[2] Excluem-se desta regra as palavras em que a inserção de uma vogal epentética para preencher um núcleo vazio (Mateus & Andrade, 2000) origina uma sílaba fonética, como na palavra *técnica* pronunciada como [ˈtɛkɨnika] (cf. Veloso, 2024).

identificaram as sílabas como elemento em contraste (cf. Tabelas 5 e 6, respostas 13 e 44, respetivamente). Estas confusões revelam um conhecimento ainda incompleto ou instável sobre os níveis de análise linguística.

Assim, relativamente à questão (ii), os resultados obtidos revelaram, de forma clara, a interferência da ortografia nas tarefas de consciência fonológica, tal como documentado na literatura (Castelo, 2012; Carroll et al., 2012; Jaskolski & Moyle, 2023; Veloso, 2003). Esta influência manifestou-se de forma transversal, tanto na consciência silábica como na segmental, na identificação de rimas e no conhecimento fonológico explícito, sendo mais acentuada em palavras cuja ortografia é menos transparente. Estes padrões de resposta não se limitam a dificuldades pontuais, mas antes resultam de diferentes vias de processamento ativadas durante a execução das tarefas — nomeadamente o acesso concorrente à representação fonológica, à informação acústica do estímulo e, frequentemente, às representações ortográficas dos itens lexicais. Tal como sublinhado por Freitas et al. (2015), este fenómeno de ativação concorrente permite explicar tanto a variabilidade dos desempenhos observados como a persistência dos erros motivados por interferência ortográfica, especialmente em contextos em que a relação entre fonologia e grafia é opaca. Desta forma, reforça-se que as tarefas de consciência fonológica requerem frequentemente a mobilização simultânea (ou alternada) destas diferentes fontes de informação linguística, o que deve ser contemplado na interpretação dos resultados e nas propostas de formação para futuros professores.

É particularmente relevante que, apesar de a maioria dos participantes afirmar ter contactado com conteúdos de Fonética ou Fonologia no ensino básico ou secundário, o conhecimento metafonológico revelado permanece muito limitado. Tal como refere Castelo (2012), o simples contacto com os temas não garante a aquisição de competências metalinguísticas, sendo necessário um ensino estruturado, explícito e recorrente, de modo a capacitar os futuros professores para nomear e explicar fenómenos fonológicos.

Embora os estudantes de mestrado tenham apresentado resultados ligeiramente superiores em diversas tarefas, como na identificação da sílaba tónica ou na justificação de diferenças fonológicas, as diferenças entre os dois grupos não são suficientemente expressivas para indicar uma melhoria consistente. Isto poderá refletir não só lacunas na formação prévia como também a ausência de continuidade no aprofundamento de conteúdos fonológicos ao longo da formação no ensino superior.

Estes resultados sublinham a necessidade urgente de reforçar o ensino da componente fonológica da língua não só ao longo da escolaridade obrigatória, mas também nos cursos de formação inicial de professores. Mais do que uma introdução superficial à fonética, importa assegurar que os alunos desenvolvem competências metafonológicas que lhes permitam compreender e manipular conscientemente a linguagem oral, separando claramente os modos oral e escrito da língua.

6. Considerações finais

Com este estudo, procurámos compreender até que ponto estudantes de licenciatura e de mestrado em cursos de formação de professores dominam competências metafonológicas essenciais para o ensino da leitura e da escrita. Através de diferentes tarefas, analisámos o seu desempenho, com o objetivo de identificar fragilidades que possam comprometer a sua futura intervenção pedagógica neste domínio.

Os resultados mostraram maior facilidade em tarefas associadas à consciência fonológica holística, como a contagem silábica e a identificação de rimas. Por contraste, as tarefas que implicam uma consciência fonológica mais analítica — como a segmentação fonémica e a explicitação de contrastes fonológicos — foram marcadas por dificuldades consistentes, que apontam para um domínio ainda incipiente de competências metafonológicas. Estes resultados coincidem com os de Castelo (2012), sendo particularmente relevantes por envolverem estudantes de cursos orientados para o ensino nos anos iniciais.

As respostas dos participantes revelaram dois padrões recorrentes de resposta: a interferência da ortografia e a confusão entre níveis de análise linguística, em especial entre sílabas e segmentos. Estes padrões sugerem uma perceção fonológica condicionada por representações gráficas, fenómeno já documentado noutros estudos (Castelo, 2012; Carroll et al., 2012; Jaskolski & Moyle, 2023; Veloso, 2003) e uma apropriação pouco

diferenciada dos conceitos fonológicos fundamentais. Mesmo quando há intuições corretas, a dificuldade em consciencializá-las e em nomeá-las revela fragilidades ao nível do conhecimento metalinguístico. Este desfasamento entre o saber intuitivo e o conhecimento explícito levanta questões importantes sobre a forma como a fonologia tem vindo a ser (ou a não ser) ensinada. Apesar de os estudantes de mestrado terem revelado um desempenho ligeiramente superior, as diferenças observadas não são suficientemente consistentes para indicar um progresso sustentado e para garantir o domínio de competências metafonológicas.

Mais do que destacar lacunas, os resultados deste estudo sublinham a urgência de reforçar, de forma sistemática, o ensino da fonologia ao longo da formação académica dos futuros professores, não a limitando a um único momento ou unidade curricular. Este reforço deve assumir uma abordagem progressiva e explícita, ancorada na investigação científica e na prática educativa, capaz de transformar o conhecimento implícito em saber operativo (Leite, 2022; Silva, 2022). Apenas dessa forma será possível capacitar os docentes para intervir com segurança e rigor no desenvolvimento da literacia nos seus alunos, promovendo um ensino da leitura e da escrita que tenha em conta as especificidades linguísticas e ortográficas do português europeu.

Agradecimentos

As autoras agradecem aos revisores anónimos as suas sugestões, que contribuíram para a melhoria deste manuscrito.

Financiamento

A presente investigação foi apoiada por fundos nacionais portugueses e por fundos comunitários europeus atribuídos pela Fundação para a Ciência e a Tecnologia (Portugal) ao inED – Centro de Investigação e Inovação em Educação (UIDP/05198/2020), ao Centro de Linguística da Universidade do Porto (UIDB/00022/2020), ao LE@D-Laboratório de Educação a Distância e eLearning (UIDB/04372/2020) e ao Centro de Linguística da Universidade de Lisboa (UIDB/00214/2020).

Referências

Alegria, J., & Morais J. (1979). Le développement de l'habileté d'analyse phonétique consciente de la parole et l'apprentissage de la lecture. *Archives de Psychologie*, *47*(183), 251–270.

Alves, D. C. (2013). *Efeito das propriedades segmentais em tarefas de consciência segmental, de leitura e de escrita* [Tese de doutoramento, Faculdade de Letras, Universidade de Lisboa]. Repositório da Universidade de Lisboa. http://hdl.handle.net/10451/8629

Amorim, C., & Castelo, A. (2025). A promoção da consciência fonológica no jardim de infância e no 1.º CEB: estudo exploratório sobre práticas educativas, crenças e necessidades formativas. *Revista Portuguesa de Educação, 38*(1), e25011. http://doi.org/10.21814/rpe.33964

Araújo, I. (2004). *A percepção do acento em português: Descrição, implicações e aplicações para o ensino do português como língua materna* [Dissertação de mestrado, Faculdade de Letras da Universidade de Lisboa]. Repositório da Universidade de Lisboa

Cabeleira, S., & Correia, S. (2004). Ditongos, tritongos e hiatos – intuição e propostas de divisão silábica (trabalho com crianças dos 10 aos 13 anos em PE). In T. Freitas & A. Mendes (Orgs.), *Atas do XIX Encontro Nacional da Associação Portuguesa de Linguística (Lisboa, 1–3 de outubro de 2003)* (pp. 351–362). Associação Portuguesa de Linguística.

Carroll, J., Gillon, G., & McNeill, B. (2012). Explicit phonological knowledge of educational professionals. *Asia Pacific Journal of Speech, Language and Hearing*, *15*(4), 231–244. https://doi.org/10.1179/136132812804731820.

Castelo, A. (2004). *A proeminência secundária no português europeu*. [Dissertação de mestrado, Faculdade de Letras da Universidade de Lisboa]. Repositório da Universidade de Lisboa. http://hdl.handle.net/10451/44226

Castelo, A. (2012). *Competência metafonológica e sistema não consonântico no português europeu: descrição, implicações e aplicações para o ensino do português como língua materna*. [Tese de doutoramento, Faculdade de Letras da Universidade de Lisboa]. Repositório da Universidade de Lisboa. http://hdl.handle.net/10451/7454.

Castelo, A. (2023). Competência metafonológica na educação pré-escolar, básica e secundária: alguns (re)conhecimentos e desafios. *Palavras*, 60–61. https://doi.org/10.61248/palavras.vi60-61.170

Costa, T., Rodrigues, C., & Freitas, M. J. (2021). Consoantes (não) soantes em final de palavra: Dados da escrita de crianças portuguesas. *Revista Linguagem & Ensino, 24*(4), 843–867. https://doi.org/10.15210/rle.v24i4.21206.

Direção-Geral da Educação (2018). *Aprendizagens essenciais*: Português [Currículo escolar]. https://www.dge.mec.pt/sites/default/files/Curriculo/

Duarte, I. (2008). *O conhecimento da língua: Desenvolver a consciência linguística*. Ministério da Educação, Direção Geral de Inovação e Desenvolvimento Curricular.

Freitas, M. J., Alves, D., & Costa, T. (2007). *O conhecimento da língua: desenvolver a consciência fonológica*. ME/DGIDC.

Freitas, M. J., Afonso, C., & Castelo, A. (2015) Quantas sílabas tem "cháv(e)na"? Sobre o impacto dos apagamentos vocálicos na segmentação silábica de crianças no ensino pré-escolar. *Linguística - Revista de Estudos Linguísticos da Universidade do Porto*, *10*, 31–58. https://ler.letras.up.pt/uploads/ficheiros/14017.pdf

Giménez, A., Sánchez, A., Flores, A., & Luque, J. L. (2023). Teachers' opinions about the teaching of reading in Spain. *Psicología Educativa*, *29*(1), 65–73. https://doi.org/10.5093/psed2022a11.

Hulme, C., Snowling, M., Caravolas, M., & Caroll, J. (2005). Phonological skills are (probably) one cause of success in learning to read: A comment on Castles and Coltheart. *Scientific Studies of Reading*, *9*, 351–365.

Im, M. (2025). *Peso silábico na perceção dos falantes nativos do PE: um estudo experimental*. [Tese de doutoramento, Faculdade de Letras da Universidade do Porto]. Repositório Aberto da Universidade do Porto. https://hdl.handle.net/10216/166627

Jaskolski, J. E., & Moyle, M. J. (2023). Professional Development in Phonological Awareness for Early Childhood Educators in Low-Income, Urban Classrooms: A Pilot Study Examining Dosage Effects. *Annals of Dyslexia*, *73*(3), 440-468. https://doi.org/10.1007/s11881-023-00289-1.

Leite, I. (2022). A importância da consciência fonémica na aprendizagem da leitura e da escrita. In R. A. Alves & I. Leite (Coords.), *Ensino da leitura e da escrita baseado em evidências* (pp. 237–250). Fundação Belmiro de Azevedo (EDULOG).

Leite, I., Leite, C., Pereira, M. F., & Lemos, G. C. (2022). *Como estão a ser preparados os futuros professores para o ensino da leitura e da escrita*? Fundação Belmiro de Azevedo (EDULOG).

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*, 201–212. https://doi.org/10.1016/0022-0965(74)90101-5.

Lopes, J., Spear-Swerling, L., Oliveira, C., Velasquez, M., Almeida, L., Araújo, L., Zibulsky, J., & Cheesman, E. (2014). *Ensino da leitura no 1º ciclo do ensino básico. Crenças, conhecimentos e formação dos professores*. Fundação Francisco Manuel dos Santos.

Lopes, J. A. (2022). Ensino e aprendizagem da leitura: Fundamentos e aplicações. In R. A. Alves & I. Leite (Coords.), *Ensino da leitura e da escrita baseado em evidências* (pp. 83–98). Fundação Belmiro de Azevedo (EDULOG).

Mateus, M. H. M., & Andrade, E. (2000). *The phonology of Portuguese*. Oxford University Press.

Miranda, A. R., & Matzenauer, C. L. (2010). Aquisição da fala e da escrita: Relações com a fonologia. *Cadernos de Educação*, *35*, 359–405.

Morais, J. (2009). Representações fonológicas na aprendizagem da leitura e na leitura competente. In A. Fiéis & M. A. Coutinho (Org.), *XXIV Encontro Nacional da Associação Portuguesa de Linguística. Textos*

Seleccionados (pp. 7–21). Associação Portuguesa de Linguística. https://apl.pt/wp-content/uploads/2017/09/1-Morais.pdf
Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323–331.
Pereira, I. (2020). Acento de palavra. E. P. Raposo, M. Barbosa, M. A. Martins, A. M. Brito, I. Duarte, & I. Ribeiro (Eds.), *Gramática do português* (Vol. 3, pp. 3399–3425). Fundação Calouste Gulbenkian.
Rodrigues, C., & Lourenço-Gomes, M. C. (2016). Estudo longitudinal da proficiência ortográfica no 2.º e 4.º anos de escolaridade – estruturas /e/, /eI/ e /oU/. *Revista Diacrítica*, *30*(1), 115–136.
Rodrigues, C., & Lourenço-Gomes, M. C. (2018). Representação ortográfica de núcleos nasais na escrita do 2.º e 4.º ano do ensino básico: Dados do português europeu. In C. Lazzarotto-Volcão & M. J. Freitas (Orgs.), *Estudos em fonética e fonologia: Coletânea em homenagem a Carmen Matzenauer* (pp. 357–386). CRV.
Santos, R. N. (2013). *Aquisição de grupos consonânticos e seu impacto nos desempenhos escritos no 1.º ciclo do Ensino Básico.* [Dissertação de mestrado, Faculdade de Letras da Universidade de Lisboa].
Santos, R., Freitas, M. J., & Veloso, J. (2014). Grupos consonânticos na escola: desenvolvimento fonológico e conhecimento ortográfico. *Diacrítica*, *28*(1), 407–436. http://hdl.handle.net/10451/33812
Silva, A. C. (2022). Consciência fonológica e conhecimento das letras. In R. A. Alves & I. Leite (Coords.), *Ensino da leitura e da escrita baseado em evidências* (pp. 167–183). Fundação Belmiro de Azevedo (EDULOG).
Silva, I. L., Marques, L., Mata, L., & Rosa, M. (2016). *Orientações curriculares para a educação pré-escolar*. Ministério da Educação/Direção-Geral da Educação. https://www.dge.mec.pt/ocepe/sites/default/files/Orientacoes_Curriculares.pdf
Treiman, R. (2022). Aprender a escrever palavras. In R. A. Alves & I. Leite (Coords.), *Ensino da leitura e da escrita baseado em evidências* (pp. 205–216). Fundação Belmiro de Azevedo (EDULOG).
Vale, A. P. (2022). Métodos fónicos sistemáticos no ensino da leitura. In R. A. Alves & I. Leite (Coords.), *Ensino da leitura e da escrita baseado em evidências* (pp. 217–235). Fundação Belmiro de Azevedo (EDULOG).
Veloso, J. (2003). Da influência do conhecimento ortográfico sobre o conhecimento fonológico: estudo longitudinal de um grupo de crianças falantes nativas de português europeu. [Tese de doutoramento, Faculdade de Letras da Universidade do Porto]. Repositório Aberto da Universidade do Porto. https://hdl.handle.net/10216/18030
Veloso, J. (2005). A língua na escrita e a escrita da língua. Algumas considerações gerais sobre transparência e opacidade fonémicas na escrita do português e outras questões. *Da Investigação às Práticas. Estudos de Natureza Educacional, 6*(1), 49–69.
Veloso, J. (2022). Fonologia e ortografia do português europeu. In R. A. Alves & I. Leite (Coords.), *Ensino da leitura e da escrita baseado em evidências* (pp. 336–369). Fundação Belmiro de Azevedo (EDULOG).
Veloso, J. (2024). Aspetos da notação formal em Fonologia: Breve relance sobre as suas motivações e alcance e exercício preliminar de aplicação de notação declarativa ao português. *Revista Linguística, 20*(3), 304–330. https://doi.org/10.31513/linguistica.2024.v20n3a65582.

Anexo 1

Este questionário foi elaborado por Clara Amorim, professora da Escola Superior de Educação de Viana do Castelo e investigadora do Centro de Linguística da Universidade do Porto, e por Adelina Castelo, professora da Universidade Aberta e investigadora do Centro de Linguística da Universidade de Lisboa. Tem como objetivo recolher informações sobre o conhecimento fonológico, com a finalidade de fundamentar propostas construtivas úteis no âmbito da formação (inicial e contínua) de professores.

Não há respostas certas nem erradas; o importante é responder às questões de forma sincera.

Ao responder a este questionário, declara que concorda com o tratamento das informações recolhidas apenas para efeitos de investigação científica, sendo garantido o anonimato e a confidencialidade dos dados.

Caso tenha alguma dúvida, sugestão ou queira colaborar mais ativamente neste projeto, poderá contactar-nos através de **cfamorim@ese.ipvc.pt**.

Agradecemos desde já a sua resposta.

1. Com que género se identifica?
a. Mulher
b. Homem
c. Não binário
d. Prefiro não responder

2. Qual é a sua idade?
a. < 18
b. 18-21
c. 22-25
d. 26-29
e. 30-33
f. 34-37
g. 37-40
h. > 40
i. Prefiro não dizer

3. Que curso frequenta?
a. Licenciatura em Educação Básica
b. Mestrado em Educação Pré-Escolar e Ensino do 1.º CEB
c. Mestrado em Ensino do 1.º CEB e Português, História e Geografia de Portugal no 2.º CEB
d. Mestrado em Ensino do 1.º CEB e Matemática e Ciências Naturais no 2.º CEB

4. Qual é a sua língua materna?
a. Português europeu (de Portugal)
b. Português do Brasil
c. Outra

4.1. Se respondeu outra, indique qual é a sua língua materna. Caso contrário, escreva "não se aplica".

_____

5. Ao longo do seu percurso escolar, estudou Fonética e/ou Fonologia?
a. Sim
b. Não

5.1. Se respondeu sim, em que contexto estudou Fonética e/ou Fonologia?
Se respondeu não, selecione "não se aplica".
a. No ensino básico
b. No ensino secundário
c. Na licenciatura em Educação Básica
d. Noutro curso superior
e. Na aprendizagem de uma língua estrangeira
f. Não se aplica

5.1.1. Se selecionou "outro contexto", indique qual. Caso contrário, escreva "não se aplica".
_____

6. Quantas sílabas tem a palavra *esmiuçar*?
a. uma
b. duas
c. três
d. quatro
e. cinco
f. seis
g. sete

7. Quantas sílabas tem a palavra *eventualidades*?
a. uma
b. duas
c. três
d. quatro
e. cinco
f. seis
g. sete

8. Quantas sílabas tem a palavra *laboratório*?
a. uma
b. duas
c. três
d. quatro
e. cinco
f. seis
g. sete

9. Quantas sílabas tem a palavra *têm*?
a. uma
b. duas
c. três

d. quatro
e. cinco
f. seis
g. sete

10. Qual é a sílaba tónica da palavra *bênção*?
Escreva-a em letras maiúsculas (ex.: a sílaba tónica de *mesa* é *ME*).

_____

11. Qual é a sílaba tónica da palavra *residencial*?
Escreva-a em letras maiúsculas (ex.: a sílaba tónica de *mesa* é *ME*).

_____

12. Qual é a sílaba tónica da palavra *resmungam*?
Escreva-a em letras maiúsculas (ex.: a sílaba tónica de *mesa* é *ME*).

_____

13. Qual é a sílaba tónica da palavra *ressentimento*?
Escreva-a em letras maiúsculas (ex.: a sílaba tónica de *mesa* é *ME*).

_____

14. Qual é a sílaba tónica da palavra *interrogar*?
Escreva-a em letras maiúsculas (ex.: a sílaba tónica de *mesa* é *ME*).

_____

15. Quantos sons tem a palavra *cadeira*?
a. um
b. dois
c. três
d. quatro
e. cinco
f. seis
g. sete

16. Quantos sons tem a palavra *carroça*?
a. um
b. dois
c. três
d. quatro
e. cinco
f. seis
g. sete

17. Quantos sons tem a palavra *harpa*?
a. um
b. dois
c. três
d. quatro

e. cinco
f. seis
g. sete

18. Quantos sons tem a palavra *sangue*?
a. um
b. dois
c. três
d. quatro
e. cinco
f. seis
g. sete

19. Quantos sons tem a palavra *táxi*?
a. um
b. dois
c. três
d. quatro
e. cinco
f. seis
g. sete

20. Considerando a pronúncia do português europeu padrão, assinale a(s) palavra(s) que conté(ê)m um ditongo nasal.
a. estudam
b. bom
c. mães
d. irmão
e. pente
f. armazém
g. muito
h. jardim

21. Considerando a pronúncia do português europeu padrão, assinale a(s) palavra(s) que conté(ê)m mais letras do que sons.
a. chato
b. armistício
c. hospital
d. exato
e. queda
f. computador
g. parte
h. melhor

22. Considerando a pronúncia do português europeu padrão, assinale a(s) palavra(s) que conté(ê)m mais sons do que letras.
a. vêm
b. milho

c. vejo
d. fixar
e. exceto
f. muito
g. veem
h. encher

23. Indique a palavra que rima com *órgão*.
a. órfão
b. orégão
c. berbigão
d. outorgam

24. Qual das seguintes palavras termina na sequência consoante + vogal + semivogal?
a. secretária
b. cotovia
c. arbitragem

25. Qual das seguintes palavras termina naa sequência consoante + vogal + vogal?
a. água
b. sorria
c. comeu

26. Considerando a pronúncia do português europeu padrão, que som representa a letra a negrito na palavra *leite*?
a. [i], como em *vi*
b. [e], como em *vê*
c. [ɛ], como em *pé*
d. [a], como em *pá*
e. [ɐ], como em *uma*

27. Considerando a pronúncia do português europeu padrão, que som representa a letra a negrito na palavra *musgo?*
a. [z], como em ***Z****é*
b. [s], como em ***s****ol*
c. [ʃ], como em ***ch****á*
d. [ʒ], como em ***j****á*

28. Considerando a pronúncia do português europeu padrão, que som representa a letra a negrito na palavra *pasta?*
a. [s], como em ***s****ol*
b. [ʃ], como em ***ch****á*
c. [ʒ], como em ***j****á*
d. [z], como em ***Z****é*

29. Considerando a pronúncia do português europeu padrão, que som representa a letra a negrito na palavra *paz?*
a. [z], como em ***Z****é*

b. [ʃ], como em ***ch****á*
c. [s], como em ***s****ol*
d. [ʒ], como em ***j****á*

30. Indique a(s) palavra(s) que termina(m) com os mesmos sons da palavra *mãe*.
a. sai
b. cem
c. porém
d. pães

31. Considerando a pronúncia do português europeu padrão, que sons mudam na primeira sílaba das palavras *vaca* e *faca*?
Porque é que esses sons são diferentes?

_____

_____

32. Que sons mudam na primeira sílaba das palavras *fita* e *finta*?
Porque é que esses sons são diferentes?

_____

_____

Explorando a Inteligência Artificial no ensino de línguas: Criação e classificação automática de corpus de narrativas a partir de provérbios para o ensino de PLE com Large Language Models

Exploring Artificial Intelligence in language teaching: Automatic creation and classification of narrative corpora based on proverbs for teaching PFL with Large Language Models

Jorge Baptista[1,2], *Sónia Reis*[1,2]

[1] Universidade do Algarve – Faculdade de Ciências Humanas e Sociais
[2] INESC-ID Lisboa – Human Language Technology Lab

Abstract

Artificial Intelligence (AI), particularly Large Language Models (LLMs), has been transforming language education. This study explores the application of LLMs in the teaching of Portuguese as a Foreign Language (PFL), focusing on the automatic creation, classification, and validation of a *corpus* of short narratives based on Portuguese proverbs. The objectives are: (i) to automatically generate short narratives based on proverbs, suitable for different proficiency levels as defined by the CEFR; (ii) to automatically classify the linguistic proficiency level of these narratives using traditional machine learning techniques and foundational models (LLMs), followed by validation by human evaluators; (iii) to assess the adequacy of the narratives in relation to the original proverbs, justifying their potential didactic use in the context of PFL teaching. The methodology involves generating narratives with LLMs, which are then validated using automatic tools and human experts, as well as analysing the correspondence between the proverb and the generated narrative with a view to pedagogical application.

Keywords: automatic generation of narrative *corpora*, Portuguese as a Foreign Language (PFL), Artificial Intelligence (AI), Large Language Models (LLMs), Automatic classification of proverbs.

Resumo

A Inteligência Artificial (IA), em particular os Modelos de Linguagem de Grande Escala (LLMs), tem vindo a transformar o ensino de línguas. Este estudo explora a aplicação de LLMs no ensino de Português como Língua Estrangeira (PLE), centrando-se na criação, classificação e validação automática de um *corpus* de narrativas curtas baseadas em provérbios portugueses. Os objetivos centram-se em: (i) gerar automaticamente narrativas curtas baseadas em provérbios, adequadas a diferentes níveis de proficiência definidos pelo QECR; (ii) classificar automaticamente o nível de proficiência linguística dessas narrativas, utilizando técnicas de *machine learning* tradicionais e modelos fundacionais (LLMs), com posterior validação por avaliadores humanos; (iii) avaliar a adequação das narrativas aos provérbios originais, justificando o seu potencial uso didático no contexto do ensino de PLE. A metodologia envolve a geração de narrativas com LLMs, validadas por ferramentas automáticas e por especialistas humanos, bem como a análise da correspondência entre provérbio e narrativa gerada, tendo em vista a sua aplicação pedagógica.

Palavras-chave: geração automática de *corpus* de narrativas, Português como Língua Estrangeira (PLE), inteligência artificial (IA), modelos de linguagem de grande escala (*Large Language Models*), classificação automática de provérbios.

1. Introdução

Nos últimos anos, a inteligência artificial, em particular os modelos de linguagem de grande escala (LLMs), tem vindo a transformar profundamente diversas áreas do conhecimento – e a área da educação não é exceção. Estes modelos, treinados com grandes quantidades de dados linguísticos, são capazes de responder adequadamente a instruções complexas dadas em linguagem natural, gerar e traduzir texto com um grau de complexidade que se aproxima, em muitos casos, do texto produzido por um ser humano, o que os torna ferramentas versáteis para uso pedagógico.

Por *complexidade*, entende-se aqui os aspetos linguísticos do texto cuja variação está ligada à sua maior ou menor legibilidade/interpretabilidade, tais como o comprimento e a estrutura do texto, o vocabulário, a estrutura gramatical, referências culturais e processos estilísticos e, ainda, aspetos relacionados com a coesão e coerência (Monteiro et al., 2023). Como aproximação ao conceito de complexidade, adotamos a perspetiva de Ribeiro et al. (2024), segundo a qual, embora os níveis do *Quadro Europeu Comum de Referência para as Línguas* (QECR) (Conselho da Europa, 2001; Council of Europe, 2018) constituam uma ferramenta amplamente reconhecida para classificar a proficiência em línguas estrangeiras, podem também ser utilizados para identificar o nível de proficiência linguística necessário à compreensão de um determinado texto (tradução nossa) e, nesse sentido, servirem de indicadores do nível de complexidade do texto.

No contexto educativo, os LLMs têm sido progressivamente integrados em ambientes de aprendizagem, quer como apoio à docência, na criação de materiais e atividades, quer como ferramentas ao dispor dos alunos para treino autónomo e feedback imediato (Bonner et al., 2023, Dong et al., 2024, Kasneci et al., 2023). Em particular no ensino de línguas, têm demonstrado potencial para adaptar conteúdos a diferentes níveis de proficiência, sugerir vocabulário contextualizado, simular diálogos realistas e gerar textos idênticos aos produzidos por humanos com objetivos pedagógicos diversos (Benedetto et al., 2025).

Este estudo centra-se na aplicação destas tecnologias ao ensino de Português como Língua Estrangeira (PLE), com especial destaque para a geração automática de narrativas baseadas em provérbios. Os provérbios, enquanto expressões condensadas de sabedoria popular, são recursos linguísticos e culturais particularmente ricos (Reis, 2020). Para além do seu valor lexical e sintático, transportam referências culturais que contribuem para o desenvolvimento da competência pragmática dos alunos (Reis & Baptista, 2017, Reis et al., 2021). Neste sentido, a exploração de provérbios em contexto de criação de texto narrativo, com o apoio da inteligência artificial, revela-se uma estratégia pedagógica inovadora, com potencial para promover não só a aprendizagem da língua, mas também a compreensão intercultural.

Pretende-se, assim, explorar a utilidade dos modelos de linguagem na criação de materiais didáticos para o ensino de Português como Língua Estrangeira. Concretamente, parte-se dos provérbios para gerar automaticamente, com recurso a LLMs, narrativas adaptadas a diferentes níveis de proficiência de PLE, seguindo as orientações do QECR. Para tal, foram definidos três objetivos principais.

O primeiro consiste na constituição de um *corpus* de narrativas curtas, geradas automaticamente com base em provérbios da tradição portuguesa. Estas narrativas serão elaboradas com instruções específicas, fornecidas ao modelo de linguagem, de modo a assegurar a adequação destas a diferentes níveis de proficiência linguística, desde o A1 ao C1. A diversidade temática e a riqueza cultural inerente aos provérbios visam estimular o envolvimento dos alunos e facilitar a aprendizagem de estruturas e vocabulário em contexto.

O segundo objetivo prende-se com o desenvolvimento de um método de classificação automática do nível de proficiência das narrativas assim geradas. Para tal, serão testadas duas abordagens complementares: por um lado, o uso de técnicas tradicionais de aprendizagem automática (ing. *machine learning*) com base em traços (ing. *features*) linguísticos; por outro, a utilização de LLMs enquanto classificadores. A precisão desta

classificação será posteriormente avaliada com recurso a um conjunto de avaliadores humanos, especialistas em ensino de PLE, permitindo aferir a validade das soluções automatizadas propostas.

Por fim, o terceiro objetivo centra-se na avaliação da adequação pedagógica do conteúdo das narrativas face aos provérbios que lhes deram origem. Esta análise terá em conta a coerência temática, a clareza da mensagem e o potencial didático dos textos produzidos, nomeadamente no que diz respeito à sua integração em práticas de ensino-aprendizagem do PLE. Pretende-se, com isto, assegurar que as narrativas não só mantêm o valor cultural e expressivo do provérbio, como também se revelam úteis para serem aplicadas em contextos reais de sala de aula.

A investigação desenvolveu-se, pois, em torno de três questões principais:

i. Até que ponto as narrativas geradas automaticamente são adequadas ao significado e contexto de uso dos provérbios?

ii. Estão essas narrativas alinhadas com os descritores dos níveis de proficiência definidos pelo Quadro Europeu Comum de Referência para as Línguas (QECR)?

iii. Qual é a precisão da classificação automática dos níveis de proficiência, em comparação com a avaliação realizada por especialistas humanos?

Nas secções seguintes, apresentam-se a metodologia adotada, os procedimentos de análise e os principais resultados obtidos, seguidos da respetiva discussão.

2. Metodologia

A metodologia deste estudo desenvolve-se em três etapas principais, correspondentes às fases de geração, classificação e validação de narrativas, geradas automaticamente, com base em provérbios portugueses.

Na primeira etapa, recorremos ao modelo ChatGPT-4 (OpenAI, 2023) para gerar narrativas curtas inspiradas nos provérbios do *Mínimo Paremiólogico do Português Europeu* (MP; Reis & Baptista, 2020). Esta lista de 318 provérbios resulta da análise de um conjunto de mais de 114 mil provérbios recolhidos a partir de dicionários de referência. A seleção dos provérbios baseou-se na sua disponibilidade lexical, estimada com base na frequência de ocorrência em diferentes *corpora*, incluindo manuais escolares e textos jornalísticos. Adicionalmente, dois especialistas contribuíram para a seleção dos provérbios mais representativos, validando as escolhas com recurso a motores de pesquisa e a dois questionários aplicados a um número considerável de falantes. Com base nesta lista final de provérbios muito usuais e bem conhecidos da generalidade dos falantes, o ChatGPT-4 foi instruído para gerar múltiplas narrativas para cada provérbio, adaptadas a diferentes níveis de proficiência linguística, de acordo com os descritores do QECR (Conselho da Europa, 2001; Council of Europe, 2018).

Na segunda etapa, procedemos à classificação automática do nível de proficiência das narrativas geradas. Para isso, foram utilizados diferentes instrumentos: o próprio ChatGPT-4, as ferramentas de classificação Clavis[1] (Curto, 2014; Curto et al., 2015), e Lx-Proficiency[2] (Branco et al., 2014; Santos et al., 2021), ambas baseadas na análise da complexidade de aspetos linguísticos; e um modelo de rede neuronal, treinado especificamente para o português europeu (Ribeiro et al., 2024).

A terceira e última etapa consistiu na validação manual das classificações obtidas, realizada por especialistas na área do ensino de PLE. Esta validação teve como objetivo verificar a adequação linguística, pedagógica e pragmática das narrativas aos níveis atribuídos, bem como garantir a correspondência entre o conteúdo da narrativa e o significado e valor pragmático do provérbio que a originou.

[1] https://string.hlt.inesc-id.pt/demo/classification.pl
[2] https://portulanclarin.net/workbench/lx-proficiency

A seguir, descrevemos cada uma destas etapas e os critérios subjacentes à sua implementação.

2.1. **Geração de narrativas**

A geração de narrativas constituiu a primeira etapa do estudo, envolvendo a definição de instruções específicas para orientar o modelo de linguagem. As instruções aos modelos (ing. *prompts*) foram cuidadosamente construídas de forma a tentar garantir que as narrativas assim geradas fossem o mais adequadas possível aos diferentes níveis de proficiência linguística do QECR, do A1 ao C1 (ver Anexo).

Para cada provérbio, o modelo ChatGPT-4 recebeu instruções explícitas: deveria produzir uma narrativa curta, com um máximo de 100 palavras, em português europeu, evitando o uso direto das palavras-chave associadas ao provérbio. Esta última restrição visava incentivar uma interpretação mais profunda e contextualizada do conteúdo. Para cada provérbio, fornecia-se, então, essa lista de palavras-chave. Quanto ao nível de proficiência, não foi fornecida qualquer instrução adicional ao modelo, uma vez que o objetivo era precisamente testar a capacidade deste para gerar, de forma autónoma, narrativas ajustadas aos diferentes níveis do QECR. Os provérbios foram apresentados ao modelo em blocos de 15, de modo a evitar *prompts* excessivamente longos, o que facilitou o processo de geração das narrativas. A mesma instrução de base foi submetida cinco vezes, uma para cada nível de proficiência (A1, A2, B1, B2 e C1).

Para gerar a lista de palavras-chave associadas a cada provérbio, recorremos ao mesmo modelo de linguagem que foi utilizado na criação das narrativas. As sugestões automáticas foram posteriormente revistas e ajustadas manualmente. Em média, cada provérbio foi associado a três palavras-chave, embora este número possa variar consoante, sobretudo, a extensão da expressão. Por exemplo, o provérbio *Com o fogo não se brinca* foi associado a duas palavras-chave: *fogo* e *brincar*. As restantes palavras foram excluídas por se tratar de elementos gramaticais (ou funcionais) sem carga semântica relevante. Já o provérbio *A laranja de manhã é ouro, à tarde é prata e à noite mata* foi relacionado com sete palavras-chave: *laranja*, *manhã*, *ouro*, *tarde*, *prata*, *noite* e *matar*.[3] A Tabela 1 apresenta a distribuição dos provérbios segundo o número de palavras-chave associadas.

Tabela 1

Distribuição dos Provérbios por Número de Palavras-chave

Palavras-chave	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
Provérbios	1	107	113	80	11	5	1	**318**

Verifica-se que a maioria apresenta entre 2 a 4 palavras-chave. Existem, no entanto, alguns casos fora deste padrão, em que o número de palavras-chave varia entre 1 e 7. No total, foram analisados os 318 provérbios do Mínimo Paremiológico.

Com base nestes provérbios e nas respetivas palavras-chave, foram geradas narrativas adaptadas a diferentes níveis de proficiência. Abaixo, apresentamos um exemplo de uma narrativa automaticamente gerada pelo modelo para o nível A1, a partir do provérbio *Em casa de ferreiro, espeto de pau*. As palavras-chave associadas ao provérbio – *casa*, *ferreiro*, *espeto* e *pau* – foram excluídas, para não se revelar explicitamente o provérbio subjacente.

> O Ricardo é mecânico e cuida muito bem dos carros dos seus clientes. No entanto, o seu próprio carro tem alguns problemas que ele nunca tem tempo de resolver. Os amigos acham engraçado, mas sabem que é porque ele está sempre ocupado a ajudar os outros.

[3] Ignorou-se, assim, o valor adverbial das expressões multipalavra *de manhã*, *à tarde* e *à noite*, retendo apenas os elementos lexicais principais (os nomes).

Tendo por base os 318 provérbios e os cinco níveis de proficiência definidos, foram geradas automaticamente 1 590 narrativas.[4] A Tabela 2 apresenta algumas estatísticas desses textos, nomeadamente, o comprimento (médio, mínimo e máximo) em número de caracteres e o respetivo desvio-padrão, bem como o total de palavras e a média de palavras por texto e por nível.

Tabela 2

Estatísticas do Corpus

Estatísticas	A1	A2	B1	B2	C1
Comp. médio (caracteres)	452	537	574	615	637
Comp. máximo (car.)	604	630	755	832	808
Comp. mínimo (car.)	223	406	465	514	566
Desvio padrão (car.)	67	37	35	38	36
N.º total palavras	**26 184**	**29 687**	**30 801**	**31 495**	**31 581**
N.º médio palavras	**82**	**93**	**97**	**99**	**99**

Verifica-se uma tendência de aumento gradual tanto no comprimento médio das narrativas (em número médio de caracteres), como na contagem média de palavras, à medida que se progride nos níveis de proficiência. Esta tendência confirma o aumento expectável da complexidade textual à medida que se avança nos níveis de proficiência linguística. Adicionalmente, o desvio-padrão do comprimento em número de caracteres mantém-se relativamente estável entre os níveis, exceto para os textos do nível A1, o que parece indicar alguma consistência na extensão dos textos gerados para cada nível. A diferença no comprimento dos textos do nível A1 prende-se com a geração neste caso de vários textos muito curtos, como se vê no comprimento mínimo em caracteres observado para este nível.

2.2. **Classificação automática com base no nível de proficiência**

Para proceder à classificação automática das narrativas geradas, de acordo com o nível de proficiência linguística, tomou-se como referência (ing. *ground truth*) o nível de proficiência (A1-C1) utilizado para a geração dos textos, ainda que essa classificação pudesse não estar inteiramente correta (v. §3.4.).

Foram adotadas quatro abordagens distintas. A primeira consistiu na utilização do ChatGPT-4 numa abordagem *zero-shot*, ou seja, sem qualquer treino específico para a tarefa de classificação. A segunda abordagem recorreu ao *Clavis* (Curto, 2014; Curto et al., 2015), um classificador baseado em algoritmos clássicos de *machine learning*. A terceira envolveu o uso de uma rede neuronal, assente num modelo fundacional, desenvolvido especificamente para a classificação do nível de complexidade de textos em português europeu (Ribeiro et al., 2024). Por fim, para um subconjunto das narrativas, foi também utilizada a ferramenta *Lx-Proficiency* (Branco et al., 2014; Santos et al., 2021), uma plataforma concebida para a classificação automática de textos com base no nível de proficiência linguística.

Apresentam-se, a seguir, estas quatro abordagens com mais pormenor, de forma a clarificar os processos subjacentes a cada uma.

[4] O *corpus* de textos gerados automaticamente (1590_Narrativas_ChatGPT_Proverbios_QECR) está disponível em https://doi.org/10.13140/RG.2.2.22435.16166

2.2.1. Classificação com o ChatGPT

Para esta etapa, recorremos ao modelo ChatGPT, acedido através de uma API (Interface de Programação de Aplicações). Esta interface permite a comunicação direta com o modelo, enviando instruções e recebendo respostas. A principal vantagem da utilização da API reside na sua maior flexibilidade e controlo, permitindo personalizar os pedidos, ajustar diversos parâmetros e realizar múltiplos pedidos em simultâneo – o que não seria viável na interface convencional. Além disso, a API mostra-se particularmente eficaz no tratamento de grandes volumes de dados, como exigido no nosso estudo, que implicava a classificação automática de um número elevado de textos.

Adotámos uma abordagem *zero-shot*, ou seja, o modelo não foi previamente afinado (ing. *fine-tuned*) para a tarefa específica de classificação de textos segundo os níveis do QECR, nomeadamente, não foram fornecidos exemplos de textos anotados quanto ao nível. Solicitámos apenas que atribuísse um nível de proficiência entre A1 e C1 a cada texto, sendo cada instância processada individualmente, sem acesso à informação das classificações anteriores nem a qualquer outro tipo de contextualização adicional. A Tabela 3 apresenta a matriz de confusão dos resultados obtidos com a classificação automática realizada através do ChatGPT. Nas colunas estão os níveis dos textos produzidos pelo modelo na etapa anterior, e nas linhas o resultado da classificação. Assim, por exemplo, para os 318 textos do nível A1, o modelo considerou apenas 2 como sendo desse nível, 157 do nível A2, 159 do nível B1, e nenhum dos níveis B2 e C1.

Tabela 3

Matriz de Confusão dos Resultados daCclassificação com o ChatGPT

	A1	**A2**	**B1**	**B2**	**C1**	**#**
A1	**2**	0	0	0	0	2
A2	157	**34**	3	3	0	197
B1	159	278	**308**	308	181	1 234
B2	0	6	7	**7**	136	156
C1	0	0	0	0	**1**	1
#	318	318	318	318	318	1 590

Tabela 4

Resultados da Classificação do ChatGPT

	A1	**A2**	**B1**	**B2**	**C1**	**Total**
Precisão	0,01	0,11	0,97	0,02	0,00	**0,22**
Precisão adjacente	0,50	0,98	1,00	0,99	0,43	**0,78**

Traduzindo estes resultados em termos de *precisão* (número de classificações corretas pelo total de classificações produzidas pelo modelo) e de *precisão adjacente* (considerando também os casos classificados um nível acima ou abaixo do nível de referência), verifica-se (Tabela 4) que (i) a classificação automática com

o ChatGPT revelou uma precisão global bastante baixa (0,22),[5] com exceção dos textos do nível B1, ou seja o modelo tem uma tendência para classificar a maioria dos textos como B1, o nível médio/central. Como é natural, a precisão adjacente é mais elevada (0,78), o que significa que o desvio entre o nível de referência e o nível atribuído não é muito grande. Assim, podemos concluir que ChatGPT ainda apresenta muitas limitações na tarefa de classificação, não sendo capaz de estabelecer uma distinção clara entre os diferentes níveis do QECR.

2.2.2. Classificação com o Clavis

Nesta secção, analisamos a utilização do classificador Clavis (Curto, 2014; Curto et al., 2015), uma ferramenta treinada com um *corpus* de 237 textos retirados dos exames do Instituto Camões e previamente classificados quanto ao nível de proficiência. O Clavis extrai 52 características linguísticas de cada texto e aplica um algoritmo de aprendizagem automática para proceder à respetiva classificação, em 3 ou em 5 níveis.

Abaixo, apresenta-se a matriz de confusão relativa à classificação automática realizada com a ferramenta Clavis. A diagonal principal da matriz representa os casos corretamente classificados, enquanto os restantes valores correspondem a classificações incorretas. Por exemplo, observa-se que, no nível B1, o modelo conseguiu classificar corretamente 294 textos. Na Tabela 6, encontra-se um resumo com os principais indicadores de desempenho: o modelo acertou 340 casos (verdadeiros positivos) e cometeu 1.250 erros de classificação (falsos positivos). A precisão global foi de 21,4%, e a precisão adjacente – que, como se disse, considera aceitáveis as classificações atribuídas a níveis imediatamente próximos do correto – foi de 46,4%. Este último valor sugere que, embora o desempenho global seja limitado, o modelo revelou alguma capacidade de aproximação ao nível correto em muitos casos.

Tabela 5

Matriz de Confusão da Classificação com o Clavis (Curto et al., 2015)

Clavis	A1	A2	B1	B2	C1	#
A1	**46**	37	24	213	224	544
A2	0	**0**	0	24	55	79
B1	272	281	**294**	79	37	963
B2	0	0	0	**2**	2	4
C1	0	0	0	0	**0**	0
#	318	318	318	318	318	1 590

[5] Na avaliação quantitativa do grau de concordância, usou-se apenas a *concordância bruta* (n.º de casos de acordo/total de casos). Já para a interpretação qualitativa, recorreu-se à escala de Landis e Koch (1977), desenvolvida para os valores de Cohen Kappa (Cohen, 1960), métrica que não foi usada nesta situação já que leva em conta a classificação que seria obtida por mero acaso.

Tabela 6

Resultados da Classificação do Clavis

	A1	A2	B1	B2	C1	Total
Precisão	0,14	0,00	0,92	0,01	0,00	**0,22**
Precisão adjacente	0,14	1,00	0,92	0,25	0,01	**0,47**

Os resultados apresentados na Tabela 6 revelam um desempenho semelhante ao observado na abordagem anterior, com valores baixos de precisão global (22%) e valores moderados de precisão adjacente (47%), talvez porque o *corpus* de treino fosse pouco adequado à tarefa de classificação. Efetivamente este *corpus* não se encontra balanceado (ing. *balanced*; McEnery et al., 2006, pp. 16–19) por níveis, o que afeta a robustez do modelo. Além disso, os dados de anotação manual incluídos no treino evidenciaram baixa concordância entre anotadores, o que fragiliza ainda mais a fiabilidade das classificações produzidas.

2.2.3. Classificação com modelo de rede neuronal

Neste ponto, comparamos os resultados obtidos com o Clavis com os de um novo classificador desenvolvido por Ribeiro et al. (2024). Este classificador foi treinado sobre uma versão ampliada do mesmo *corpus* Camões utilizado pelo Clavis, agora com 598 textos – mais do que o dobro do conjunto de dados original. Além da ampliação do *corpus*, esta abordagem integra uma tecnologia mais recente, baseada numa rede neuronal profunda, e sustentada pelo modelo fundacional *Albertina PT-PT*, desenvolvido por Rodrigues et al. (2023). Trata-se de um modelo especificamente concebido para o processamento do português europeu, o que o torna particularmente adequado para tarefas de classificação linguística em contextos educativos. O modelo foi afinado (ing. *fine-tuned*) com a versão alargada do *corpus* acima referida.

Como se pode observar na Tabela 7, o número de acertos obtido por este classificador é superior ao verificado com o Clavis, totalizando 385 verdadeiros positivos. No entanto, as principais dificuldades do modelo concentram-se nos níveis mais avançados: nenhum texto foi corretamente classificado como B2 ou C1, o que evidencia limitações na diferenciação das produções mais complexas.

Tabela 7

Matriz de Confusão da Classificação com a Rede Neuronal (Ribeiro et al., 2024) & o Modelo Fundacional Albertina (Rodrigues et al., 2023)

RN	A1	A2	B1	B2	C1	#
A1	102	46	37	16	5	206
A2	57	64	62	75	79	337
B1	159	208	219	227	234	1 047
B2	0	0	0	0	0	0
C1	0	0	0	0	0	0
#	318	318	318	318	318	1 590

Tabela 8

Resultados da Classificação da Rede Neuronal

	A1	A2	B1	B2	C1	Total
Precisão	0,32	0,20	0,69	0,00	0,00	**0,24**
Precisão adjacente	0,50	1,00	0,88	0,71	0,00	**0,62**

A precisão global do modelo foi de 24,2%, enquanto a precisão adjacente atingiu 61,9%, revelando uma capacidade considerável de aproximação ao nível correto, mesmo quando a classificação exata não foi atingida.

Comparação de resultados: De um modo geral, em termos de precisão, os resultados dos dois classificadores utilizados são bastante fracos (aprendizagem automática clássica & Clavis: 0,22; rede neuronal + Albertina-PT: 0,24), ainda que o segundo tenha utilizado um *corpus* de treino com praticamente o dobro do tamanho. Mais expressiva foi a diferença observada na precisão adjacente. Neste caso, o Clavis registou 47%, ao passo que o novo classificador alcançou 62% – um ganho significativo, que demonstra maior sensibilidade desta tecnologia às subtilezas linguísticas entre níveis contíguos. Ainda assim, ambos os modelos evidenciam fragilidades na identificação de produções de níveis mais avançados: o Clavis classificou corretamente apenas quatro textos nos níveis B2 e C1 (dois em cada), e o novo classificador não atribuiu corretamente nenhum texto a esses níveis. Estes resultados sugerem que a distinção entre os níveis superiores continua a representar um desafio para estes sistemas de classificação automática.

2.2.4. Classificação automática com o Lx-Proficiency

Nesta secção, e a título meramente ilustrativo, exploramos o desempenho da Lx-Proficiency na tarefa de atribuição automática de níveis de proficiência a um subconjunto de textos do nosso *corpus*.

Este sistema foi inicialmente desenvolvido (Branco et al., 2014) de um modo semelhante ao do sistema Clavis, como um sistema de aprendizagem supervisionada com base em quatro traços (ing. *features*) selecionados (índice de Flesch, proporção de substantivos, comprimento médio de palavras em número de sílabas e comprimento médio de frases em número de palavras). O serviço é disponibilizado numa aplicação *on-line*.[6] Posteriormente, Santos et al. (2021) desenvolveram um modelo do ChatGPT2 afinado (ing. *fine-tuned*) com uma versão maior do *corpus* do Instituto Camões, com 500 textos, e que passou a ser usado como base daquela aplicação.

Dadas as limitações na utilização da ferramenta Lx-Proficiency, cuja interface não permite processar em bloco o conjunto dos 1 590 textos gerados, esta foi aplicada apenas aos 75 textos de 15 provérbios, selecionados aleatoriamente (esta seleção de provérbios será apresentada mais adiante). Veja-se, na Tabela 9, a distribuição das classificações atribuídas.

[6] https://portulanclarin.net/workbench/lx-proficiency

Tabela 9

Matriz de Confusão da Classificação com o Lx-Proficiency (Branco et al., 2014; Santos et al., 2021)

Lx-Proficiency	A1	A2	B1	B2	C1	#
A1	**8**	0	7	0	0	15
A2	5	**0**	10	0	0	15
B1	4	0	**11**	0	0	15
B2	2	0	13	**0**	0	15
C1	2	0	12	0	**1**	15
#	21	0	53	0	1	75

A precisão global da Lx-Proficiency foi de 27%, ligeiramente superior à registada pelos classificadores previamente analisados. A precisão adjacente, por sua vez, situou-se nos 64%, o que revela uma performance moderada na identificação dos níveis de proficiência, com alguma capacidade de aproximação ao nível correto.

Tabela 10

Resultados daCclassificação do Lx-Proficiency

	A1	A2	B1	B2	C1	Total
Precisão	0,38	-	0,21	-	1,00	**0,27**
Precisão adjacente	0,62	-	0,64	-	1,00	**0,64**

Em síntese, e de um modo geral, a precisão global foi baixa em todos os classificadores aqui testados, situando-se entre os 22% do Clavis e os 27% do Lx-Proficiency, o que demonstra as dificuldades persistentes na classificação precisa de textos por nível segundo o QECR. No entanto, os valores de precisão adjacente, especialmente no caso do ChatGPT (78%) e do classificador neuronal (62%), revelam já uma capacidade de acerto, o que pode ser particularmente útil em contextos pedagógicos onde a gradação por níveis contíguos seja razoavelmente aceitável. Outro dado comum às diferentes abordagens diz respeito à dificuldade na distinção dos níveis mais avançados, nomeadamente B2 e C1. Nenhum dos modelos foi capaz de classificar corretamente um número significativo de textos nesses níveis, o que sugere que a complexidade lexical e estrutural destas produções ainda não é plenamente capturada pelos sistemas automáticos, mesmo quando suportados por modelos de língua relativamente recentes e especificamente desenvolvidos para o português europeu.

Apesar destas limitações, os resultados obtidos oferecem pistas para futuras melhorias nos processos de classificação automática de textos. A utilização de *corpora* mais extensos, adequadamente balanceados, e a diversificação/complexificação das instruções aos modelos (ing. *prompts*) poderão contribuir para o desenvolvimento de ferramentas mais eficazes e sensíveis às nuances dos diferentes níveis de proficiência linguística em PLE.

3. Classificação manual

Avançamos agora para a fase de classificação manual, que visava validar e complementar os resultados das abordagens automáticas previamente descritas. A classificação manual foi realizada com base nos mesmos 15 provérbios utilizados na fase anterior, selecionados aleatoriamente do Mínimo Paremiológico e totalizando

75 textos. Os textos foram avaliados e classificados manualmente por dois anotadores, com conhecimento aprofundado do QECR e do ensino de PLE. Os 15 provérbios utilizados nesta fase do estudo são os seguintes:

Tabela 11

Lista dos 15 Provérbios da Seleção Aleatória

ID	Provérbios
ID_102	*Em boca fechada não entra mosca.*
ID_108	*Enquanto o pau vai e vem folgam as costas.*
ID_111	*Errar é humano.*
ID_112	*Faz o que eu digo e não faças o que eu faço.*
ID_135	*Manda quem pode.*
ID_176	*O prometido é devido.*
ID_199	*Os fins justificam os meios.*
ID_212	*Para bom entendedor, meia palavra basta.*
ID_215	*Para quem é, bacalhau basta.*
ID_240	*Quem corre por gosto não cansa.*
ID_249	*Quem está mal que se mude.*
ID_291	*Recordar é viver.*
ID_305	*Tristezas não pagam dívidas.*
ID_306	*Tudo está bem quando acaba bem.*
ID_314	*Vão-se os anéis, ficam os dedos.*

Para a avaliação da complexidade linguística das narrativas geradas automaticamente, foram consideradas cinco categorias principais, adaptadas da definição dos níveis de complexidade do projeto iRead4Skills[7] (Monteiro et al., 2023) e articuladas com as orientações do QECR. As categorias em análise foram:

i. Comprimento e estrutura do texto – Varia de textos curtos, até 50 palavras, com estruturas simples e lineares (nível 1), até textos longos e complexos, com estruturas intricadas e variação significativa de tópicos, exigindo leitura atenta e conhecimento prévio (nível 5).

ii. Vocabulário – Vai de palavras simples e frequentes, relacionadas com o quotidiano e objetos concretos, sem uso de siglas ou estrangeirismos (nível 1), até vocabulário mais complexo, com palavras raras, jargões específicos e uso intensivo de estrangeirismos e siglas não explicadas (nível 5).

iii. Complexidade gramatical – Vai de frases curtas e simples, com estruturas de coordenação básicas, uso predominante da voz ativa e tempos verbais limitados ao presente e aos pretéritos simples (nível 1), a frases muito complexas, com coordenação e subordinação, construções gramaticais mais complexas como a voz passiva impessoal e o uso de tempos verbais compostos, incluindo o pretérito mais-que-perfeito (nível 5).

[7] https://zenodo.org/records/10459090

iv. Referências culturais e estilísticas – Vai de textos sem qualquer figura de estilo ou referência cultural, caracterizados por uma linguagem denotativa, clara e direta (nível 1), a textos densamente marcados por figuras de estilo mais complexas (como metáforas ou analogias sofisticadas) e referências culturais ou históricas específicas, cuja interpretação requer conhecimento prévio e elevado grau de inferência (nível 5).

v. Coesão e coerência – Vai de textos com cadeia referencial completa, sem omissões, e com repetição de termos para garantir clareza, organizados segundo uma sequência temporal linear (nível 1), a textos com coesão e coerência altamente complexas, marcados por uso extensivo de omissões, pronomes e sequências temporais intricadas, que requerem uma atenção acrescida por parte do leitor (nível 5).

Cada categoria foi avaliada numa escala de Likert de 1 a 5, em que 1 corresponde a "muito fácil" e 5 a "muito difícil". Esta grelha permitiu uma análise sistemática da complexidade de cada narrativa, contribuindo para a aferição da sua adequação ao nível de proficiência-alvo. A atribuição do nível global ao texto[8] resultou da soma das pontuações atribuídas a cada uma das cinco categorias anteriormente descritas, segundo o critério seguinte: A1 ={5,8}, A2={9,12} B1={13,16} B2={17,20} e C1={21,25}.

A seguir, apresenta-se a matriz de confusão relativa à concordância entre os dois avaliadores na classificação manual das narrativas segundo os níveis de proficiência. Esta matriz permite observar o alinhamento (ou divergência) entre os juízos atribuídos por cada anotador.

Tabela 12

Acordo entre Dois Anotadores (nas Colunas o Anotador 1 e nas Linhas o Anotador 2)

	A1	**A2**	**B1**	**B2**	**C1**
A1	**4**	6	3	2	
A2	0	**11**	4		
B1	1	3	**11**	0	
B2		2	6	**7**	0
C1		2	4	9	**0**

A taxa de concordância exata foi relativamente baixa, situando-se em torno dos 44% (ver Tabela 13), o que reflete a complexidade da tarefa e a subjetividade envolvida na avaliação de produções linguísticas com base em múltiplos critérios.

[8] Adotou-se, assim, uma abordagem analítica e indutiva, partindo do princípio de que a identificação de cada uma das categorias individuais é, por si só, uma tarefa complexa, mas que oferece maior sustentação à classificação final. Considerou-se, ainda, que a atribuição direta de um grau de dificuldade global seria mais suscetível a enviesamentos.

Tabela 13

Taxa de Acordo entre Anotadores

	A1	A2	B1	B2	C1	Total
Precisão	0,19	0,00	0,21	0,00	0,00	**0,44**
Precisão adjacente	0,19	0,00	0,40	0,00	0,00	**0,81**

Apesar dessa dificuldade, a precisão adjacente foi significativamente mais elevada, atingindo os 81%. Este valor indica que, mesmo quando os avaliadores não coincidiram no nível exato a atribuir ao texto, a grande maioria das discrepâncias ocorreu entre níveis contíguos, o que evidencia um grau considerável de alinhamento e consistência interpretativa.

3.2 **Acordo entre anotador e referência**

Com o intuito de validar as avaliações manuais, esta secção apresenta os dados comparativos entre a classificação de cada anotador e o nível de referência definido para a geração de cada texto. Nas tabelas seguintes, apresentam-se os dados relativos à comparação entre cada anotador e a referência, com base nas classificações manuais atribuídas às 75 narrativas.

Tabela 14

Acordo Anotadores-Referência (à Esquerda Anotador1 e à Direita o Anotador 2)

Anot 1	A1	A2	B1	B2	C1	#	Anot 2	A1	A2	B1	B2	C1	#
A1	**9**	1		3	2	15	**A1**	**5**	10				15
A2	3	**10**	2			15	**A2**		**12**	2			14
B1	3	1	**10**	1		15	**B1**		3	**13**	9	3	28
B2		1	1	**8**	5	15	**B2**				**6**	12	18
C1		2	2	3	**8**	15	**C1**					**0**	0
#	15	15	15	15	15	75	#	5	25	15	15	15	75

Tabela 15

Matriz de Confusão das Anotações (à Esquerda, o Anotador1 e, à Direita, o Anotador 2)

Anot1	A1	A2	B1	B2	C1	Total	Anot2	A1	A2	B1	B2	C1	Total
Precisão	0,60	0,67	0,67	0,53	0,53	**0,60**	**Precisão**	1,00	0,48	0,87	0,40	0,00	**0,48**
PAdj.	0,80	0,80	0,87	0,80	0,87	**0,83**	**PAdj.**	1,00	1,00	1,00	1,00	0,80	**0,96**

À esquerda, observa-se a matriz de confusão do Anotador 1 em relação à referência, com uma taxa de concordância exata de 60% e uma taxa de concordância adjacente de 83%. Estes valores indicam um acordo moderado entre as classificações atribuídas pelo anotador e os níveis definidos como referência, mas elevado no que diz respeito à distinção entre níveis contíguos.

À direita, é apresentada a comparação entre o Anotador 2 e a referência, com uma taxa de precisão também moderada, mas inferior à do Anotador 1 (48%), e uma taxa de precisão adjacente bastante mais elevada, quase perfeita (96%). Apesar da menor coincidência no nível exato, os dados sugerem que as classificações do Anotador 2 se mantiveram muito próximas das da referência, com a maioria das discrepâncias situadas entre níveis vizinhos, o que evidencia uma boa sensibilidade ao grau de dificuldade relativo das narrativas.

3.3 **Classificação manual texto-provérbio**

A última tarefa realizada neste estudo consistiu em atribuir manualmente a cada texto 1 dos 15 provérbios do mesmo conjunto anteriormente referido. Os anotadores foram os mesmos que realizaram a tarefa anterior. Trata-se de uma tarefa de estabelecimento de correspondência, pois os anotadores conheciam a lista dos provérbios a partir dos quais foram gerados os textos. O objetivo foi avaliar a adequação do texto ao provérbio. Utilizou-se uma escala de Likert de 0 a 2, de acordo com as diretrizes que se apresentam abaixo:

0: O texto não corresponde a nenhum provérbio da lista
1: O texto corresponde, mas de forma inadequada a um dos provérbios da lista, ou pode aplicar-se-lhe mais do que um provérbio
2: O texto corresponde perfeitamente a um dos provérbios da lista

Apresentamos a seguir os resultados da correspondência entre os textos e os provérbios, feita pelo Anotador 1. Nas colunas indica-se o n.º de identificação de referência do provérbio que serviu de base para a geração do texto (ver Tabela 16) e nas linhas o n.º de identificação atribuído pelo anotador. Para 3 textos, o Anotador 1 não conseguiu determinar um provérbio da lista a que pudesse fazer corresponder o texto gerado (coluna '0').

Cada célula da matriz mostra o número de provérbios atribuídos correta ou incorretamente a um determinado texto.

Tabela 16

Verificação da Correspondência entre o Texto e o Provérbio, Feita pelo Anotador 1

Anot1	0	102	108	111	112	135	176	199	212	215	240	249	291	305	306	314	
102		**5**															5
108	**1**		**3**												**1**		5
111				**5**													5
112					**5**												5
135	**1**					**3**									**1**		5
176							**5**										5
199							**1**	**2**							**2**		5
212									**5**								5
215										**0**					**5**		5
240	**1**										**4**						5
249												**5**					5
291													**5**				5
305														**5**			5
306															**5**		5
314																**5**	5
Total	3	5	3	5	5	3	6	2	5	0	4	5	5	5	14	5	

A diagonal mostra os casos de correspondência, onde as células a **verde** indicam que o anotador acertou todas as cinco vezes, atribuindo corretamente aos cinco textos o provérbio correspondente. Já as células a **amarelo** indicam que houve acertos, mas não em todos os textos. As células fora da diagonal revelam os erros de atribuição, indicados a **vermelho**. O total de casos corretamente assinalados nesta avaliação foi de 62, resultando numa precisão de 83%, o que pode ser considerado um acordo quase perfeito.

A Tabela 17 mostra os resultados da atribuição de correspondência entre os textos e os provérbios, realizada pelo Anotador 2.

Tabela 17

Verificação da Correspondência entre o Texto e o Provérbio, Feita pelo Anotador 2

Anot2	0	102	108	111	112	135	176	199	212	215	240	249	291	305	306	314	
102		**5**															5
108			**5**														5
111				**5**													5
112					**5**												5
135	**1**					**4**											5
176							**5**										5
199								**4**			**1**						5
212									**5**								5
215										**5**							5
240											**5**						5
249												**5**					5
291													**5**				5
305														**5**			5
306															**5**		5
314																**5**	5
Total	1	5	5	5	5	4	5	4	5	5	6	5	5	5	5	5	

Com um total de 73 verdadeiros-positivos, a precisão da classificação foi de 97%, o que demonstra uma precisão quase perfeita na atribuição dos provérbios aos textos, com apenas dois casos em que os provérbios foram incorretamente identificados.

A elevada qualidade da classificação por parte dos dois anotadores pode ser explicada pelo facto de se trabalhar com um conjunto limitado de 15 provérbios e previamente conhecidos pelos avaliadores, o que terá facilitado o processo de correspondência entre textos e provérbios.

Comparando as classificações de ambos os anotadores, como se pode observar na Tabela 18, foram registados 63 casos de concordância, o que se traduz numa taxa de acordo de 0,84 – valor que representa um nível bastante elevado de consistência entre os dois anotadores.

Tabela 18

Concordância entre os Dois Anotadores

Anot1/2	0	102	108	111	112	135	176	199	212	215	240	249	291	305	306	314
0	1		1								1					
102		5														
108			3													
111				5												
112					5											
135						3										
176							5	1								
199								2								
212									5							
215										0						
240											4					
249												5				
291													5			
305														5		
306			1			1		1		5	1				5	
314																5
Total	1	5	5	5	5	4	5	4	5	5	6	5	5	5	5	5

3.4 Narrativas vs. Referencial Camões: análise linguística

Uma outra tarefa desenvolvida consistiu em verificar em pormenor se os fenómenos linguísticos presentes nos textos gerados estavam adequados ao nível de proficiência atribuído. Como meros exemplo, apresentam-se a seguir alguns fenómenos gramaticais identificados nos textos dos níveis A1 e A2, que foram gerados com base no provérbio *Em boca fechada não entra mosca*.

> [ID_102, **A1**] O Tiago estava com os amigos a conversar sobre um assunto delicado. Ele tinha vontade de dar a sua opinião, mas percebeu que poderia causar confusão. Decidiu ouvir os outros em vez de falar, e assim evitou uma discussão desnecessária.
>
> [ID_102, **A2**] Durante uma reunião, o João ouviu uma opinião com a qual não concordava. Quis responder imediatamente, mas lembrou-se de que era melhor refletir antes de falar. Esperou até o fim da reunião para dar a sua opinião de forma ponderada e construtiva.

No texto de nível A1, observa-se o uso do pretérito perfeito do indicativo (*evitou*), tempo verbal associado ao nível A2 conforme o inventário de gramática[9] do Referencial Camões PLE (Direção de Serviços de Língua e Cultura, 2017). Identificam-se também estruturas como o complexo (ou perífrase) verbal "estar a + infinitivo",

[9] https://www.instituto-camoes.pt/activity/centro-virtual/referencial-camoes-ple

mas com o verbo *estar* no pretérito imperfeito; a preposição *sobre* com função de "expressão de noção situacional (sobre)"; e a locução conjuncional *em vez de,* a introduzir uma oração (*em vez de falar*) – todos elementos que, no Referencial, são associados ao nível A2.

No texto classificado como A2, surgem fenómenos de maior complexidade, como o uso de um verbo auxiliar modal no pretérito perfeito + infinitivo (*quis responder*); o pronome relativo *a qual*, que o Referencial apresenta associado ao nível B2; e a preposição *durante*, empregue para fazer a situação de um evento (a oração principal) no tempo de outro evento (*a reunião*) – um fenómeno explicitamente descrito no Referencial como próprio do nível B1.

Seria fastidioso estar aqui a alistar todos os fenómenos detetados em que se verificou uma não correspondência entre as estruturas observadas nos textos gerados artificialmente e as orientações do Referencial. Esse trabalho de anotação prossegue ainda, a fim de se poderem tirar conclusões mais sólidas sobre este aspeto.

4. Conclusões e trabalho futuro

Podemos concluir que, de um modo geral, o ChatGPT se revela capaz de produzir textos que são, na sua maioria, adequados aos diferentes níveis de proficiência do QECR para que foram solicitados, mesmo numa abordagem *zero-shot*. Contudo, identificámos algumas situações em que se observa a introdução de fenómenos linguísticos que, segundo o Referencial Camões, seria mais apropriado explorar em níveis superiores.

O modelo consegue igualmente produzir textos apropriados aos provérbios, mantendo, na maioria dos casos, uma relação coerente entre o conteúdo gerado e o sentido do provérbio.

Verificámos ainda que os humanos não são particularmente eficazes na classificação do nível de proficiência dos textos, mas que os modelos de linguagem também não o são, independentemente dos dados em que foram treinados ou da abordagem adotada – seja baseada em redes neuronais ou em métodos mais tradicionais de aprendizagem automática.

Do ponto de vista pedagógico, consideramos que estas ferramentas têm elevado potencial para poderem ser utilizadas, mas a sua aplicação requer validação humana, especialmente quando os textos são integrados em contextos de aprendizagem reais.

Para trabalho futuro, pretendemos envolver mais anotadores nas tarefas manuais aqui apresentadas, com o objetivo de (a) contribuir para uma avaliação mais robusta dos níveis de proficiência associados e (b) confirmar a adequação das narrativas aos provérbios que lhes deram origem, mas sem os revelar previamente a esses anotadores.

Agradecimentos

Os autores agradecem igualmente ao Professor Eugénio Ribeiro a utilização do seu modelo de redes neuronais (Ribeiro et al., 2024) para a classificação automática das narrativas inspiradas em provérbios analisadas neste estudo.

Financiamento

Este trabalho foi parcialmente financiado por fundos nacionais através da FCT (Referência: UIDB/50021/2020, DOI: 10.54499/UIDB/50021/2020) e pela Comissão Europeia (Projeto: iRead4Skills, Número da subvenção: 1010094837, Tópico:HORIZON-CL2-2022-TRANSFORMATIONS-01-07, DOI: 10.3030/101094837).

Referências

Benedetto, L., Gaudeau, G., Caines, A., & Buttery, P. (2025). Assessing how accurately large language models encode and apply the Common European Framework of Reference for Languages. *Computers and Education: Artificial Intelligence*, *8*, 100353. https://doi.org/10.1016/j.caeai.2024.100353

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English with Technology*, *1*(23), 23–41.

Branco, A., Rodrigues, J., Silva, J., Costa, F., & Vaz, R. (2014). Assessing automatic text classification for interactive language learning. In *Proceedings International Conference on Information Society (i-Society 2014)* (pp. 70–78). IEEE. https://doi.org/10.1109/i-Society.2014.7009014

Cohen, Jacob (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Conselho da Europa. (2001). *Quadro Europeu Comum de Referência para as Línguas – Aprendizagem, ensino, avaliação*. Edições ASA.

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment, companion volume with new descriptors*. Council of Europe. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Curto, P. (2014). *Classificador de textos para o ensino de português como segunda língua* [Dissertação de mestrado não publicada]. Instituto Superior Técnico-Universidade de Lisboa.

Curto, P., Mamede, N., & Baptista., J. (2015). Automatic text difficulty classifier. In *CSEDU: Proceedings of the International Conference on Computer Supported Educatio*n (pp. 36–44). SCITEPRESS. https://doi.org/10.5220/0005428300360044

Direção de Serviços de Língua e Cultura. (2017). *Referencial Camões PLE – Português Língua Estrangeira*. Camões, Instituto da Cooperação e da Língua, I.P. https://www.instituto-camoes.pt/images/REFERENCIAL_ebook.pdf

Dong, B., Bai, J., Xu, T., & Zhou, Y. (2024). Large language models in education: A systematic review. In *Proceedings 2024 6th International Conference on Computer Science and Technologies in Education (CSTE)* (pp. 131–134). IEEE. https://doi.org/10.1109/CSTE62025.2024.00005

Ghosh, S. & Srivastava, S. (2021). *EPiC: Employing proverbs in context as a benchmark for abstract language understanding*. ArXiv preprint.

Grosso, M., Soares, A., Sousa F. & Pascoal, J. (2011). *QuaREPE – Quadro de Referência para o Ensino Português no Estrangeiro – Documento Orientador*. Ministério da Educação.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllemeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, A., Sailer, M., Schmidt, A., Seidel, T., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochem Med*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Monteiro, R., Amaro, R., Correia, S., Pintard, A., Gauchola, R., Moutinho, M., & Blanco Escoda, X. (2023). *iRead4Skills – Complexity Levels* (V1.0). Zenodo. https://doi.org/10.5281/zenodo.10459090

OpenAI. (2020). *GPT-3 via API* [Modelo de linguagem de grande escala]. https://platform.openai.com/

OpenAI. (2023). *ChatGPT* (versão GPT-4) [Modelo de linguagem de grande escala]. https://chat.openai.com/

Reis, S., & Baptista, J. (2017). Os provérbios em manuais de ensino de português língua não materna. In V. Pinheiro & G. H. Paetzold (Orgs.), *Anais do XI simpósio brasileiro de tecnologia da informação e da linguagem humana* (pp. 247–255). SBC.

Reis, S., & Baptista, J. (2020). Determinação de um mínimo paremiológico do português europeu. *Acta Scientiarum. Language and Culture*, *2*(42), e52114.

Reis, S. (2020). *Expressões proverbiais do português-usos, variação formal e identificação automática* [Tese de doutoramento não publicada]. Universidade do Algarve.

Reis, S., Baptista, J., & Mamede, N. (2021). Provérbios portugueses usuais: Distribuição em *corpora*. In E. Ruiz, T. Torrent, J. Souza, M. Madruga, R. Rodrigues, C. Barros & D. Claro (Orgs.), *Anais do XIII simpósio brasileiro de tecnologia da informação e da linguagem humana* (pp. 325–334). SBC.

Ribeiro, E., Mamede, N., & Baptista, J. (2024). Automatic Text Readability Assessment in European Portuguese. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese* (Vol. 1, pp. 97–107). Association for Computational Linguistics.

Rodrigues, J., Gomes, L., Silva, J., Branco, A., Santos, R., Cardoso, H., & Osório, T. (2023). Advancing neural encoding of Portuguese with transformer Albertina PT-*. In N. Moniz, Z. Vale, J. Cascalho, C. Silva & R. Sebastião (Eds.), *Progress in Artificial Intelligence. EPIA 2023* (part I, pp. 441–453). Springer. https://doi.org/10.1007/978-3-031-49008-8_35

Santos, R., Rodrigues, J., Branco, A., & Vaz, R. (2021). Neural text categorization with transformers for learning Portuguese as a second language. In G. Marreiros, F. S. Melo, N. Lau, H. Lopes Cardoso & L. Reis (Eds.), *Progress in Artificial Intelligence. EPIA 2021* (pp. 715–726). Springer. https://doi.org/10.1007/978-3-030-86230-5_56

Anexo

Este anexo apresenta um exemplo de *prompt* elaborado para orientar a geração automática de narrativas com recurso a um modelo de linguagem de grande escala (LLM).

Prompt:

Considera que és um especialista em Ensino de Português como Língua Estrangeira e que tens a tarefa de construir textos adequados ao nível de proficiência esperado para estudantes desse nível.

Vais receber abaixo uma linha de exemplo com um provérbio português, precedido do respetivo ID e uma lista de palavras-chave que *NÃO* devem ser usadas no texto a produzir. Aqui está um exemplo de um provérbio (input):

ID_075 Casa roubada, trancas à porta. casa,roubar,tranca,porta

O total de provérbios é 318. No entanto, vou apresentar-te somente 15 provérbios de cada vez.

Instrução:

Constrói uma pequena narrativa, em português europeu (respeitando o Novo Acordo Ortográfico), com um máximo de 100 palavras, que conte uma história adequada ao provérbio indicado e que o texto seja adequado a leitores com um nível de proficiência *A1* do Quadro Comum Europeu de Referência para as Línguas (QECR). Nesse texto, *NÃO* deves usar as palavras-chave indicadas.

Os textos do Nível A1 (Iniciante) podem ser descritos assim: Textos simples e curtos, frases simples, vocabulário básico, temas quotidianos (saudações, informações pessoais, compras).

Aqui estão os provérbios a considerar e a lista de palavras-chave que *NÃO* devem ser usadas no texto a produzir:

ID_001 A beleza está nos olhos de quem a vê. beleza,olho,ver
ID_002 A carne é fraca. carne,fraco
ID_003 A cavalo dado não se olha ao dente. cavalo,dar,olhar,dente
ID_004 A culpa morreu solteira. culpa,solteiro
ID_005 A curiosidade matou o gato. curiosidade,matar,gato
ID_006 A descer todos os santos ajudam. descer,santos,ajudar
ID_007 A esperança é sempre a última a morrer. esperança,sempre,último,morrer
ID_008 A experiência é a mãe da ciência. experiência,mãe,ciência
ID_009 A falar é que a gente se entende. falar,gente,entender
ID_010 A fé move montanhas. fé,mover,montanha
ID_011 A fome é o melhor tempero. fome,melhor,tempero
ID_012 A galinha da vizinha é sempre melhor que a minha. galinha,vizinho,melhor,minha
ID_013 A idade não perdoa. idade,perdoar
ID_014 A intenção é que conta. intenção,contar
ID_015 A justiça tarda mas não falta. justiça,tardar,faltar

Atenção:

(i) o texto deverá estar em PORTUGUÊS EUROPEU: não te esqueças dos artigos antes dos nomes próprios (e.g. *o* João) e do pronome possessivo (*o* seu amigo), usa apenas vocabulário do português europeu.

(ii) Não deves inserir nenhum provérbio e não podes usar as palavras que formam o provérbio que te foi indicado.

(iii) Os textos não podem ter nenhum newline (parágrafo).

A referência metalinguística no processo de escrita colaborativa

Metalinguistic reference in the collaborative writing process

Luís Filipe Barbeiro[1], *Célia Barbeiro*[2]

[1] Escola Superior de Educação e Ciências Sociais, Instituto Politécnico de Leiria / CELGA-ILTEC
[2] Escola Dr. Correia Mateus / CELGA-ILTEC

Abstract

The collaborative writing process activates metalinguistic reference during the interaction that participants engage in to generate and make decisions about the linguistic-textual units that will form the text. The purpose of the study presented in this article is to analyse metalinguistic reference in the collaborative writing process of basic education students. It takes as its theoretical framework the complexity of the writing process, considering the different linguistic-textual levels and components involved, and the particular characteristics of the process when a text is produced collaboratively. Methodologically, the interaction of students from different school years during the process was recorded. The analysis focused on the presence of manifestations of metalinguistic reference, the mechanisms that it activates, the linguistic-textual levels or domains involved, and the metalinguistic terms used. In addition to the overall scores, the results for the different years of schooling, second, fourth, sixth and eighth, were considered in order to see trends in progression. The results show the presence of a metalinguistic reference in the process analysed, with an extensive use of metalinguistic terms. The reference covers the different linguistic-textual levels and domains, with an emphasis on punctuation. Metalinguistic terms are used from a functional point of view to serve operations and decisions and to guarantee the homogeneity of the text among the participants. The comparison between the different years shows a significant decrease in the weight of spelling from the second to the fourth year and the increasing importance of the level associated with the domains of genre, text and discourse. The study also shows the relevance of the format of the collaborative writing task used for the results obtained in the different domains and for the activation of knowledge related to grammar learning.

Keywords: writing, collaborative writing, writing process, metalanguage, metalinguistic consciousness

Resumo

O processo de escrita colaborativa ativa a referência metalinguística na interação que os participantes colocam em prática para gerar e tomar decisões sobre as unidades linguístico-textuais que formarão o texto. O estudo apresentado no presente artigo teve como objetivo analisar a referência metalinguística no processo de escrita colaborativa de alunos do ensino básico. Toma como enquadramento teórico-concetual a complexidade do processo de escrita quanto a níveis linguístico-textuais que são objeto de operações e quanto a componentes ativadas e as características que o processo adquire quando a produção de um texto é realizada em conjunto. Metodologicamente, procedeu ao registo da interação de alunos de diferentes anos de escolaridade no decurso do processo. A análise incidiu na presença de manifestações de referência metalinguística, nos mecanismos ativados para a sua construção, nos níveis ou domínios linguístico-textuais sobre que incide e nos termos metalinguísticos utilizados. Além dos valores globais, considerou os valores dos diversos anos de escolaridade, segundo, quarto, sexto e oitavo, para apreender tendências de progressão. Os resultados evidenciam a presença da referência metalinguística no processo analisado, com amplo recurso a termos metalinguísticos. A referência abrange os diversos níveis e domínios linguístico-textuais, salientando-se o domínio da pontuação. Os termos metalinguísticos são utilizados numa

perspetiva funcional ao serviço das operações e decisões e para garantir a homogeneidade do produto textual entre os participantes. A comparação entre os diversos anos revela a diminuição significativa do peso da ortografia do segundo para o quarto ano e o relevo crescente do nível associado aos domínios do género, texto e discurso. Emerge também do estudo a relevância do formato da tarefa de escrita colaborativa posta em prática para os valores alcançados pela referência metalinguística nos diversos domínios e para a ativação do conhecimento ligado à aprendizagem gramatical.

Palavras-chave: escrita, escrita colaborativa, processo de escrita, metalinguagem, consciência metalinguística

1. Introdução

A escrita colaborativa com interação direta e síncrona entre os participantes (Lowry et al., 2004) implica a construção conjunta do texto, que vai acontecendo com o desenrolar dessa interação. A colaboração manifesta-se por meio da apresentação de propostas de formulação e reformulação que incidem sobre as unidades linguístico-textuais que formarão o texto (Barbeiro, 2019, 2023) e por meio da realização de escolhas e tomadas de decisão. A interação que tem lugar ativa a referência metalinguística, ou seja, a referência às próprias unidades linguístico-textuais que são objeto de decisão.

O presente artigo toma como campo de estudo a interação que se desenvolve na escrita colaborativa, focando-se, de modo específico, na presença da referência metalinguística. Tem como objetivo analisar a referência metalinguística no processo de escrita colaborativa de alunos do ensino básico, de diferentes anos de escolaridade, em relação à sua presença no processo, aos mecanismos que ativa, aos domínios linguístico-textuais sobre que incide e aos termos metalinguísticos que são utilizados. Procura ainda apreender eventuais progressões ao longo da escolaridade do ensino básico, em relação aos aspetos em foco.

Associadas a esses objetivos, surgem, por conseguinte, algumas questões, para cuja resposta o estudo apresentado pretende contribuir: 1) Qual a presença da referência metalinguística no processo de escrita colaborativa analisado?; 2) Quais os mecanismos de referência metalinguística predominantes?; 3) Quais os domínios em que é ativada a referência metalinguística?; 4) Quais os termos mais frequentes?; 5) Existem diferenças significativas entre os anos de escolaridade?; 6) Existe relação com a extensão textual?; 7) Existem termos distintivos entre os anos de escolaridade considerados? Para obter respostas a estas questões, investigou-se a interação que teve lugar numa tarefa de escrita colaborativa no segundo, quarto, sexto e oitavo anos de escolaridade, na modalidade de coescrita referida (Lowry et al., 2004).

Na sua organização, além desta introdução, previamente à apresentação do estudo empírico, o artigo contempla o enquadramento teórico, focado no processo de escrita e na escrita colaborativa, na perspetiva da interação e da referência metalinguística ativada no processo. Em relação ao estudo, explicita-se a metodologia, considerando os participantes, os procedimentos e tarefa implicados e a análise realizada. Segue-se a apresentação dos resultados, focando-se, primeiramente, na presença da referência metalinguística, em termos gerais, e prosseguindo para indicadores mais específicos como a relação com a extensão textual, a distribuição pelos mecanismos de referência e domínios linguístico-textuais e a apreensão da saliência quantitativa e distintiva dos diversos termos, considerada em ligação aos anos de escolaridade. Na secção final, dedicada à discussão e conclusão, estabelece-se a relação dos resultados com as questões de investigação, com os apresentados por outros estudos, e perspetivam-se implicações para o ensino-aprendizagem.

2. Enquadramento concetual

O enquadramento concetual do estudo considera a complexidade do processo de escrita e a sua colocação em prática na modalidade de escrita colaborativa. A complexidade do processo manifesta-se quanto aos níveis linguístico-textuais envolvidos, à diversidade de componentes ou atividades que são chamadas a atuar e de operações que os escreventes realizam para construir a forma linguístico-textual que surgirá perante o leitor. A

modalidade de escrita colaborativa ativa a referência metalinguística para que a realização de operações e a tomada de decisões sejam feitas em conjunto.

2.1. **Processo de escrita: níveis linguístico-textuais, componentes e operações**

O processo constitui uma dimensão essencial da escrita. De facto, o processo de escrita dá origem a um resultado textual que era desconhecido antes de o processo ter lugar. Tal acontece quer no caso da produção escrita pela inteligência artificial (IA) generativa, quer no caso da escrita humana. Contudo, o contraste entre estas duas "fontes" de textos estende-se também à dimensão processual, evidenciando algumas particularidades do processo de escrita humano, que estão em relevo no presente artigo. No momento atual, esse contraste estabelece-se em relação aos níveis linguístico-textuais considerados no processamento, às componentes ativadas e às operações realizadas. As particularidades da escrita humana associadas a estes aspetos estão relacionadas com a possibilidade de o sujeito que escreve os poder processar de forma refletida, consciente, e de se poder referir explicitamente às unidades envolvidas nesse processamento.

Para evidenciar essas particularidades, considere-se a geração da escrita pela IA, perante quem observa o ecrã: o texto aparece pelo fluir de caracteres, formando palavras, em direção ao final do texto – é, de facto, essa a imagem associada à geração de texto pelos Large Language Models (LLM) no seu nível de desenvolvimento atual. A geração do texto, no nível em que tem lugar, é governada pela probabilidade, no âmbito do modelo de linguagem de grande escala, e situa-se num nível elementar da organização linguística, incidindo na (probabilidade da) presença de cada carácter ou sequências de caracteres correspondentes a palavras (Akhtar, 2024; Pereira & Moura, 2024).

No caso da escrita humana, o processo de escrita não emerge como um fluir contínuo de unidades processado ao nível das letras (caracteres) ou sequer das palavras. Em vez de ser governado pela probabilidade estatística, o avanço do texto escrito faz-se pelo processamento que coloca em relação diferentes níveis linguístico-textuais. Esse processamento envolve diversidade de componentes ou subprocessos e de operações que vão sendo feitas ao longo do processo, as quais refletem um percurso de decisões e escolhas (Barbeiro, 1999, 2001, 2019, 2023; Calil & Myhill, 2020; O' Donnell, 2013).

Os níveis linguístico-textuais sobre que podem incidir as operações e escolhas incluem: as unidades estruturais do texto; as unidades do discurso, que se situam num nível superior à oração (Martin & Rose, 2007; Rose & Martin, 2012), quer se trate de parágrafos ou de outras sequências discursivas e informacionais; as unidades oracionais; os sintagmas; as palavras e os elementos que as formam (Barbeiro, 2023). Além da diversidade de níveis, são ativados na escrita diferentes componentes ou subprocessos: planificação, redação ou textualização e revisão (Barbeiro, 1999, 2019; Berninger et al., 1996; Horning & Becker, 2006; Flower & Hayes, 1981; Keen, 2020). Para dar conta do processo de escrita, devemos conjugar a consideração dos níveis linguístico-textuais com as componentes. Essa conjugação faz emergir fases do processo de escrita, designadamente em relação ao texto tomado na sua globalidade. Assim, a planificação do texto, como unidade global, ocorre numa fase de pré-escrita. Nesta fase, antes de colocar em prática a textualização, quem escreve cria uma representação mental do resultado, que pode ser mais ou menos apoiada num trabalho prévio de organização e estruturação, de pesquisa e seleção de informação. A fase de escrita propriamente dita, tendo por referência o texto global, ativa a componente de redação ou colocação em texto escrito, linear, das unidades antes representadas apenas mentalmente ou, eventualmente, também apoiadas em anotações e representações gráficas, que podem assumir uma configuração não linear. A fase de pós-escrita torna saliente a componente de revisão que, por meio da releitura e avaliação da versão do texto já escrita, procede à verificação quanto a aspetos linguísticos, formais e de conteúdo.

A sequência processual apresentada, que, como referido, considera a unidade global do texto, não é absoluta. Devem ser tidas em conta duas salvaguardas que remetem para a complexidade que o processo de escrita humana pode apresentar. A primeira, que já ficou implícita, diz respeito ao facto de as componentes em causa (planificação, redação e revisão) poderem incidir sobre unidades situadas noutros níveis textuais que não

a globalidade do texto. Deste modo, capítulos, secções, ou outras passagens relativas a conteúdos específicos poderão ser objeto da sua própria planificação, textualização e revisão, de forma mais ou menos autonomizada. Tal poderá acontecer ainda em relação a unidades como parágrafos e períodos ou mesmo frases. A segunda é relativa ao facto de, em qualquer um dos níveis linguístico-textuais, o processo humano de escrita não consistir num fluir de caracteres e palavras, sem recuos, em direção ao final. Além de incidirem sobre diversidade de níveis e de colocarem em ação critérios associados a essa multiplicidade linguística e comunicacional de níveis, as operações e escolhas focadas numa passagem textual não ficam encerradas com o registo ou inscrição de uma unidade no suporte textual. A componente de revisão pode dar origem a correções e reformulações e à recursividade, ou seja, à reativação das componentes de planificação e de redação, em resultado da avaliação efetuada ou como exploração de novas possibilidades (Barbeiro, 2001, 2003; Tarchi et al., 2024).

Tendo como referência o texto já escrito, as operações manifestam-se pelas modificações realizadas. Estas modificações podem consistir em adições, substituições, deslocamentos ou mudanças de ordem dos elementos e supressões ou apagamentos. A amplitude destas operações também pode corresponder a diferentes níveis, pois as unidades envolvidas podem ser diversas: letras, palavras, sintagmas, frases, sequências discursivas, períodos, parágrafos, secções, capítulos, partes, etc. Mas, verdadeiramente, a realização de operações ou alterações para a construção do texto não necessita de esperar pela inscrição das unidades linguísticas no suporte de escrita, seja o papel, o ecrã, ou outro. Essas operações podem ter lugar mentalmente, antes de o escritor/escrevente as passar para o suporte (Barbeiro, 2003; Calil, 2017; Felipeto & Calil, 2021).

Na complexidade do processo, ressalta-se que há duas forças em interação. Uma das forças atua, em grande medida, de forma automatizada, inconsciente, ativando recursos (nomeadamente, conhecimentos prévios e os correspondentes termos e formulações linguísticas) e a execução de ações (inscrição ou transcrição das formulações linguísticas para o suporte de escrita, por meio do desenho das letras ou batimento das teclas, para fazer aparecer as palavras segundo determinada forma, que se pretende que corresponda à representação ortográfica). A segunda força atua de forma consciente, deliberada. A sua ação pode emergir em resultado de procura intencional da estruturação textual e da formulação linguística mais adequadas para construir o significado no texto, por relação com o seu contexto (Jewitt, 2006), ou em resultado de dúvidas e incertezas surgidas no decurso do processo, exigindo a resolução de problemas, tomadas de decisão refletidas e, eventualmente, novas procuras. O diferimento espacial e temporal que caracteriza a escrita, ou seja, o facto de o processo não ter lugar perante o leitor e em interação direta com ele, potencia a dimensão de procura e consideração de possibilidades para a decisão refletida quanto às soluções que serão adotadas no texto, quer em relação à estrutura do texto, quer em relação a formulações linguísticas específicas. Por outro lado, o próprio facto de, como referimos no início desta secção, o processo de escrita ir inscrevendo unidades e formulações linguísticas que eram desconhecidas antes de o processo ter lugar potencia o aparecimento de dúvidas e desencadeia novas possibilidades. Efetivamente, o texto que vai sendo inscrito passa a constituir um elemento da equação, isto é, emerge como um elemento que também deve ser considerado, pois configura as (novas) possibilidades quanto ao que escrever a seguir e, eventualmente, pode conduzir à sua própria reformulação.

Em suma, na sua complexidade, o processo de escrita, pelos níveis linguístico-textuais que envolve, pelo diferimento em relação à leitura, abre um campo alargado para a consideração de possibilidades e para tomada de decisão consciente. As operações e as escolhas processam-se na mente de quem escreve, por meio do seu pensamento interior, mas podem manifestar-se no suporte de escrita, como revelam os estudos dos rascunhos (Fabre, 1986; Fabre-Cols, 2002), ou os estudos dos batimentos nas teclas dos computadores (O'Donnell, 2013; Wengelin et al., 2019). Podem também ser verbalizadas, como revelam os estudos baseados nos protocolos que solicitam a quem escreve que vá pensando em voz alta (Gerova, 2022; Latif, 2021). As unidades linguísticas envolvidas nas operações, nos diferentes níveis, constituem unidades operatórias para o sujeito, sendo objeto de alterações e decisões, e, na verbalização, podem emergir no discurso, por meio da referência metalinguística.

2.2. **Escrita colaborativa: interação e referência metalinguística**

A colaboração constitui uma estratégia para a construção de um texto que é frequentemente posta em prática quer no contexto da escola, quer no contexto profissional. A colaboração pode ser concretizada segundo diversas modalidades (Brien & Brady, 2003; Lowry et al., 2004; Sharples et al., 1993), considerando o formato, o momento ou a componente em que a colaboração é efetivada. Quanto ao formato, Lowry et al. (2004) sistematizam quatro categorias de base: i) escrita colaborativa com redator único (*group single-author writing*): a tarefa de redigir o texto recai sobre um elemento do grupo; ii) escrita colaborativa sequencial (*sequential single writing*): cada elemento intervém dando continuidade ao produto escrito repartido por secções; iii) escrita colaborativa paralela (*parallel writing*): o produto planeado também é dividido em partes, cuja elaboração é repartida entre os membros do grupo, que trabalham em paralelo; e iv) escrita colaborativa reativa (*reactive writing*), que também podemos designar de *coescrita*: todo o processo de escrita do texto é realizado por meio da interação direta e síncrona entre os participantes. Com base nestes quatro formatos, Lowry et al. (2004) preveem a possibilidade de existência de formatos mistos.

Quanto ao momento do processo de escrita, a colaboração pode ser ativada em relação à totalidade do processo ou em relação a momentos específicos nos quais está em causa a resolução de problemas particulares. Estes problemas podem situar-se nos diferentes níveis linguístico-textuais, por exemplo, a organização do texto segundo a estrutura de determinado género ou a adoção de uma determinada perspetiva discursiva, questões que afetam o texto na sua globalidade, ou a resolução de uma dúvida de natureza ortográfica, cujo alcance se circunscreve a uma palavra particular.

No que se refere às componentes, a colaboração pode ser obtida em relação à conceção ou planificação, à redação ou textualização, e à revisão. A perspetiva das componentes também pode remeter para alguns formatos. Por exemplo, em relação ao texto na sua globalidade ou a partes específicas, é frequentemente posto em prática o formato sequencial, em que a colaboração é ativada para a tarefa de revisão, após a redação individual de uma primeira versão do texto. Ainda numa perspetiva sequencial, em vez de atuar no final, a colaboração pode ser chamada a contribuir no âmbito da planificação, para ajudar o autor a organizar a selecionar conteúdos e a estruturar o texto, seguindo-se a redação individualizada.

Uma dimensão que pode atravessar diferentes modalidades de colaboração consiste na interação. Esta constitui a força motriz da escrita reativa ou coescrita, mas pode estar presente ou ser chamada a complementar outros formatos. Na coescrita, é no contexto dessa interação direta que os participantes apresentam propostas e sugestões, reagem às propostas dos outros elementos do grupo, discutem e contrapõem argumentos, negoceiam e tomam decisões. O texto resultante é o produto das escolhas e decisões realizadas nessa interação no âmbito do grupo. Devido ao lugar que a interação ocupa na coescrita, foi esta a modalidade de colaboração adotada no presente estudo.

Ao incidir sobre a escrita, sobre as suas unidades e relações, o discurso da interação ativa a referência metalinguística, em contexto. Na situação que corresponde ao contexto imediato da produção escrita (Barbeiro, 2022, 2024), a referência metalinguística pode recorrer a diversos mecanismos (Barbeiro et al., 2022): à dêixis (*esta aqui não está bem*) e correferência (*aquela que ela disse é boa*), à utilização metalinguística das palavras (como, por exemplo, em *'Liz' está mal escrito*) e aos próprios termos metalinguísticos *(temos de fazer uma 'frase' comprida*). Na sequência de estudos anteriores, que exploraram indutivamente dados de interação na escrita colaborativa para estabelecer estes mecanismos (Barbeiro, 2012; Barbeiro et al., 2022), mantiveram-se estas categorias para realizar a análise, no presente estudo. Os estudos anteriores mostraram capacidade descritiva e analítica das categorias, sem sobreposição, mas admitindo combinação entre elas nos segmentos discursivos em que concretiza a referência metalinguística.

O conhecimento metalinguístico, com os respetivos termos faz parte da aprendizagem gramatical realizada na escola. Emerge, assim, a questão de saber em que medida a referência metalinguística que ocorre na interação da escrita colaborativa mobiliza o conhecimento gramatical, designadamente os termos que são alvo dessa aprendizagem. O presente artigo incide precisamente sobre os mecanismos de referência metalinguística que

são ativados no processo de escrita colaborativa, numa modalidade de coescrita, desempenhado por participantes de diferentes níveis de escolaridade e, consequentemente, de aprendizagem gramatical.

3. Metodologia

3.1. Participantes

Os participantes do estudo são alunos que frequentavam o ensino básico, no segundo, quarto, sexto e oitavo anos de escolaridade. Em cada ano, participaram 24 alunos, perfazendo um total de 96 alunos. As idades médias eram de 7;10, 9;10; 11,9 e 14;0, para os anos indicados, respetivamente. Em relação à distribuição por género, existe em cada ano um número igual de rapazes e de raparigas (12, perfazendo um total de 48 elementos de cada um dos géneros).

A amostra foi constituída em escolas da cidade de Leiria que os participantes frequentavam. Os seus locais de residência variavam entre a zona urbana e as freguesias da periferia da cidade.

3.2. Procedimentos e tarefa

Em cada ano, os participantes foram organizados em grupos para a realização da tarefa de escrita colaborativa. Esta consistiu na escrita de um relato correspondente a uma visita fictícia de três amigos a Leiria. A escolha da própria cidade dos participantes como o local da visita deveu-se à preocupação em assegurar que todos os elementos tinham um nível de conhecimento do conteúdo suscetível de ser integrado no texto que lhes permitisse apresentar propostas e contribuir para a seleção e tomada de decisões no grupo.

Nos diversos anos de escolaridade, cada grupo incluiu três elementos, que eram colegas da mesma turma. A constituição de grupos de três elementos teve como objetivo diversificar as possibilidades de interação, para além das relações entre elementos de um par, procurando minimizar a possibilidade de domínio do processo por um elemento, e sem demasiado risco de diluição da participação individual que poderia acontecer em grupos maiores. A adoção deste formato permitiu explorar o eventual contraste com formatos em que a escrita colaborativa é realizada por díades e cada elemento desempenha papéis específicos (por exemplo, ditante e escrevente, como ocorre em Barbeiro et al., 2022), A escolha dos elementos para cada grupo fez-se de forma aleatória, com algumas restrições, designadamente a de que todos os grupos deveriam ser mistos, ou seja, teriam dois elementos de um género e o terceiro do outro, e a de que, no final, deveria haver um número igual de participantes de cada género.

Em relação ao desempenho escolar, não se colocaram pré-condições ou exigências específicas, além da indicação do professor de que os participantes em causa apresentavam um nível de competência nas tarefas habituais de escrita que lhes permitia construir um texto. A amostra foi constituída em turmas de escolas do sistema público e também numa escola do sistema de ensino particular, em todos os anos de escolaridade. A inclusão dos dois sistemas procurou integrar no estudo a diversidade de contextos socioeducativos (neste artigo, contudo, não são colocados em foco eventuais contrastes entre os desempenhos dos alunos dos dois sistemas).

A recolha de dados para a realização do estudo foi autorizada pelas direções dos agrupamentos ou escolas. Com a constituição da amostra e explicação genérica da tarefa e do seu propósito, obteve-se a concordância dos alunos e recolheram-se autorizações dos respetivos encarregados de educação para a sua participação no estudo, incluindo a autorização para registo de áudio e vídeo, com objetivos de investigação.

Em relação aos procedimentos adotados para a concretização da atividade e da recolha de dados, antes da tarefa propriamente dita de produção escrita em grupo, o investigador conduziu uma entrevista em grupo sobre a relação dos alunos com a leitura e escrita, a fim de permitir a habituação à presença da câmara de vídeo e microfone. Ao passar-se para a tarefa de escrita colaborativa, o investigador distribuiu folhas individuais para que cada aluno registasse o texto do grupo e facultou uma folha com a indicação dos nomes dos protagonistas (João, Rui e Lena). Sobre estes, apenas era dito que eram três amigos que fizeram uma visita a Leiria. Como

apoio à componente de geração de conteúdo, foi também facultado um conjunto de seis postais ilustrados com imagens da cidade de Leiria, deixando-se claro que estes postais serviam apenas para ajudar e que os alunos poderiam escrever o texto como entendessem melhor. A instrução específica dada para a realização da tarefa consistiu em indicar que os alunos deveriam escrever um texto a contar a visita que estas personagens fizeram a Leiria. O texto deveria ser um texto único de todo o grupo e não textos individuais, para o que o grupo deveriadiscutir, colaborar e chegar a acordo. Era também indicado que cada elemento deveria escrever o texto na sua folha individual (não havendo a atribuição de papéis diferenciados, designadamente escrevente ou redator e ditantes). Deste modo, além de contributos para a elaboração textual, cada participante era confrontado com a tarefa de inscrição ou transcrição, no âmbito da redação, designadamente quanto à ortografia.

A atividade de cada grupo foi registada por meio de áudio, estando o microfone sobre a própria mesa de trabalho, para permitir uma boa captação da interação verbal, e por meio de vídeo, com a câmara mais afastada para captar todos os elementos do grupo. As atividades foram realizadas numa sala da escola, para a qual os alunos de cada grupo se deslocaram, para permitir a realização dos registos. A condução de cada atividade foi realizada pelo investigador. Além dos dados recolhidos por meio dos registos, procedeu-se à recolha das folhas individuais em que cada participante registou o texto.

Na sequência da atividade, os registos da interação foram transcritos. Na transcrição, foram indicados o código do participante (de S1 a S96, para os alunos) e aspetos contextuais e comportamentais. A transcrição foi realizada manualmente, primeiramente a partir dos registos áudio, seguida de verificação com base nos registos vídeo. Teve um caráter ortográfico, tendo sido marcadas as pausas e a entoação exclamativa ou interrogativa.

3.3. **Análise**

Os dados correspondentes ao discurso da interação entre os participantes durante a escrita colaborativa foram analisados com vista à identificação de ocorrências de referência metalinguística, em cada um dos anos. A análise das manifestações encontradas foi aprofundada, considerando o mecanismo de referência metalinguística posto em prática e o nível linguístico-textual da unidade em causa.

Em relação aos mecanismos de referência metalinguística, tendo em conta que a referência foi realizada num determinado contexto situacional, adotaram-se as categorias anteriormente mencionadas: a dêixis e correferência, a utilização metalinguística das palavras no discurso e o recurso aos termos metalinguísticos propriamente ditos.

No que diz respeito aos níveis e domínios linguístico-textuais, adotaram-se na análise as seguintes categorias: i) género, texto e discurso, agrupando os domínios superiores à frase; ii) gramática (integrando a fonologia, morfossintaxe, relações semântico-lexicais); e decorrentes da representação escrita, consideraram-se os domínios: iv) ortografia; v) pontuação; e vi) configuração gráfica.

A análise teve em conta a distribuição da frequência pelos anos de escolaridade integrados no estudo, procurando apreender contrastes entre esses anos, designadamente perspetivando uma eventual progressão com a aprendizagem escolar. Para a verificação da existência de diferenças significativas entre os anos de escolaridade, recorreu-se ao teste estatístico de análise de variância unidirecional, uma vez verificados os requisitos de normalidade e homogeneidade exigidos por este teste. Para determinar quais os grupos ou anos que apresentam diferenças estatísticas entre si, aplicaram-se os testes de comparações múltiplas (com recurso ao teste de Tukey, escolhido por o número de grupos não ser elevado e o tamanho das amostras ser igual, permitindo abranger todas as comparações par a par).

Em complemento da análise de variância, determinou-se o tamanho do efeito por meio do indicador de eta ao quadrado (η^2). O cálculo dos indicadores estatísticos foi realizado com recurso ao programa SPSS-29. Considerando as dinâmicas de interação variadas que podem existir no interior dos grupos, a análise foi realizada ao nível do grupo e não de cada elemento individual, o que se reflete nos graus de liberdade dos testes estatísticos.

Para a análise de contrastes com a progressão escolar, recorreu-se ainda a instrumentos da linguística de *corpus*, nomeadamente para determinar a existência de termos distintivos (*keywords*) entre os anos de escolaridade, no conjunto dos termos metalinguísticos utilizados. Uma *palavra distintiva* ou *keyword* corresponde a uma palavra com uma frequência significativamente mais elevada num *corpus* (o *corpus* alvo) face à sua ocorrência noutro (*corpus* de referência). Para realizar a análise, constituíram-se *corpora* com os termos metalinguísticos de cada ano de escolaridade. A procura de termos distintivos foi efetuada com recurso ao programa *Antconc* (Anthony, 2024), que, além dos valores de frequência absoluta, no *corpus*, disponibiliza os valores de frequência normalizada (cálculo do número de ocorrências por milhão de palavras) e os valores do teste estatístico relativo à "distintividade" (*keyness*).

4. Resultados

A apresentação dos resultados incidirá sobre as três vertentes consideradas: a referência metalinguística, quer numa perspetiva global, quer considerando os mecanismos de referência metalinguística; os domínios ou níveis linguístico-textuais em que se situam as unidades; e os termos metalinguísticos que emergem no discurso. Em cada vertente, será tida em conta a perspetiva de contraste entre os níveis de escolaridade em análise no estudo.

4.1. A referência metalinguística

Esta primeira subsecção foca-se na presença da referência metalinguística e nos mecanismos utilizados pelos participantes para efetuar a referência às unidades linguísticas ao longo da interação na escrita colaborativa. Na Tabela 1, são apresentados os valores das ocorrências de referência metalinguística, para cada ano de escolaridade.

Tabela 1

Referência Metalinguística

Níveis de escolaridade	N.º	Média	Desv. padrão
2.º	851	106,4	78,9
4.º	968	121,0	48,5
6.º	997	124,6	55,3
8.º	958	119,8	55,1
Global	3774	117,9	58,0

Os valores apresentados na Tabela 1 mostram que a referência metalinguística está amplamente presente no processo de escrita colaborativa analisado, em todos os níveis de escolaridade. A análise estatística (análise de variância) não revela diferenças significativas entre as médias dos diversos anos de escolaridade ($F_{3,28}$=0,139, p=0,936).

Os resultados apresentados na Tabela 1 e objeto da análise estatística referida no parágrafo anterior não têm em conta o facto de os textos produzidos apresentarem uma extensão crescente com a progressão nos anos de escolaridade (com médias de 81,1, 156,5, 177,8 e 237,9 palavras, respetivamente para os anos indicados). Estas diferenças surgem como significativas ($F_{3,28}$=9,496, p=0,000) e o tamanho do efeito associado a esta variável relativa ao nível de escolaridade é muito grande ($\eta 2$= 0,504). Em termos estatísticos, aproximadamente, 50,4% da variância total da variável dependente, o número de palavras, pode ser explicada por esta variável independente. Na Tabela 2, apresentam-se os resultados da relação entre a extensão textual e a referência metalinguística, por meio do respetivo rácio (n.º de referências por palavra).

Tabela 2

Relação com a Extensão Textual

Níveis de escolaridade	Extensão (média)	Ref.ª metal.	Rácio (média)	Desv. Padrão
2.º	81,1	851	1,31	0,69
4.º	156,5	968	0,79	0,23
6.º	177,8	997	0,71	0,28
8.º	237,9	958	0,53	0,20
Global	163,3	3774	0,83	0,48

Observa-se no rácio apresentado na Tabela 2 um decréscimo dos valores com o aumento do nível de escolaridade, com diferença estatisticamente significativa entre o segundo e o oitavo anos ($F_{3,28}$=5,374, p=0,005, η2= 0,365); a magnitude do efeito permanece muito grande (36,5% de variância explicada). Por outras palavras, o recurso à referência metalinguística não acompanha proporcionalmente o aumento da extensão textual, ou seja, a sua ativação passa a não ser requerida num grau tão elevado.

O passo seguinte é respeitante à distribuição das ocorrências de referência metalinguística pelos mecanismos por meio dos quais ela é construída. Na Tabela 3, são apresentados os resultados das três categorias estabelecidas (dêixis/correferência, utilização metalinguística das unidades e recurso a termos metalinguísticos).

Tabela 3

Frequência dos Mecanismos de Referência Metalinguística

Mecanismos	Níveis de escolaridade				Total
	2.º	4.º	6.º	8.º	
Dêixis e correferência	84	78	92	106	360
	11%	*8%*	*10%*	*11%*	*10%*
Utilização metalinguística	198	171	190	240	799
	25%	*19%*	*18%*	*26%*	*22%*
Termos metalinguísticos	569	719	715	612	2615
	64%	*73%*	*72%*	*63%*	*68%*

A Tabela 3 mostra que as três categorias consideradas estão presentes em todos os níveis de escolaridade. A frequência de cada categoria nos diferentes anos de escolaridade apresenta valores totais (e médios) e também proporcionais próximos. Efetivamente, a análise estatística confirma a inexistência de contrastes significativos entre os anos de escolaridade (resultados da análise de variância unidirecional: $F_{3,28}$=0,203, p=0,893, para a referência metalinguística por meio de dêixis ou correferência; $F_{3,28}$=0,485, p=0,695, para os da utilização metalinguística das unidades; $F_{3,28}$=0,378, p=0,770, para os valores médios de termos metalinguísticos).

Considerando as proporções, os resultados evidenciam que o uso de termos metalinguísticos é muito significativo e predominante, em todos os níveis de escolaridade. Importa, por conseguinte, aprofundar os resultados em relação aos níveis linguístico-textuais a que esses termos estão ligados e em relação aos próprios termos metalinguísticos que os participantes utilizam na interação colaborativa. Esse aprofundamento será realizado nas subsecções seguintes.

4.2. **Níveis e domínios linguístico-textuais dos termos metalinguísticos**

A Tabela 4 apresenta a distribuição dos resultados relativos aos termos metalinguísticos pelos níveis e domínios considerados (não foram integrados na Tabela duas ocorrências, pois não dizem respeito ao texto em construção, mas ao ritmo e volume de voz na interação entre os participantes, ou seja, ao domínio da Elocução).

Tabela 4

Níveis e domínios linguístico-textuais dos termos metalinguísticos

Níveis e domínios	Níveis de escolaridade				Total
	2.º	4.º	6.º	8.º	
Género, texto e discurso	24	48	88	93	253
	4%	*7%*	*12%*	*15%*	*10%*
Gramática	32	33	34	32	131
	6%	*5%*	*6%*	*5%*	*5%*
Ortografia	211	126	88	82	507
	37%	*18%*	*12%*	*13%*	*19%*
Pontuação	204	388	416	333	1341
	36%	*54%*	*58%*	*54%*	*51%*
Configuração gráfica	97	123	89	72	381
	17%	*17%*	*12%*	*12%*	*15%*

Os resultados revelam a saliência do domínio da pontuação, que é preponderante a partir do quarto ano de escolaridade. No segundo ano, também apresenta valor saliente, mas ligeiramente inferior ao domínio da ortografia. Em relação a este domínio, observa-se um claro decréscimo a partir do quarto ano. Em contrapartida, além da pontuação, aumentam os valores relativos à frequência e ao peso proporcional dos termos ligados aos domínios do nível superior à oração e à frase, ou seja, termos ligados aos domínios do género, do texto e do discurso. Os termos relativos ao processamento da configuração gráfica estão presentes em todos os anos de escolaridade, segundo valores aproximados, designadamente proporcionais. A gramática (morfologia e sintaxe) apresenta a menor saliência global, com resultados próximos nos diferentes anos, quer quanto às médias de ocorrências, quer quanto aos valores proporcionais.

Neste panorama de resultados, a significância estatística dos contrastes entre os anos de escolaridade apenas é alcançada em dois casos. Um dos casos diz respeito ao nível correspondente ao género, texto e discurso, em que os valores médios de ocorrências não atingem o limiar de significância ($F_{3,28}$=2,462, p=0,083), mas, quando se consideram os valores proporcionais, o contraste significativo já emerge, dando conta da proporção crescente entre o segundo e o oitavo anos ($F_{3,28}$=4,470, p=0,011, $\eta 2$= 0,324), com um tamanho do efeito que alcança um valor muito grande (32,4% da variância explicada). O segundo caso é relativo ao domínio da Ortografia: os valores médios de ocorrências não apresentam diferenças significativas ($F_{3,28}$=1,771, p=0,176), mas a redução do peso proporcional entre o segundo e o oitavo anos surge como estatisticamente significativa ($F_{3,28}$=3,803, p=0,021), sendo a diferença entre estes dois níveis considerada significativa pelo teste de comparações múltiplas de Tukey.

Além dos números, de que a Tabela 4 nos apresenta o panorama global e relativo a cada ano de escolaridade, quanto à presença de termos metalinguísticos no discurso da interação, interessa dirigir o olhar para o processamento em que se integra a ativação desses termos. De seguida, ilustrar-se-á o processamento que está em causa com alguns exemplos para cada domínio, que, no entanto, não esgotam o tipo de processamento que é feito. Nos extratos de interação apresentados, assinalam-se as propostas de formulação linguística para construir o texto por meio de itálico e os termos metalinguísticos em foco no domínio considerado por meio de negrito; outros termos e enunciados metalinguísticos que também surjam são assinalados com recurso a sublinhado.

Os extratos (1), do 2.º ano, e (2), do 6.º ano, são relativos ao domínio da pontuação. Os dois exemplos apresentados ilustram como a consideração de alternativas é um dos aspetos que emerge no processamento da pontuação a inscrever no texto. Nestes casos, constitui objeto de escolha a possibilidade de utilização de ponto final ou de dois pontos ou ainda de ponto e vírgula. Surgem em foco o tratamento linguístico-discursivo a dar à enumeração (1) e à explicitação (2), com reflexos na própria pontuação e na extensão da frase. Sublinhe-se o papel da modelação que o trabalho em aula representa, como se encontra em (1), pela referência às composições

feitas na sala, e a hesitação ou dúvida quanto às possibilidades existentes para construir a pontuação de uma passagem do texto (neste caso, em (2), envolvendo o ponto final, os dois pontos e o ponto e vírgula), o que ativa o processamento consciente e o recurso à referência metalinguística. Os extratos evidenciam ainda a recorrência dos termos na interação, decorrente de dúvidas, consideração de alternativas e confirmação entre os elementos do grupo, o que contribui para explicar a elevada frequência do domínio da pontuação que surge na Tabela 4.

(1)

S13: *foram ó castelo de Leiria e viram*
S14: *lá muitas coisas*/ **dois pontinhos**
S13: **ponto**/ e depois
S14: não/ não/ **dois pontinhos**/ e agora vais tu dizer o que eles viram// não te lembras que é como nas composições que nós fazemos lá em cima na sala?
(G05, 2.º ano)

(2)

S50: *que era/ deu uma sugestão*/ **dois pontos**/ *que era ir a um arquivo/ que era*
S49: *que era*/ deixa-me cá ver/ está bem
S50: *que era*
S51: mas *sugestão* não leva **ponto**/ se continuamos a frase
S49: *sugestão* leva **dois pon.../ dois pontos** não/ **ponto e vírgula**/ leva **ponto e vírgula**/ olha
S51: *deu uma sugestão*
S50: *que era ir a um arquivo*
S49: não sei se realmente leva **ponto e vírgula**/ hã?
(G17, 6.º ano)

Os extratos (3) e (4), do 6.º ano, e o (5), do 8.º ano, incidem sobre a ortografia. Mostram que no processo de escrita colaborativa os participantes efetuam a verificação da norma ortográfica entre os diversos elementos, como acontece em (3), e que procuram resolver os problemas e dúvidas que emergem com a ativação (ou a reminiscência) de regras acedidas na aprendizagem escolar, como se pode considerar que ocorre em (4), com a tentativa de justificação por parte de S57. A aprendizagem gramatical escolar emerge ainda em (4), por meio do termo *homónima*. Estes extratos mostram ainda, para além da emergência do problema ou dúvida, a consciência ou descoberta da natureza convencional da norma ortográfica, neste caso, associada à sua mudança ao longo do tempo e à coexistência de formas num determinado momento, como acontece com as formas *Lis* e *Liz* (que ainda era visível no nome de um estabelecimento que aparecia numa das imagens) ou *Dinis*/*Diniz,* caso referido adicionalmente por S85 (5).

(3)

S68: *um lugar para nos hospedarmos*/ mal escrito {indica a palavra "hospedarnos" na folha de S69}
S67: leva um **agá**/ leva um **agá**
S68: *hospedarmos*
S67: é com **agá**
(G23, 6.º ano)

(4)

S57: *uma gaivota* não é com **letra grande**/ é com **letra pequena**/ porque gaivota não é pássaro nem objeto/ não é aí {refere-se ao texto de S56}

S56: uma palavra homónima/ gaivota
(G19, 6.º ano)

(5)

S85: *do rio Lis*
S86: eh/ pá!/ eu não sei se é com **esse** ou com **zê**
S85: é com **esse**
S85: é com/ um
S87: olhem! mas pode ser com **esse** ou **zê**
S85: mas pode/ pois pode/ pois pode!
S86: pois é
S85: Dom Dinis também é com **zê** ou com **esse**
(G29, 8.º ano)

Os excertos (6) e (7) colocam em foco o termo metalinguístico *frase*, integrado no domínio gramatical. Nestes excertos, do 2.º e 8.º anos, revela-se a plasticidade concetual que lhe corresponde e a necessidade da contextualização dos critérios nas escolhas discursivas de cada texto. Enquanto em (6), no excerto do 2.º ano, a argumentação é orientada para a escrita de frases grandes, em (7), no 8.º ano, a argumentação é já orientada no sentido contrário.

(6)

S3 *viram/ viram* o quê?
S2: *o Turismo*
S3: o
S1: *Tu.../* tão pequeninas as **frases**!
(…)
S1: *eles viram o centro comercial Dom Dinis*
S3: *eles*
S2: tem que ser bem grande!
S1: pois
S2: é sempre pequeninas!
S1: oh!/ atão calhou!/ olha!/ faz!
(G01, 2.º ano)

(7)

S77 *como eram curiosos foram ver o que era*
S76: *como eram?!* {tom depreciativo}
S77: sim/ porque
S76: ó pá!/ *curiosos*/ pomos só *curiosos foram ver*
(…)
S77: *foram ver*
S76: *curiosos foram ver*// ponto outra vez?/ não!
S77: sim/ não faz mal!
(…)
S77: deve-se fazer **frases** pequeninas
(G26, 8.º ano)

Os excertos (8) e (9) são também retirados da interação dos 2.º e 8.º anos, respetivamente, e mostram que, além da presença quantitativa evidenciada pelos valores da Tabela 4, a procura e atribuição de valor semiótico à configuração gráfica têm início desde cedo, no percurso escolar, como indica o exemplo (8) do 2.º ano de escolaridade.

(8)

S6 deixa-se uma **linha**/ hã?
S4: deixa/ deixamos uma **linha** e
S5: eu não deixo
S6: então não deixes!
S4: é/ porque a gente vamos começar outra vez/ era uma vez não sei quê/ não sei quê mais
(…)
S4: depois não se percebe/ olha/ como é que depois eles vão ver que é uma história/ duas e três?
S2: sim/ não faz mal!
(…)
S5: isto é só uma história
(G02, 2.º ano)

(9)

S80: *di/ri/gi/mo-nos para lá*/ o **á** aqui é/ é arredondado para dar efeito de exclamação
(G27, 8.º ano)

O aparecimento na interação dos termos metalinguísticos considerados no nível que integra os domínios relativos ao género, texto e discurso é ilustrado nos excertos (10), do 2.º ano, (11) e (12) do 8.º ano. Como se pode observar em (10), os traços linguístico-discursivos de alguns géneros, designadamente da narrativa, emergem já no 2.º ano.

(10)

S16 primeiro é ***composição*** {título}
S17: ***composição***/ ***composição***
(…)
agora o que dizemos?
S16: *de Leiria*
S18: u*m dia*
S16: vamos começar *eles foram ao castelo* (…)
S18: *um dia!*
S16: não/ *eles foram*
S17: *um dia*// começamos sempre pelo *um dia*
(G06, 2.º ano)

(11)

S80 como é que começamos?/ é verdade!/ temos que fazer **composição**/ temos que fazer/ temos que fazer uma **introdução**
(G29, 8.º ano)

(12)

S91: acabou a **apresentação**!
S92: sim/ **apresentação** já tá/ agora vamos começar/ isto é mais tipo dum **diário**/ tamos a contar uma **história**
(G31, 8.º ano)

4.3. **Os termos metalinguísticos**

Nas subsecções anteriores, foram apresentados os valores globais e por domínios da ocorrência de termos linguísticos. Nos excertos ilustrativos, alguns dos termos particulares surgiram integrados na interação da escrita colaborativa. A presente subsecção focar-se-á nos próprios termos metalinguísticos. Na globalidade, ocorreram 175 termos metalinguísticos diferentes. Na Tabela 5, são apresentados os 20 termos metalinguísticos mais frequentes em cada ano de escolaridade.

Tabela 5

Termos Metalinguísticos mais Frequentes em Cada Ano de Escolaridade

	Níveis de escolaridade							
	2.º		**4.º**		**6.º**		**8.º**	
	Termos	**N.º**	**Termos**	**N.º**	**Termos**	**N.º**	**Termos**	**N.º**
1	ponto final	85	ponto	116	ponto	159	ponto	143
2	ponto	47	vírgula	101	vírgula	125	vírgula	118
3	vírgula	40	parágrafo	52	parágrafo	61	parágrafo	55
4	parágrafo	38	dois pontos	51	ponto final	60	ideia	16
5	linha	37	linha	36	linha	20	esse	15
6	letra maiúscula	18	ponto e vírgula	27	ideia	19	dois pontos	13
7	letra grande	17	ponto de exclamação	21	travessão	18	história	13
8	nome	15	frase	20	nome	17	zê	13
9	parênteses	14	ponto final	19	dois pontos	13	ponto final	12
10	sê	14	letra maiúscula	18	narrador	13	travessão	12
11	á	14	letra grande	17	composição	11	ó	12
12	agá	13	parênteses	15	letra grande	11	palavra	11
13	letra	13	travessão	15	parênteses	11	reticências	10
14	acento	10	zê	15	pegado	11	texto	10
15	dois pontinhos	10	letra	14	frase	10	nome	9
16	frase	10	esse	10	aspas	9	diálogo	7
17	história	10	agá	9	tracinho	9	fim	7
18	vê	10	ideia	9	zê	7	aspas	6
19	zê	10	letra pequena	9	diálogo	6	frase	6
20	letra pequena	9	composição	7	palavra	6	ponto de exclamação	6
…	…	…	…	…	…	…	…	…

A observação da Tabela 5 permite ressaltar alguns aspetos: i) as primeiras posições são ocupadas por termos ligados à pontuação, designadamente os termos *ponto (final)* e *vírgula*; ii) ocorrem variações terminológicas, designadamente *ponto* e *ponto final*, *letra maiúscula* e *letra grande*, e a variação por meio do diminutivo (*dois pontinhos*); iv) os nomes de letras estão presentes em todos os anos de escolaridade, indiciando que emerge a necessidade de resolução de problemas de natureza ortográfica, sobretudo, ou também relativa ao desenho das letras; v) *frase* constitui uma unidade que emerge em todos os anos de escolaridade, nestas posições de frequência, o que remete para a sua função operatória no processo, desde cedo; vi) o domínio relativo à pontuação, bastante marcante, apresenta, nestas posições mais frequentes, um alargamento de termos, a partir do quarto ano de escolaridade (*ponto e vírgula*, *ponto de exclamação*, *travessão*, *reticências*); vii) os termos

relativos ao género, texto e discurso apresentam também um alargamento quando se considera a progressão na escolaridade (de *história* e *nome*, que se refere à necessidade de colocar os nomes dos protagonistas, no segundo ano, passa-se ao aparecimento também de *ideia*, *composição*, *narrador*, *diálogo*, *texto*, *fim*, nos outros anos).

A evolução, para que os resultados da Tabela 5 apontam, quanto aos termos metalinguísticos, faz surgir a questão da existência de termos distintivos, ou seja, termos com uma frequência estatisticamente elevada num *corpus* face às ocorrências noutro *corpus*, tomado como referência. Considerando o contraste, que já emergiu por diversas vezes, entre o segundo ano e os restantes anos de escolaridade considerados, procurou-se apreender se esse contraste se manifestava por meio da existência de termos metalinguísticos estatisticamente distintivos. A Tabela 6 apresenta os resultados da análise estatística quanto à sua distintividade (*keyness*), entre os termos do segundo ano, tomados como *corpus* de referência, e os termos do quarto ano, tomados como *corpus* alvo.

Tabela 6

Termos Metalinguísticos Distintivos entre o Segundo e o Quarto Anos

Termos distintivos (*Keywords*)	Freq. ref.	Freq. alvo	Freq. norm. ref.	Freq. norm. alvo	*Keyness (Likelihood)*
dois pontos	4	51	7029,877	70931,850	38,682
ponto e vírgula	0	27	0	37552,156	31,938
ponto de exclamação	0	21	0	29207,232	24,760
ponto	47	116	82601,054	161335,188	18,479
travessão	0	15	0	20862,309	17,629
vírgula	40	101	70298,770	140472,879	16,686

Como se pode observar na tabela, todos os termos pertencem ao domínio da pontuação. O caráter distintivo advém do aparecimento de novos termos neste domínio entre o segundo e o quarto anos (*ponto e vírgula*, *ponto de exclamação* e *travessão*) e da maior frequência absoluta e normalizada (valores por milhão de palavras) alcançadas no quarto ano por outros termos (*dois pontos*, *ponto*, *vírgula*). O reforço da frequência destes termos revela que, longe de se tornar automatizada na construção colaborativa do texto, no quarto ano, a pontuação requer o processamento consciente e intencional por parte dos elementos do grupo na escrita conjunta.

A análise dos termos distintivos foi alargada aos *corpora* específicos de termos metalinguísticos do segundo e oitavo anos. A Tabela 7 apresenta os resultados.

Tabela 7

Termos Metalinguísticos Distintivos entre o Segundo e o Oitavo Anos

Termos distintivos (*Keywords*)	Freq. ref.	Freq. alvo	Freq. norm. ref.	Freq. norm. alvo	*Keyness (Likelihood)*
ponto	47	143	82601,054	233660,131	52.084
vírgula	40	118	70298,770	192810,458	39.878
travessão	0	12	0	19607,843	15.892
texto	0	12	0	16339,869	13.227

No contraste entre o segundo e o oitavo anos, os termos distintivos continuam a situar-se em larga medida no domínio da pontuação. A forte frequência dos termos *ponto* e *vírgula* no oitavo ano confirma que este domínio continua a requerer processamento consciente na coescrita de um texto. Ainda neste domínio, continua a emergir o aparecimento de novos termos, como *travessão* (que permanece com uma frequência contrastante em relação ao segundo ano). Além dos termos ligados à pontuação, o contraste distintivo alarga-se ao nível superior à frase, por meio do termo *texto*.

5. Discussão e conclusão

O estudo apresentado incidiu sobre a presença da referência metalinguística na interação da escrita colaborativa. Em resposta à questão genérica 1), os resultados da análise efetuada confirmam que a referência esteve amplamente presente na interação ocorrida na tarefa de escrita colaborativa posta em prática, que se caracterizou por ter um caráter de coescrita ou escrita reativa, na designação de Lowry et al. (2004). Ressalta-se também nos resultados, relativamente à questão 2), que o uso de termos metalinguísticos constituiu o mecanismo a que os participantes recorreram de forma preponderante. Mesmo sendo possível o recurso à dêixis e correferência, pelo facto de os elementos do grupo partilharem o mesmo contexto de situação, e à utilização metalinguística ou autorreferencial das unidades em foco, foi o uso de termos metalinguísticos que predominou, em todos os níveis de escolaridade, desde o segundo ao oitavo ano.

No contexto da interação orientada para a tarefa de escrita conjunta e participada do texto, os termos metalinguísticos ativados adquiriram funcionalidade ao serviço dessa tarefa. O que emergiu de forma saliente não foi a descrição gramatical das propriedades e relações das unidades em causa, mas o recurso aos termos para operacionalizar e conduzir as atividades e proceder à tomada de decisões e realização de escolhas que a escrita implica. A explicitação de propriedades e relações emergiu, sobretudo, quando constituiu argumento para essa decisão, como é caso da regra ortográfica de escrita com maiúscula ou das regras associadas à utilização de sinais de pontuação particulares. De qualquer modo, as unidades linguísticas estão como objetos perante os sujeitos e o conhecimento sobre elas pode ser ativado, para além da funcionalidade restrita, como foi ilustrado no excerto (4), em que um aluno, perante os diversos significados de *gaivota* refere a relação de homonímia, embora sem aprofundar a análise e a explicitação do seu conhecimento.

O alargamento e aprofundamento do conhecimento gramatical não se encontra na primeira linha, na sequência da ativação dos termos metalinguísticos na escrita colaborativa realizada pelos alunos. Concluindo-se que nessa ativação predomina a dimensão funcional, deve ser assinalado que existe na tarefa e na interação o substrato para esse aprofundamento. A modalização da tarefa com outro participante, o professor, pode reforçar a presença e aprofundamento do conhecimento metalinguístico. De facto, os resultados da análise da tarefa de escrita colaborativa com participação e condução do professor, como a realizada em Barbeiro e Barbeiro (2023), mostram não só que são ativados os termos metalinguísticos ligados à geração de propostas e realização de escolhas, mas que o professor procura (re)ativar o conhecimento gramatical associado a essas propostas e decisões, sobretudo quando foram objeto de ensino-aprendizagem, em aulas anteriores. Nalguns casos, além da reativação dos termos e do conhecimento em si, o professor procura evidenciar o valor ou efeito para a construção do significado associando-o à explicitação terminológica do recurso (Barbeiro & Barbeiro, 2023).

Em relação à questão 3), focada nos níveis linguístico-textuais em que é ativada a referência metalinguística e, de modo particular, os termos metalinguísticos, os resultados mostram que estes se situam nos diversos domínios: o nível superior à frase (incluindo termos ligados ao género, texto e discurso), o domínio gramatical e os domínios respeitantes à representação escrita (ortografia, pontuação e configuração gráfica). Globalmente, o domínio que se salienta é o da pontuação.

A saliência quantitativa do domínio da pontuação, confirmada também em resposta à questão 4), vai além da encontrada noutros estudos, designadamente em Barbeiro et al. (2022), que também analisou quantitativamente a presença de termos metalinguísticos na interação da escrita colaborativa. Neste estudo, o domínio da pontuação também recebe um número relevante de ocorrências de termos metalinguísticos, mas fica atrás do domínio da ortografia. A tentativa de explicação para a diferença nos resultados poderia ser procurada no facto de em Barbeiro et al. (2022) os níveis de escolaridade incluídos no estudo se limitarem ao 2.º e 4.º ano, níveis em que o domínio ortográfico está em foco. Contudo, no presente estudo, mesmo nestes anos, o domínio da pontuação ou está muito próximo dos resultados da ortografia (2.º ano, com proporções de 36% face a 37%, respetivamente para a pontuação e ortografia) ou os suplanta claramente (54% e 18%, respetivamente). Uma parte relevante da explicação poderá residir num outro fator: o formato ou modalidade da escrita colaborativa em cada um dos estudos. Em Barbeiro et al. (2022), a tarefa de escrita é realizada em

pares e apenas um dos participantes recebe a função de redator. No presente estudo, como deixámos expresso, a tarefa é realizada em grupos de três elementos e todos devem redigir o texto comum numa folha individual. Para que o texto seja comum, incluindo as decisões sobre a pontuação, o desempenho da tarefa mostra que é necessário que os alunos a tornem manifesta ou explicitem, para os três elementos. Concomitantemente, a verificação da igualdade da pontuação vai sendo verificada e controlada no decurso da tarefa, dando também origem à apresentação de argumentação (como os excertos (1) e (2) permitiram observar). A diferença encontrada faz, assim, ressaltar a relevância da modalidade de colaboração adotada em cada tarefa.

Num patamar de frequência bastante abaixo da pontuação, encontram-se as ocorrências de termos associados aos domínios da ortografia, da configuração gráfica e do género, texto e discurso. Num patamar ainda mais baixo, surgem as ocorrências de termos do domínio da gramática.

Uma outra questão que esteve em foco no estudo, questão 5), diz respeito à existência de diferenças significativas entre os níveis ou anos de escolaridade considerados. Em relação aos indicadores em valores absolutos, nomeadamente os valores médios da referência metalinguística, assim como os relativos à distribuição pelos mecanismos de referência e pelos domínios linguístico-textuais, as diferenças encontradas não são estatisticamente significativas. Os contrastes significativos surgem em função de dois parâmetros: a relação com a extensão textual e a distribuição proporcional pelos domínios. Efetivamente, a extensão média dos textos produzidos nos diversos anos não é homogénea, mas apresenta uma progressão crescente significativa. Quando se considera o rácio correspondente à relação da referência metalinguística com a extensão textual, em resposta também à questão 6), os valores passam a apresentar uma orientação decrescente, que se mostra significativa entre os anos mais afastados, ou seja, entre o segundo e o oitavo. Esta orientação decrescente da relação com a dimensão textual indica que a referência metalinguística na interação não acompanha proporcionalmente o aumento de palavras dos textos produzidos. Uma vez que a ativação da referência metalinguística corresponde a um processamento consciente e deliberado do aspeto em causa, pode considerar-se que a diminuição do rácio reflete um incremento da automatização que caracteriza a evolução do processo de escrita ao longo do ensino básico (Barbeiro, 1999; Cerni & Job, 2024; Graves, 1983, 1984; Kim, 2020; Van Reybroeck et al., 2017). Os termos distintivos entre os anos de escolaridade, colocados em foco pela questão 7), realçam também a saliência quantitativa dos termos do domínio da pontuação. Em relação aos pesos proporcionais dos domínios, emerge o peso crescente dos termos metalinguísticos de nível suprafrásico, relativos ao género, texto e discurso, e a diminuição do peso da ortografia. Os resultados evidenciam, assim, o aumento do foco do processamento em aspetos discursivo-textuais, com o avançar da escolaridade. Confirmam também que o processamento ortográfico evolui no sentido de requerer, num grau cada vez menor, a ação consciente no processo de escrita.

Os resultados do presente estudo e as conclusões que sustentam podem ser perspetivados no sentido das implicações para o ensino-aprendizagem. A implicação imediata diz respeito à confirmação do potencial da escrita colaborativa e da interação que nela é ativada para o desenvolvimento da competência de escrita. No processo colaborativo, os participantes podem contar com o contributo dos outros elementos do grupo para a resolução consciente de problemas nos diversos níveis linguístico-textuais, para o alargamento das possibilidades consideradas na construção textual, para a tomada de decisão ou escolha entre possibilidades e para a ativação e partilha de conhecimento por meio do qual se procura fundamentar essas escolhas.

Uma parte do conhecimento que é ativado estabelece a relação entre a competência de escrita e o conhecimento metalinguístico, podendo favorecer a aprendizagem nas duas vertentes. Na vertente do conhecimento gramatical, no âmbito da interação da escrita colaborativa, os termos metalinguísticos e o conhecimento gramatical em geral não ficam encerrados na esfera gramatical (Sousa, 1993), mas adquirem um valor funcional, no processo de escrita e das operações e escolhas que envolve. Esta relação funcional entre o conhecimento gramatical e a escrita encontra eco noutros estudos, nomeadamente os que têm vindo a ser desenvolvidos por D. Myhill e colegas (Myhill & Jones, 2015, Myhill & Newman, 2016; Myhill et al., 2016, 2020). Deve, no entanto, ter-se presente que, na tarefa de coescrita entre os alunos que foi posta em prática, o alcance quanto à explicitação e aprofundamento do conhecimento gramatical foi limitado, nomeadamente em

comparação com formatos de tarefas de escrita colaborativa em que o professor constitui um dos participantes (Barbeiro & Barbeiro, 2023).

Como, numa perspetiva mais geral, o confronto dos resultados com os de outros estudos mostrou que o formato da tarefa de escrita colaborativa é relevante para o processamento realizado e para as potencialidades quanto à ativação e explicitação do conhecimento, incluindo o gramatical, podemos pensar em possibilidades de reconfiguração da tarefa para potenciar essa explicitação e aprofundamento. Por exemplo, tendo como base a tarefa de coescrita posta em prática neste estudo, alargando-a, pode solicitar-se que os alunos registem escolhas que estiveram em causa no decurso do processo. No final, com a partilha dos textos efetuada para a turma, os alunos poderão ser chamados a partilhar também algumas decisões que estiveram em discussão. No âmbito dessa partilha, poderá alargar-se e aprofundar-se a reflexão, orientando-a para a discussão de cada uma das possibilidades e das suas potencialidades específicas para a construção do significado. Sob a orientação e condução do professor, essa discussão pode ser feita em associação à consciência metalinguística e mobilizando o conhecimento gramatical.

Em relação ao desenvolvimento do estudo, há uma dimensão de aprofundamento que se justifica. De facto, a amplitude do presente estudo incluiu os diversos domínios. Importa analisar, de forma específica, o processamento que é feito em relação a cada um dos níveis e domínios linguístico-textuais, incluindo o género, o texto, o discurso, a gramática e os domínios da representação escrita, ortografia, pontuação e configuração gráfica.

Referências

Akhtar, Z. (2024). Unveiling the evolution of generative AI (GAI): A comprehensive and investigative analysis toward LLM models (2021–2024) and beyond. *Journal of Electrical Systems and Information Technology*, *11*, 22. https://doi.org/10.1186/s43067-024-00145-1

Anthony, L. (2024). AntConc (Versão 4.3.1) [Software Informático]. Waseda University. https://www.laurenceanthony.net/software.html

Barbeiro, L. (1999). *Os alunos e a expressão escrita*: *Consciência metalinguística e expressão escrita.* F. C. Gulbenkian.

Barbeiro, L. (2001). Profundidade do processo de escrita. *Educação & Comunicação*, *5*, 64–76.

Barbeiro, L. (2003). *Escrita: Construir a aprendizagem*. Instituto de Educação e Psicologia da Universidade do Minho.

Barbeiro, L. (2012). Conhecimento sobre a língua e criatividade. In J. A. B. Carvalho, L. F. Barbeiro, L. Á. Pereira, & A. Silva (Orgs.), *Aula de língua: Interação e reflexão* (pp. 133–158). ESECS - Instituto Politécnico de Leiria, CIEd – Universidade do Minho, CIDTFF – Universidade de Aveiro.

Barbeiro, L. (2019). Escrita: Tecer e esculpir o texto. *Letras de Hoje*, *54*(2), 221–230.

Barbeiro, L. (2022). Os processos na atividade de escrita: Estudo com base na escrita colaborativa. *Acta Scientiarum. Language and Culture*, *44*(1), e57804. https://doi.org/10.4025/actascilangcult.v44i1.57804

Barbeiro, L. (2023). A frase como mar de escolhas para aprender a escrever. In V. W. Pereira, L. F. Barbeiro, R. Guaresi, & K. Forneck (Orgs.), *Ensino da compreensão leitora e da produção escrita: Unidades linguísticas em interação com foco na frase* (pp. 60–85). Editora Fonema e Grafema.

Barbeiro, L. (2024). A projeção das experiências de ensino e aprendizagem nos títulos dos relatórios: processos, participantes e contexto. In M. Abreu, I. Dias, T. Ribeiro, M. Barroso, H. Pinto, & A. Neves. *Conferência Internacional Investigação, Práticas e Contextos em Educação: Livro de Atas* (pp. 25–33). Polytechnic University of Leiria.

Barbeiro, L., & Barbeiro, C. (2023). A dimensão metalinguística na reescrita conjunta. In P. Silva, A. Pinto, & C. Marques (Org.), *Discurso académico: Conhecimento disciplinar e apropriação didática* (pp. 19–37). Grácio Editor.

Barbeiro, L., Pereira, L. A., Calil, E., & Cardoso, I. (2022). Termos metalinguísticos e operações de natureza gramatical na escrita colaborativa dos alunos do ensino básico. *Tejuelo*, *35*(2), 45–76. https://doi.org/10.17398/1988-8430.35.2.45

Berninger, V., Whitaker, D., Feng, Y., Swanson, H., & Abbott, R. (1996). Assessment of planning, translating, and revising in junior high writers. *Journal of School Psychology, 34*(1), 23–52.

Brien, D. L., & Brady, T. (2003). Collaborative practice: Categorising forms of collaboration for practitioners. *Text*, *7*(2), 1–11. https://doi.org/10.52086/001c.32496.

Calil, E. (2017). Rasura oral comentada: Definição, funcionamento e tipos em processos de escritura a dois. In C. L. C. Silva, A. Del Ré, & M. Cavalcante (Orgs.), *A criança na/com a linguagem: Saberes em contraponto* (pp. 161–192). UFRGS.

Calil, E., & Myhill, D. (2020). Dialogue, erasure and spontaneous comments during textual composition: What students' metalinguistic talk reveals about newly-literate writers' understanding of revision. *Linguistics and Education, 60*, 100875. https://doi.org/10.1016/j.linged.2020.100875

Cerni, T., & Job, R. (2024). Spelling processing during handwriting and typing and the role of reading and visual-motor skills when typing is less practiced than handwriting. *Reading and Writing: An Interdisciplinary Journal, 37*, 205–237. https://doi.org/10.1007/s11145-023-10418-2

Fabre, C. (1986). Des variantes de brouillon au cours préparatoire. *Études de Linguistique Appliquée*, *62*, 59–79.

Fabre-Cols, C. (2002). *Réécrire à l'école et au collège : De l'analyse des brouillons à l'écriture accompagnée.* ESF.

Felipeto, C., & Calil, E. (2021). A rasura não visível capturada em processo de escrita colaborativa na sala de aula. *Manuscrítica: Revista de Crítica Genética*, *44*, 92–103.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365–387. https://doi.org/10.2307/356600.

Gerova, G. (2022). The think-aloud method in EFL writing: A study with two Bulgarian students. *British Journal of Education*, *10*(2), 47–59.

Graves, D. (1983). *Writing: Teachers and children at work*. Heinemann.

Graves, D. (1984). *A researcher learns to write: Selected articles and monographs*. Heinemann.

Horning, A., & Becker, A. (Eds.). (2006). *Revision: History, theory, and practice*. Parlor Press; The WAC Clearinghouse.

Keen, J. (2020). Writing revision: Evidence for learning. *Changing English: Studies in Culture and Education*, *27*(2), 121–136. https://doi.org/10.1080/1358684X.2019.1653169

Kim, Y.-S. G. (2020). Interactive dynamic literacy model: An integrative theoretical framework for reading and writing relations. In R. Alves, T. Limpo, & M. Joshi (Eds.), *Reading-writing connections: Towards integrative literacy science* (pp. 11–34). Springer.

Latif, M. (2021). Remodeling writers' composing processes: Implications for writing assessment. *Assessing Writing*, *50*, *100547.* https://doi.org/10.1016/j.asw.2021.100547

Lowry, P. B., Curtis, A., & Lowry, M. R. (2004). Building a taxonomy and nomenclature of collaborative writing to improve interdisciplinary research and practice. *Journal of Business Communication*, *41*(1), 66–99.

Myhill, D., & Jones, S. (2015). Conceptualizing metalinguistic understanding in writing. *Cultura y Educación*, *27*(4), 839–867. https://doi.org/10.1080/11356405.2015.1089387

Myhill, D., & Newman, R. (2016). Metatalk: Enabling metalinguistic discussion about writing. *International Journal of Education Research*, *80*, 177–187. https://doi.org/10.1016/j.ijer.2016.07.007

Myhill, D., Jones, S., & Wilson, A. (2016). Writing conversations: Fostering metalinguistic discussion about writing. *Research Papers in Education*, *31*(1), 23–44. https://doi.org/10.1080/02671522.2016.1106694

Myhill, D., Watson, A., & Newman, R. (2020). Thinking differently about grammar and metalinguistic understanding in writing. *Bellaterra Journal of Teaching and Learning Language and Literature*, *13*(2). https://doi.org/10.5565/rev/jtl3.870

O'Donnell, M. J. (2013). A dynamic view of choice in writing: composition as text evolution. In L. Fontaine, T. Bartlett, & G. O'Grady (Eds.), *Systemic functional linguistics: exploring choice* (pp. 247–266). Cambridge University Press.

Pereira, I., & Moura, S. (2024). Compreendendo a inteligência artificial generativa na perspectiva da língua. In *Anais Colóquio Interdisciplinar em Cognição e Linguagem*. Universidade do Norte Fluminense Darcy Ribeiro (UENF). https://www.even3.com.br/anais/vii-coloquio-interdisciplinar-cognicao-linguagem-393557/752031-compreendendo-a-inteligencia-artificial-generativa-na-perspectiva-da-lingua/

Sharples, M., Goodlet, J., Beck, E., Wood, C., Easterbrook, S., & Plowman, L. (1993). Research issues in the study of computer supported collaborative writing. In M. Sharples (Ed.), *Computer supported collaborative writing* (pp. 9–28). Springer.

Sousa, M. (1993). *A interpretação de textos nas aulas de português*. Edições ASA.

Tarchi, C., Villalón, R., Vandermeulen, N., Casado-Ledesma, L., & Fallaci A. (2024). Recursivity in source-based writing: A process analysis. *Reading and Writing: An Interdisciplinary Journal*, *37*, 2571–2593.

Van Reybroeck, M., Penneman, J., Vidick, C., & Galand, B. (2017). Progressive treatment and self-assessment: effects on students' automatisation of grammatical spelling and self-efficacy beliefs. *Reading and Writing: An Interdisciplinary Journal*, *30*, 1965–1985. https://doi.org/10.1007/s11145-017-9761-1

Wengelin, Å., Frid, J., Johansson, R., & Johansson, V. (2019). Combining keystroke logging with other methods: Towards an experimental environment for writing process research. In E. Lindgren, & K. Sullivan (Eds.), *Observing writing: Insights from keystroke logging and handwriting* (pp. 30–49). Brill.

Leite de Vasconcelos, entre publicações, reimpressões e reedições

Leite de Vasconcelos, between publications, reprints and reeditions

Vânia Carvalheiro[1]

[1]Faculdade de Letras da Universidade de Lisboa

Abstract

The vast range of José Leite de Vasconcelos' work is reflected and described in the article published by Isabel Vilares Cepeda in *José Leite de Vasconcellos: Livro do Centenário (1858–1958)*, (1960). The author published his articles in several reviews and journals, from the 70s of the 19th-century until his death, in the 1940s. A brief analysis of Isabel Cepeda's article demonstrates that there was a high incidence of reissues and reprints over the years. Using the aforementioned article as the basis for the present study, the aim is to understand how and to what extent Leite de Vasconcelos reused his articles in subsequent works, as well as to identify the decades in which certain themes were most explored.

***Keywords*:** José Leite de Vasconcelos, editions, reissues, textual criticism, genetic criticism.

Resumo

A vastíssima obra de José Leite de Vasconcelos encontra-se listada e descrita no artigo publicado por Isabel Vilares Cepeda em *José Leite de Vasconcellos: Livro do Centenário (1858–1958)*, datado de 1960. O autor publicou artigos em variados periódicos, desde a década de 70 do século XIX até à data do seu falecimento, na década de 40 do século XX. Uma rápida leitura do artigo de Isabel Cepeda é suficiente para constatar uma elevada incidência de reedições e reimpressões ao longo dos anos. Tendo o artigo referido como base para o presente estudo, pretende perceber-se de que forma e até que ponto Leite de Vasconcelos reutilizou os seus artigos em obras subsequentes, assim como identificar as décadas em que temas específicos foram mais explorados.

Palavras-chave: José Leite de Vasconcelos, edições, reedições, crítica textual, crítica genética.

1. Introdução

A vida de Leite de Vasconcelos foi dedicada ao labor científico, tendo publicado assiduamente o resultado do seu trabalho e das suas pesquisas em periódicos desde a década de 70 do século XIX. Alguns destes trabalhos são, ainda hoje, obras de referência nos diversos campos de estudo aos quais se dedicou. Estes artigos e publicações formam uma extensa listagem, devidamente descrita por Isabel Cepeda no artigo datado de 1960 (Cepeda, 1960). Mediante uma análise detalhada do referido artigo, pretende verificar-se o que foi reeditado ou de alguma forma reintegrado noutra obra, reimpresso, ou seja, novamente impresso de forma autónoma, e o que, por algum motivo, permaneceu com uma única edição.

O artigo de Cepeda segue uma estrutura semelhante ao *Indículo das Obras de José Leite de Vasconcelos*, de Moses Amzalak, publicado em 1924 (Amzalak, 1924), no qual as publicações foram agrupadas por secções sob a orientação de Leite de Vasconcelos. Seguindo a mesma metodologia, o artigo de Cepeda está dividido em

sete secções temáticas, com as respetivas subsecções. Assim, a secção I. poesia, agrupa toda a sua obra poética publicada, sem quaisquer subdivisões. A secção II. etnografia portuguesa foi dividida nas subsecções: A) generalidades e síntese; B) a terra; C) o povo; D) etnografia lusitana; E) arqueologia propriamente dita, com o apêndice: história da arqueologia portuguesa; F) etnografia moderna, com as subdivisões: a) ergologia e b) folclore; e, finalmente, G) Museu etnológico português. A secção III. filologia, é composta pelas subdivisões: A) glotologia geral; B) latim; C) língua nacional, com as subdivisões: a) trabalhos doutrinais, b) história e crítica e c) literatura; D) onomatologia, com as subdivisões: a) antroponímia, b) toponímia e c) nomes de pessoas e nomes geográficos ao mesmo tempo; E) dialetologia portuguesa propriamente dita; F) galego; G) linguagens das raias portuguesa e espanhola; H) um dialeto espanhol; I) provençal; e, finalmente, J) vários textos. No que concerne à secção IV. poligrafia, foi dividida nas subdivisões: A) numismática, medalhística, tesserologia; B) biografias; e C) outros trabalhos. Seguem-se as secções V. publicações periódicas, VI. *Opúsculos* e, por fim, a secção VII. críticas bibliográficas. Os títulos publicados são agrupados em cada secção e subsecção por ordem cronológica.

A análise terá como foco o que foi reeditado ou reimpresso e o que permaneceu somente na sua publicação original em cada secção. Serão incluídas informações do *Indículo* de Moses Amzalak (Amzalak, 1924), dos *Opúsculos* (Vasconcelos, 1928a, 1928b, 1929b, 1931b, 1938b, 1938c, 1985), da PORBASE (PORBASE, s.d.), assim como de alguns manuscritos no espólio do autor à guarda do Museu Nacional de Arqueologia.[1]

Leite de Vasconcelos publicou artigos e obras de extensão variável, podendo ser encontrados artigos de um parágrafo somente ou obras extensas com mais de 600 páginas. No presente artigo, considerar-se-ão os títulos publicados independentemente da sua dimensão ou extensão na contabilização. A par da indicação anterior, serão consideradas compilações as obras *Baladas do Ocidente* (Vasconcelos, 1885a), *Nuvens* (Vasconcelos, 1898), *De Terra em Terra* (Vasconcelos, 1927b, 1927c), *Mês de Sonho* (Vasconcelos, 1926), *Religiões da Lusitânia* (Vasconcelos, 1897g, 1905b, 1913d), *Opúsculos* (Vasconcelos, 1928a, 1928b, 1929b, 1931b, 1938b, 1938c, 1985), *Etnografia Portuguesa* (Vasconcelos, 1933c, 1936, 1941b) e *Ensaios Etnográficos* (Vasconcelos, 1891b, 1903a, 1906a, 1910), em virtude de agruparem em si títulos previamente publicados de forma autónoma.

Não serão analisadas as secções V. publicações periódicas, considerando que, por vezes, constituem o ponto de partida de publicação, assim como a secção VI. *Opúsculos* (Vasconcelos, 1928a, 1928b, 1929b, 1931b, 1938b, 1938c, 1985), considerada compilação de títulos anteriormente publicados.

No que diz respeito a títulos que foram divididos e reeditados em separado, considerar-se-ão como incluídos no volume que agrupa o maior número de parcelas do título, conforme publicado inicialmente.

2. Poesia

No âmbito da poesia, Leite de Vasconcelos publicou 67 títulos entre 1875, "A naiade" (Vasconcelos, 1875a), e 1924, "À beira-mar" (Vasconcelos, 1924a). Para a sua última poesia, publicada na revista *O Soneto Neolatino* (Vasconcelos, 1929–1933), não é apresentada a data exata, sabendo-se, no entanto, que a revista foi publicada com a data 1929–1933 (Alonso Montero, 2011, p. 14). Publicou oito títulos na década de 70, 38 na década de 80, contabilizando o poema "Arcades Ambo", inicialmente publicado nas *Baladas do Ocidente* (Vasconcelos, 1885a), 12 na década de 90, quatro na década de 1900, dois na década de 1910, dois na década de 1920 e o último soneto publicado entre 1929 e 1933.

As poesias foram compiladas em dois volumes. O primeiro, *Baladas do Ocidente* (Vasconcelos, 1885a), publicado em 1885 pela Livraria Portuense Editora, contém a reedição de 13 poemas, aos quais, conforme referido, juntou "Arcades Ambo". O segundo, *Nuvens* (Vasconcelos, 1898), foi publicado nos anos 90 do século XIX pela Livraria Chardron, e contém a reedição de nove poemas. Treze dos seus títulos publicados foram

[1] Uma palavra de agradecimento à Dra. Lívia Cristina Coito pelo cuidado e carinho com que acedeu às minhas solicitações durante as minhas visitas à biblioteca do Museu Nacional de Arqueologia e sem cuja preciosa ajuda este estudo não teria sido possível.

reimpressos em separatas, sendo dez deles também incluídos nas compilações. No que concerne a reedições, falta somente referir que cinco títulos foram incluídos em publicações diversas. Finalmente, são 37 os títulos de poesia publicados apenas uma vez.

Tabela 1

Poesia

Uma edição	37
Separatas	13
Publicações várias	5
Baladas do Ocidente	13
Nuvens	9

3. Etnografia portuguesa

O estudo do povo e da cultura portuguesa foi o catalisador das muitas pesquisas que Leite de Vasconcelos efetuou ao longo da vida. As obras que publicou foram divididas nas secções A) generalidades; B) a terra; C) o povo; D) etnografia lusitana; E) arqueologia propriamente dita; F) etnografia moderna e G) Museu etnológico português, analisadas em seguida.

3.1. A) Generalidades

Ao olhar globalmente para a secção, e excluindo a obra *Etnografia Portuguesa* (Vasconcelos, 1933c, 1936, 1941b), por conter títulos inicialmente publicados de forma autónoma, contabilizam-se apenas nove títulos: quatro publicados na década de 1910, um na década de 1920 e quatro na década de 1930.

Introduziu dois títulos no volume V dos *Opúsculos* (Vasconcelos, 1938b) e o artigo "O Museu Etnográfico dos Açores" (Vasconcelos, 1930e) no volume VII (Vasconcelos, 1938c) da mesma compilação. No que concerne aos artigos "Importância da Etnografia" (Vasconcelos, 1919b), "Considerações Gerais e Sumárias acerca das Fontes de Investigação Etnográfica" (Vasconcelos, 1930b) e "Fontes de Investigação Etnográfica" (Vasconcelos, 1937a), são incorporados no volume I de *Etnografia Portuguesa* (Vasconcelos, 1933c). No âmbito de reedições, deve referir-se o artigo "Etnografia Portuguesa (Tentame de Sistematização)", reeditado na *Revista Lusitana* (Vasconcelos, 1934a). Restam apenas dois títulos que permaneceram como publicação única.

3.2. B) A terra

No âmbito da secção relativa à terra, Leite de Vasconcelos publicou, entre 1878, "O Presbitério de Vila Cova" (Vasconcelos, 1878–1879), e 1940/1941, "Condados da Idade-Média" (Vasconcelos, 1940–1941), 45 títulos, contabilizando "Excursão pelo Baixo Alentejo" (Vasconcelos, 1927c), entre breves artigos e obras de grande envergadura. Dois títulos foram publicados na década de 1870, três na década de 1880, seis na década de 1890, três na década de 1900, nove na década de 1910, nove na década de 1920, doze na década de 1930 e apenas um na década de 1940.

Publicou 13 títulos em separatas autónomas, sendo onze destas separatas incorporadas na compilação *De Terra em Terra* (Vasconcelos, 1927b, 1927c). Quatro separatas foram incluídas no volume I (Vasconcelos, 1927b), às quais se juntaram os títulos: "Notícias transmontanas do princípio do século XVII" (Vasconcelos, 1894a), "Antiguidades da Beira" (Vasconcelos, 1896), "Uma Viagem à Beira" (Vasconcelos, 1902i), "Trás-os-Montes Etnográfico" (Vasconcelos, 1920d), "Geografia Tradicional da Beira" (Vasconcelos, 1921b) e "Aspetos do Minho" (Vasconcelos, 1924c). No volume II (Vasconcelos, 1927c), foram agrupadas sete separatas, às quais

juntou o título "Excursão Arqueológica a Alcácer-do-Sal" (Vasconcelos, 1895b), perfazendo, por isso, 18 títulos do total publicado. Foram ainda adicionados o inédito "Excursão pelo Baixo Alentejo" (Vasconcelos, 1927c), a separata *Excursão Interamnense* (Vasconcelos, 1917b), publicada pela Imprensa Nacional em 1917, e a separata *Excursão Extremenha* (Vasconcelos, 1917c), publicada em Lisboa no mesmo ano, que serão analisadas na respetiva secção.

Quanto aos títulos reeditados ou desenvolvidos em *Etnografia Portuguesa* (Vasconcelos, 1933c, 1936, 1941b), podem contabilizar-se 11 títulos: dois no volume II (Vasconcelos, 1936) e nove no volume III (Vasconcelos, 1941b). No que concerne aos *Opúsculos*, no volume I (Vasconcelos, 1928a) incluiu a nota 3 da página 170 do título "Entre Tejo e Odiana" (Vasconcelos, 1916b); no volume II (Vasconcelos, 1928b) incluiu o glossário de "Uma Excursão ao Soajo" (Vasconcelos, 1882f); no volume III (Vasconcelos, 1931b) inclui "*Annotationes ad Geographiam Lusitanam*" (Vasconcelos 1897a), a nota2 da página 30 e a etimologia de "Meixedo, Meixendo" de "Por Trás-os-Montes" (Vasconcelos, 1917e); no volume V (Vasconcelos, 1938b) incluiu "Geografia Antiga do Algarve" (Vasconcelos, 1913a), a segunda parte do artigo "Celorico da Beira" (Vasconcelos, 1930a) e "Divisões geográficas" (Vasconcelos, 1930c); finalmente, no volume VII (Vasconcelos, 1938c) incluiu "Em Trás-os-Montes" (Vasconcelos, 1884b), perfazendo um total de nove títulos adicionados aos *Opúsculos*. São dois os títulos incluídos em *Ensaios Etnográficos*, a saber: "O Presbitério de Vila Cova" (Vasconcelos, 1878–1879), adicionado ao volume I (Vasconcelos, 1891b), e "Em Évora" (Vasconcelos, 1888b), adicionado ao volume IV (Vasconcelos, 1910). O artigo "Geografia da Lusitânia na Época Proto-histórica" (Vasconcelos, 1903b) é incluído no volume II das *Religiões da Lusitânia* (Vasconcelos, 1905b). São ainda reeditados, em publicações diversas, os títulos "O Presbitério de Vila Cova" (Vasconcelos, 1878–1879), "Uma Excursão ao Soajo" (Vasconcelos, 1882f), "Campo de Coimbra" (Vasconcelos, 1929a) e "Excursão pelo Baixo Alentejo", previamente publicado como inédito em *De Terra em Terra* (Vasconcelos, 1927b).

Cinco títulos são publicados somente uma vez, contabilizando o inédito "Excursão pelo Baixo Alentejo", datado de 1897, introduzido em *De Terra em Terra* (Vasconcelos, 1927c). De referir que dois destes cinco títulos são: *Memórias de Mondim da Beira (Para a História do Concelho deste Nome)*, publicado pela Imprensa Nacional em 1933 (Vasconcelos, 1933d), um volume de 471 páginas; e o prefácio à obra de Manuel de Vasconcelos *A Vila de Canaveses — Notas para a sua História* (Vasconcelos, 1935b).

3.3. **C) O povo**

No que concerne ao povo português, Leite de Vasconcelos publicou 25 títulos entre 1887, "A Nacionalidade Portuguesa" (Vasconcelos, 1887), e 1941, "Origem do Povo Português" (Vasconcelos, 1941c). Sete títulos foram publicados na década de 1880, um na década de 1890, três na década de 1900, um na década de 1910, quatro na década de 1920, oito na década de 1930 e o último na década de 1940.

Quanto a estes títulos, elaborou cinco separatas e incluiu oito títulos no volume V dos *Opúsculos* (Vasconcelos, 1938b). Na *Etnografia Portuguesa* foram introduzidos dois artigos no volume I (Vasconcelos, 1933c), três artigos no volume II (Vasconcelos, 1936) e, finalmente, "Os Saloios" (Vasconcelos, 1939) no volume III (Vasconcelos, 1941b), perfazendo um total de seis artigos. O título "*Les Celtes de la Lusitanie Portugaise*" (Vasconcelos, 1902b) foi incluído no volume II das *Religiões da Lusitânia* (Vasconcelos, 1905b).

Foram dez os títulos publicados apenas uma vez, sendo oito deles correspondentes aos seus primeiros artigos sobre o tema.

3.4. **D) Etnografia lusitana**

Leite de Vasconcelos publicou 31 títulos neste âmbito. O primeiro, *Portugal Pré-histórico* (Vasconcelos, 1885d), em 1885, e a sua última publicação datada de 1934 corresponde ao título "Inscrição Ibérica do Algarve" (Vasconcelos, 1934b). Publicou um título na década de 1880, onze na década de 1890, nove na década de 1900, um na década de 1910, sete na década de 1920, um na década de 1930 e o seu último artigo não se encontra datado.

No que diz respeito a reedições nos *Opúsculos* o artigo "*Onomasticon Lusitanien*" (Vasconcelos, 1900d) foi incluído no volume III (Vasconcelos, 1931b) e foram incluídos treze artigos no volume V (Vasconcelos, 1938b).

Os títulos "*Les Monnais de la Lusitanie Portugaise*" (Vasconcelos, 1900c), "*Onomasticon Lusitanien*" (Vasconcelos, 1900d), "*Signification Religieuse, en Lusitanie, de Quelques Monnais Percées d'un Trou*" (Vasconcelos, 1905c) e "Medicina dos Lusitanos" (Vasconcelos, 1925c) foram reeditados em publicações diversas. Foram publicados em separata 12 títulos e outros tantos foram publicados apenas uma vez, correspondendo cinco destes títulos aos últimos títulos publicados.

3.5. **E) Arqueologia propriamente dita**

Esta é uma secção cuja análise é dificultada pelo facto de Leite de Vasconcelos não ter conseguido terminar a publicação planeada dos *Opúsculos*, que incluiria volumes para as secções de arqueologia, numismática, epigrafia, biografias, assim como um volume de miscelânea.

Para esta secção, procede-se à análise dos títulos que realmente foram reimpressos ou reeditados pelo autor, sendo, no final, feita a análise dos dados passíveis de recolher nos documentos existentes no espólio e que permitem ter uma perspetiva do que teria pensado incluir no volume dos *Opúsculos*.

No âmbito da arqueologia, entre breves notícias e obras mais extensas, publicou 241 títulos. O seu primeiro artigo data de 1880–1881, "Notas de Pré-História" (Vasconcelos, 1880–1881b) e o seu último artigo data de 1940, "Lápide Romana de São Miguel de Odrinhas, Concelho de Sintra" (Vasconcelos, 1940b). Escreveu seis títulos na década de 1880, 85 na década de 1890, 72 na década de 1900, 37 na década de 1910, 26 na década de 1920, 13 na década de 1930, um na década de 1940 e, finalmente, contabiliza-se um título publicado sem data. Leite de Vasconcelos refere ainda, nos documentos preparatórios para o volume dos *Opúsculos* de arqueologia, um "artigo publicado num jornal da Faculdade de Letras", que, no entanto, não foi possível identificar, nem o respetivo ano de publicação.

Para os 241 títulos publicou 56 separatas, onde se incluem títulos como: "Estudos sobre Panóias" (Vasconcelos, 1897c), com duas separatas publicadas, ambas em Lisboa e datadas de 1897; "Arqueologia do Alto-Minho" (Vasconcelos, 1900a), também publicada em Lisboa, em 1900; "Objetos Paleolíticos do Casal do Monte" (Vasconcelos, 1914–1915b), publicada pela Imprensa da Universidade de Coimbra, em 1915; "Descobridores de Monsanto" (Vasconcelos, 1922a), publicada em Lisboa e datada de 1922 ou "Castros Lusitânicos" (Vasconcelos, 1933a), publicada em Lisboa em 1931.

São dez os títulos incluídos em publicações diversas, nove deles reeditados no *Arqueólogo Português*.

O artigo "Inscrição Inédita de Mercúrio em Moura e Vários Costumes Sepulcrais da Época Romana em Portugal" (Vasconcelos, 1892) foi reeditado no volume V dos *Opúsculos* (Vasconcelos, 1938b). A parte referente à etimologia de "Évora" do artigo "Antiquités du Sud de Portugal" (Vasconcelos, 1899), a parte referente ao étimo de "Torre de Gandufe", a etimologia de "Rominha" em "Coisas Velhas" (Vasconcelos, 1917a) e a etimologia de Bragança do artigo "Museu Municipal de Bragança" (Vasconcelos, 1897f) foram incluídos no volume III dos *Opúsculos* (Vasconcelos, 1931b). Conforme referido anteriormente, as separatas *Excursão Interamnense* (Vasconcelos, 1917b) e *Excursão Extremenha* (Vasconcelos, 1917c) foram ambas reeditadas nos volumes I e II da obra *De Terra em Terra* (Vasconcelos, 1927b, 1927c). Ficaram com uma edição apenas 182 títulos.

As anotações de Leite de Vasconcelos, sobre os títulos que pretenderia incluir no seu volume de arqueologia dos *Opúsculos* não são muito precisas, ao contrário do que acontece para o volume de numismática. No entanto, analisar-se-ão as informações possíveis de reunir a partir dos dois planos preparatórios para este volume. Num plano inicial, datado de 19 de agosto de 1935, Leite de Vasconcelos indica que pretenderia incorporar o artigo "Borges de Figueiredo e a Arqueologia Portuguesa" (Vasconcelos, 1890), ao qual adicionaria outros artigos. Seguir-se-ia o "Discurso Inaugural, Pronunciado em Coimbra em 21 de Setembro de 1930, na 1.ª Sessão (Conjunta) dos Congressos de Antropologia e Arqueologia Pré-histórica e do Instituto

Internacional de Antropologia" (Vasconcelos, 1933b), seguido do artigo "Para a História da Arqueologia Portuguesa" (Vasconcelos, 1926b), tendo anotado "Desde a Idade-Média até os Fins do Século XIX" (Vasconcelos s.d.a). Pretenderia terminar esta secção com o artigo relativo à fundação do Museu Etnológico e, em seguida, reeditar a globalidade dos seus artigos organizados por épocas. Dada a vastidão de publicações nesta área, termina a anotação com a consideração de que talvez devesse usar a totalidade do volume para os artigos de arqueologia, iniciando-o com o artigo "História da Arqueologia em Portugal", deixando o artigo sobre Borges de Figueiredo para o volume de biografias e o discurso para o volume de miscelânea.

Num plano de data posterior, indica ter feito, no ano de 1938, um plano de história da arqueologia portuguesa que terminava com a fundação do Museu Etnológico. Nesse plano incluiria um Apêndice com breves notícias biográficas de arqueólogos falecidos depois da fundação do Museu, que remeteriam para o volume das biografias. A este plano segue-se um documento, datado de 28 de dezembro de 1938, onde indica que pretenderia incluir na história da arqueologia portuguesa os artigos: "Significação do Museu Etnológico Português" (Vasconcelos, 1912b), "Discurso Inaugural, Pronunciado em Coimbra em 21 de Setembro de 1930, na 1.ª Sessão (Conjunta) dos Congressos de Antropologia e Arqueologia Pré-histórica e do Instituto Internacional de Antropologia" (Vasconcelos, 1933b), "Borges de Figueiredo e a Arqueologia Portuguesa" (Vasconcelos, 1890), assim como um artigo publicado num jornal da Faculdade de Letras, que não foi possível identificar até ao momento, "precedido de um: esboço da história da arqueologia portuguesa, desde o começo até fins de 1853, fundação do Museu Etnológico, […]. Com um Apêndice: notícias biográficas dos arqueólogos falecidos depois de 1923 (algumas já feitas por mim)" (Vasconcelos, s.d.a). Pretenderia também incluir o "Prolóquio" (Vasconcelos, 1922c) que escreveu para a *Revista de Arqueologia e História*, na época em que foi presidente da Associação dos Arqueólogos Portugueses (1922), que não se encontra incluído no artigo de Cepeda (1960).

3.6. **F) Etnografia moderna**

A secção de etnografia moderna encontra-se dividida em duas secções, pelo que os títulos serão analisados separadamente.

3.6.1. **a) Ergologia**

Leite de Vasconcelos escreveu 86 títulos sobre o tema. O primeiro título, "Estudo Etnográfico a Propósito da Ornamentação dos Jugos e Cangas dos Bois nas Províncias Portuguesas do Douro e Minho" (Vasconcelos, 1881), datado de 1881, e o último, "Lume e Iluminação" (Vasconcelos, 1937b), de 1937. Escreveu dois artigos na década de 1880, dois artigos na década de 1890, um artigo na década de 1900, quatro artigos na década de 1910, 73 na década de 1920 e quatro na década de 1930.

Oito destes títulos foram publicados em separatas e nove títulos foram reeditados no volume V dos *Opúsculos* (Vasconcelos, 1938b). O título *A Figa — Estudo de Etnografia Comparativa, Precedido de Algumas Palavras a Respeito do "Sobrenatural" na Medicina Popular Portuguesa* (Vasconcelos, 1925a), foi o único título publicado numa publicação diversa. No que concerne aos títulos publicados apenas uma vez contabilizam-se 75 títulos.

3.6.2. **b) Folclore**

No que concerne este tema, Leite de Vasconcelos escreveu 177 títulos. O primeiro, em 1879, "Mitologia Popular Portuguesa" (Vasconcelos, 1879a), e o seu último artigo, "Curiosidades Etnográfico-Antropológicas" (Vasconcelos, 1941a), está datado de 1941. Publicou dois títulos na década de 1870, 69 títulos na década de 1880, nove títulos na década de 1890, 19 na década de 1900, 33 na década de 1910, 28 na década de 1920, 15 títulos na década de 1930 e dois títulos na década de 1940, excluindo o *Romanceiro Português* (Vasconcelos, 1958), por ter sido publicado a título póstumo.

Destes títulos foram elaboradas 19 separatas. Foram reeditados 47 títulos nos *Ensaios Etnográficos*: três títulos no volume I (Vasconcelos, 1891b), 15 títulos no volume II (Vasconcelos, 1903a), 11 títulos no volume III (Vasconcelos, 1906a) e 18 títulos no volume IV (Vasconcelos, 1910). Foram 79 os títulos integrados nos *Opúsculos*. Três no volume I (Vasconcelos, 1928a), incluindo o artigo "Pôr Freio a Alguém" (Vasconcelos, 1920b), que não é mencionado no artigo de Isabel Cepeda, 21 títulos no volume V (Vasconcelos, 1938b) e 58 títulos no volume VII (Vasconcelos, 1938c). O artigo "Unha na Palma" (Vasconcelos, 1924f) foi reeditado em *Mês de Sonho* (Vasconcelos, 1926a). No que diz respeito a títulos reeditados noutras publicações contabilizam-se quatro títulos: "As Maias" (Vasconcelos, 1882a), "Poesia e Etnografia" (Vasconcelos, 1902f), "Uma Expressão Popular Portuguesa: "Mais Vale um Gosto que Quatro Vinténs" (Vasconcelos, 1911c) e "Lenda Popular. A Truta de Celorico da Beira" (Vasconcelos, 1929c). O artigo "No Ventre da Virgem Bela", tendo sido integrado nas *Nótulas Etnográficas* (Vasconcelos, 1917–1918) previamente contabilizadas, não foi aqui contabilizado. Ficaram com uma edição apenas 44 títulos.

3.7. **G) Museu etnológico português**

No que diz respeito ao Museu etnológico português, Leite de Vasconcelos publicou 11 títulos. O primeiro, em 1895, "Museu Etnográfico Português" (Vasconcelos, 1895c), e o último, em 1919, *Sinopse do Museu Etnológico Português* (Vasconcelos, 1919c). Publicou dois títulos na década de 1890, quatro títulos na década de 1900 e cinco títulos na década de 1910.

Em separata foram publicados dois títulos: "Museu Etnográfico Português" (Vasconcelos, 1895c), pela Tipografia de A. F. de Vasconcelos em 1894, e *Museu Etnológico Português* (Vasconcelos, 1897e), em Lisboa, datada de 1897. Em *História do Museu Etnológico Português* (Vasconcelos, 1915a), Leite de Vasconcelos compilou oito títulos e em publicações diversas foram reeditados cinco títulos: *Museu Etnológico Português* (Vasconcelos, 1897e), *Notice Sommaire sur le Musée Etnologique Portugais* (Vasconcelos, 1904–1905), *Musée Ethnologique Portugais* (Vasconcelos, 1906b), *Plano Sumário do Museu Etnológico Português* (Vasconcelos, 1906c) e *Visita ao Museu Etnológico Português* (Vasconcelos, 1906c). No que concerne aos *Opúsculos*, introduziu o artigo *Sinopse do Museu Etnológico Português* (Vasconcelos, 1919c) no volume VII (Vasconcelos, 1938c) e, segundo as suas anotações, terá planeado introduzir o artigo "Significação do Museu Etnológico Português" (Vasconcelos, 1912) no volume de arqueologia. Finalmente, contabilizam-se apenas dois títulos com uma edição apenas.

Tabela 2

Etnografia Portuguesa

	A)	**B)**	**C)**	**D)**	**E)**	**F) a)**	**F) b)**	**G)**
Uma Vez Apenas	2	5	10	12	182	75	44	2
Separatas	—	13	5	12	56	8	19	2
Publicações Várias	1	4	—	4	10	1	4	5
Etnografia Portuguesa	3	11	6	—	—	—	—	—
Ensaios Etnográficos	—	2	—	—	—	—	47	—
História do Museu	—	—	—	—	—	—	—	8
Religiões da Lusitânia	—	1	1	—	—	—	—	—
De Terra em Terra	—	18	—	—	2	—	—	—
Mês de Sonho	—	—	—	—	—	—	1	—
Opúsculos	3	9	8	14	5	9	82	1

4. Filologia

Entramos agora no âmbito da linguística, onde muitos dos títulos são, ainda hoje, fundamentais, e encontram-se organizados nas secções que se analisam abaixo.

4.1. A) Glotologia geral

No que diz respeito à glotologia geral, Leite de Vasconcelos publicou quatro títulos, todos na década de 80, e tendo todos eles sido reimpressos ou reeditados. O primeiro, "Fonética da Linguagem Infantil" (Vasconcelos, 1883–1884), foi publicado em 1883–1884 e o último, *A Evolução da Linguagem*, a sua tese de licenciatura, (Vasconcelos, 1886) foi publicada dois anos mais tarde.

Em separata, publicou apenas um artigo, "Contribuições para o Estudo da Linguagem Infantil" (Vasconcelos, 1883), na Tipografia do *Tirocínio*, em 1883–1884, e em publicações diversas integrou o artigo "Fonética da Linguagem Infantil" (Vasconcelos, 1883–1884b). No que concerne a títulos integrados nos *Opúsculos*, contabilizam-se três. A saber: "Contribuições para o Estudo da Linguagem Infantil" (Vasconcelos, 1883a), "Patologia da Linguagem" (Vasconcelos, 1884e) e *A Evolução da Linguagem* (Vasconcelos, 1886), todos no volume I (Vasconcelos, 1928a).

4.2. B) Latim

No que concerne ao Latim, publicou cinco títulos. O primeiro, "Pronúncia do "H" Latino no Século V" (Vasconcelos, 1887–1889b), foi publicado em 1887–1889, e o último, "Três Anos de Latim na Faculdade de Letras da Universidade de Lisboa" (Vasconcelos, 1915c), em 1915. Publicou um título na década de 1880, um na década de 1890, um na década de 1900, e dois na década de 1910.

Apenas um destes títulos, "Esopo Adelfo" (Vasconcelos, 1904b), datado de 1904, não foi reimpresso ou reeditado e dois títulos foram publicados em separatas. No volume I dos *Opúsculos* (Vasconcelos, 1928a) foram reeditados quatro títulos, a saber: "Pronúncia do "H" Latino no Século V" (Vasconcelos, 1887–1889b), *Quid apud Lusitanos Verbum "Aedeoli" Significaverit* (Vasconcelos, 1894b), "Da Importância do Latim" (Vasconcelos, 1911a) e "Três Anos de Latim na Faculdade de Letras da Universidade de Lisboa" (Vasconcelos, 1915c). Nenhum destes títulos foi integrado noutra publicação.

4.3. C) Língua nacional

Os títulos incluídos na secção língua nacional encontram-se divididos em três subsecções, que se analisam em seguida.

4.3.1. a) Trabalhos doutrinais

No que concerne a esta secção, foram publicados 108 títulos. O primeiro, "Linguagem Popular Portuguesa" (Vasconcelos, 1880–1881a), publicado em 1880–1881, e o último "Matéria Filológica" (Vasconcelos, 1931a), publicado em 1937. Publicou 11 títulos na década de 1880, 13 títulos na década de 1890, 20 na década de 1900, 26 na década de 1910, 22 títulos na década de 1920 e 16 na década de 1930.

Leite de Vasconcelos publicou 12 títulos em separatas. Setenta e quatro títulos foram reeditados nos *Opúsculos*: 60 títulos no volume I (Vasconcelos, 1928a), incluindo o artigo "Publicação de Textos Antigos" (Vasconcelos, 1921c), sem menção no artigo de Isabel Cepeda; nove títulos no volume III (Vasconcelos, 1931b) e cinco títulos no volume IV (Vasconcelos, 1929b). Em publicações diversas contabilizam-se oito títulos. Vinte e sete títulos foram publicados somente uma vez. É importante referir que destes 27 títulos, 16 foram publicados após a publicação dos volumes de filologia dos *Opúsculos*, ou seja, entre 1929 e 1937, tendo também, durante esta época, sido publicadas apenas duas separatas.

4.3.2. **b) História e crítica**

No âmbito da história e crítica, o autor publicou 15 títulos. O primeiro, "Currente Calamo" (Vasconcelos, 1882b), em 1882, e o último, "História da Língua Portuguesa" (Vasconcelos, 1925b), em 1925. Publicou seis títulos na década de 1880, três na década de 1890, um na década de 1900, quatro na década de 1910 e um na década de 1920.

Dos títulos incluídos nesta secção foram publicadas quatro separatas. Sete títulos foram integrados em dois volumes dos *Opúsculos*: no volume I (Vasconcelos, 1928a), os títulos "Discussão Filológica: A Palavra "Momo"" (Vasconcelos, 1912–1913) e "Discussão Filológica: Perneta" (Vasconcelos, 1919–1920); no volume IV (Vasconcelos, 1929b), Leite de Vasconcelos, incluiu "A reforma do Curso Superior de Letras e a filologia portuguesa" (Vasconcelos, 1888a), "Crítica Literária: *Os Lusíadas* de Camões" (Vasconcelos, 1889a), "*Lições Práticas de Linguagem Portuguesa*, por Cândido de Figueiredo" (Vasconcelos, 1891c), *O Gralho Depenado* (Vasconcelos, 1891c) e, finalmente, "O Dicionário da Academia" (Vasconcelos, 1914–1915a). No que diz respeito à reedição em publicações diversas, são três os títulos reeditados: "A Reforma do Curso Superior de Letras e a Filologia Portuguesa" (Vasconcelos, 1888a), "Crítica Literária: *Os Lusíadas* de Camões" (Vasconcelos, 1889a) e "O Gralho Depenado" (Vasconcelos, 1891c), com duas reedições. São oito os títulos publicados apenas uma vez.

4.3.3. **c) Literatura**

São 32 os títulos publicados no campo da literatura. O primeiro, "Um Inédito" (Vasconcelos, 1879b), publicado em 1879, e o último, "Um Dicionário Português-Latino" (Vasconcelos, 1930f), em 1930. Publicou dois títulos na década de 1870, quatro títulos na década de 1880, três na década de 1890, 12 na década de 1900, cinco na década de 1910, cinco na década de 1920 e apenas um na década de 1930.

Publicou seis títulos em separatas e incluiu nos volumes dos *Opúsculos* oito títulos: quatro no volume I (Vasconcelos, 1928a), três no volume IV (Vasconcelos, 1929b) e, finalmente, o artigo "O Cantador de Setúbal" (Vasconcelos, 1902e), no volume VII (Vasconcelos, 1938c). No que concerne a reedições em publicações diversas, contabilizam-se dois títulos: "O Cantador de Setúbal" (Vasconcelos, 1902e) e *Textos Arcaicos* Vasconcelos, 1903–1905b), com duas edições. Falta apenas referir que o artigo "Algumas Particularidades da Vida de Antero" (Vasconcelos, 1924b) foi reeditado em *Mês de Sonho* (Vasconcelos, 1926a) e que ficaram com uma publicação apenas 19 títulos.

4.4. **D) Onomatologia**

Também os títulos relativos à onomatologia foram divididos em três subsecções, as quais se analisam em seguida.

4.4.1. **a) Antroponímia**

No âmbito da antroponímia, Leite de Vasconcelos publicou 14 títulos. O primeiro em 1903–1905, "Amaral" (Vasconcelos, 1903–1905) e o último em 1940, "Um Cognome Raro na Epigrafia (Presalso)" (Vasconcelos, 1940d). Publicou dois artigos na década de 1900, três artigos na década de 1910, sete artigos na década de 1920, um na década de 1930 e um na década de 1940.

Dois títulos foram publicados em separata: "Etimologia de um Nome Ilustre" (Vasconcelos, 1921a), publicada no Porto em 1921, e "Preito Filológico, Prestado a um Insigne Orador" (Vasconcelos, 1923b), publicada em Coimbra, em 1923. Dez destes títulos foram reeditados no volume III dos *Opúsculos* (Vasconcelos, 1931b) e o artigo ""Pedro" e "Pedra"" (Vasconcelos, 1913c) foi também publicado na *Revista Lusitana* (Vasconcelos, 1913), tendo permanecido somente os seus quatro últimos títulos com uma publicação única.

4.4.2. **b) Toponímia**

No âmbito da toponímia, Leite de Vasconcelos escreveu 52 títulos. O primeiro, publicado em 1882, "Duas Etimologias (Chaminha e Vilarelho)" (Vasconcelos, 1882c), e o último, publicado pouco tempo após a sua morte, "A Província de Entre-Douro-e-Minho. Seu Nome Geral e Nomes Étnicos" (Vasconcelos, 1942). Publicou oito títulos na década de 1880, cinco na década de 1890, três na década de 1900, oito na década de 1910, 15 na década de 1920, nove na década de 1930 e quatro na década de 1940.

Cinco destes títulos deram origem a separatas: "Amostra de Toponímia Portuguesa" (Vasconcelos, 1918a, 1919a), com duas separatas autónomas, ambas publicadas pela Tipografia de Sequeira (1918 e 1919); "Safira" (Vasconcelos, 1918–1919b), publicada em Coimbra em 1919; "Riba d'Ave" (Vasconcelos, 1921–1922), publicada em Coimbra em 1913; "Alguns Topónimos Portugueses" (Vasconcelos, 1927a), publicada em Cernăuti, em 1927; e, finalmente, "A Ideia de "fonte" na Toponímia Portuguesa" (Vasconcelos, s.d.b), publicada em Turim. No volume I dos *Opúsculos* (Vasconcelos, 1928a) foi introduzido apenas o artigo "Nomes Gentílicos" (Vasconcelos, 1884d) e no volume III (Vasconcelos, 1931b) foram introduzidos 35 títulos. Apenas o artigo "Alguns Topónimos Portugueses" (Vasconcelos, 1927a) foi reeditado numa publicação diversa e contabilizam-se 15 títulos publicados somente uma vez.

4.4.3. **c) Nomes de pessoas e nomes geográficos ao mesmo tempo**

Os quatro títulos que se inscrevem neste âmbito foram, em momento posterior, reeditados. Dois foram publicados na década de 1880, um na década de 1910 e o último na década de 1920.

Foram publicadas duas separatas: "Enquissas Onomatológicas" (Vasconcelos, 1918b), publicada pela Tipografia de Sequeira em 1918, e "Nomes de Pessoas Tornados Geográficos em Portugal" (Vasconcelos, 1920–1921), publicada em Coimbra em 1923. Os quatro artigos foram integrados no volume III dos *Opúsculos* (Vasconcelos, 1931b).

4.5. **E) Dialetologia portuguesa propriamente dita**

No campo da Dialetologia, e não tendo o autor conseguido terminar o volume dos *Opúsculos* subordinado ao tema, que se encontrava em impressão à data da sua morte, faz-se o confronto com o volume editado por Maria Adelaide Vale Cintra, sendo adicionados, assim, os títulos nele reeditados.

O autor publicou 16 títulos neste âmbito. O primeiro, "Subdialeto Alentejano" (Vasconcelos, 1883c), em 1883, e o último, "Cacografia Dialetal" (Vasconcelos, 1916a), em 1916. Publicou oito títulos na década de 1880, seis na década de 1890, um na década de 1900 e outro na década de 1910.

Apesar de ter publicado 16 títulos, foram contabilizadas 24 separatas, pois alguns títulos mereceram um tratamento diferenciado, com a publicação de diversas separatas. Foram integrados seis títulos nos volumes dos *Opúsculos*. Três títulos no volume II (Vasconcelos, 1928b): "Linguagem Popular de Barcelos" (Vasconcelos, 1884c), "Dialetos Minhotos" (Vasconcelos, 1885c) e "Dialetos Interamnenses" (Vasconcelos, 1885b); o título *Mapa Dialectológico do Continente Português* (Vasconcelos, 1897d) no volume IV (Vasconcelos, 1929b) e os artigos "Dialetos Beirões" (Vasconcelos, 1884–1885b) e "Dialetos Transmontanos" (Vasconcelos, 1890–1892a, 1895a) foram integrados no volume VI (Vasconcelos, 1985). Apenas o *Mapa Dialectológico do Continente Português* (Vasconcelos, 1897d) foi integrado numa publicação diversa. Ficaram com uma publicação apenas três títulos, sendo um deles a sua tese de doutoramento, *Esquisse d'une Dialectologie Portugaise* (Vasconcelos, 1901).

4.6. **F) Galego**

No que concerne ao galego, Leite de Vasconcelos publicou oito títulos. O primeiro, em 1883–1884, "Literatura Popular Galega" (Vasconcelos, 1883–1884c), e o último, em 1909, "Miuçalhas Galegas" (Vasconcelos, 1909b). Quatro títulos foram publicados na década de 1880 e os restantes na década de 1900.

Publicou três títulos em separatas e seis títulos foram publicados nos volumes dos *Opúsculos*. No volume IV (Vasconcelos, 1929b) foram integrados os títulos "Literatura Popular Galega" (Vasconcelos, 1883–1884c), "Pronúncia Galega" (Vasconcelos, 1887–1889c), *Uma Crónica de 1404* (Vasconcelos, 1903c), "O Santuário de Terena" (Vasconcelos, 1915a) e "Miuçalhas Galegas" (Vasconcelos, 1909b). No volume VII (Vasconcelos, 1938c) foi integrado o artigo "Poesia Popular Galega" (Vasconcelos, 1883b). O título "Literatura Popular Galega" (Vasconcelos, 1883–1884c) foi também incluído no volume IV dos *Ensaios Etnográficos* (Vasconcelos, 1910). Apenas um título não foi contemplado com qualquer reimpressão ou reedição.

4.7. **G) Linguagens das raias portuguesa e espanhola**

Neste âmbito, foram publicados 19 títulos. O primeiro, em 1882, "O Dialeto Mirandês" (Vasconcelos, (1882d), e o último, em 1940, "Da Fala de Barrancos" (Vasconcelos, 1940a). Foram publicados quatro títulos na década de 1880, quatro na década de 1890, cinco na década de 1900, um na década de 1920, quatro na década de 1930 e um na década de 1940.

Foram publicadas sete separatas e integrados seis artigos no volume IV dos *Opúsculos* (Vasconcelos, 1929b): "O Dialeto Mirandês" (Vasconcelos, 1882d), "Línguas Raianas de Trás-os-Montes" (Vasconcelos, 1885–1886), "Linguagens Fronteiriças de Portugal e Espanha" (Vasconcelos, 1902c), "Sátira à Linguagem de Palaçoulo" (Vasconcelos, 1902g), "Silva Mirandesa" (Vasconcelos, 1902h) e "Linguagem Portuguesa de Alamedilla ou "Almedilha"" (Vasconcelos, 1930d). Foram três os títulos publicados em publicações diversas: "O Dialeto Mirandês" (Vasconcelos, 1882d), para o qual foi feita uma segunda edição, "Linguagens Fronteiriças" (Vasconcelos, 1902c) e "Da Fala de Barrancos" (Vasconcelos, 1940a). Ficaram com uma publicação apenas oito títulos.

4.8. **H) Um dialeto espanhol**

Para a presente secção, temos apenas o artigo "Tradições Populares e Dialeto da Estremadura Espanhola" (Vasconcelos, 1883–1884d), inicialmente publicado na *Revista de Estudos Livres* na década de 1880, integrado no volume IV dos *Opúsculos* (Vasconcelos, 1929b) e para o qual foi feita uma segunda edição.

4.9. **I) Provençal**

Nesta secção surgem apenas dois títulos, ambos publicados na década de 1900. Para ambos foram feitas separatas: "Canção de "Santa Fides" de Agen" (Vasconcelos, 1902a), publicada em Paris e datada de 1902, e "Notícia Bibliográfica do Poema Provençal de "Santa Fé"" (Vasconcelos, 1902d), publicada pela Imprensa da Universidade de Coimbra em 1902, e simultaneamente integrada no volume IV dos *Opúsculos* (Vasconcelos, 1929b).

4.10. **J) Vários textos**

Neste campo foi publicado apenas o título *Textos Românicos* (Vasconcelos, 1919, 1920c, 1921d), com três volumes publicados em 1919, 1920 e 1921, respetivamente, e dos quais não foram feitas reimpressões ou reedições.

Tabela 3

Filologia

	A)	B)	C) a)	C) b)	C) c)	D) a)	D) b)	D) c)	E)	F)	G)	H)	I)	J)
Uma Vez	—	1	27	8	19	4	15	—	3	1	8	—	—	1
Separatas	1	2	12	4	6	2	5	2	24	3	7	—	2	—

Várias	1	—	8	3	2	1	1	—	1	—	3	1	—	—
Ensaios Etnog.	—	—	—	—	—	—	—	—	—	1	—	—	—	—
Mês de Sonho	—	—	—	—	1	—	—	—	—	—	—	—	—	—
Opúsculos	3	4	74	7	8	10	36	4	6	6	6	1	1	—

5. Poligrafia

Também os títulos de poligrafia foram divididos em três secções: A) numismática, medalhística, tesserologia; B) biografias e C) outros trabalhos, que se analisam em seguida.

5.1. A) Numismática, medalhística, tesserologia

Retornamos a uma área em que Leite de Vasconcelos não conseguiu completar o seu plano para publicação de um volume dos *Opúsculos*. Deixou, no entanto, um plano a partir do qual se pode inferir o que teria pretendido reeditar. Analisa-se inicialmente o que de facto reimprimiu ou reeditou, para, em seguida, se analisar o que teria, provavelmente, sido globalmente reeditado pelo autor.

Publicou 35 títulos, o primeiro, "Numismática Nacional" (Vasconcelos, 1888c), em 1888, e o último, "Nomenclatura Numismática" (Vasconcelos, 1938a), em 1938. Publicou dois títulos na década de 1880, dez na década de 1890, dez na década de 1900, oito na década de 1910, quatro na década de 1920 e apenas um na década de 1930.

Foram 15 as separatas publicadas para estes títulos, contabilizando a separata "O Numismata Manuel Joaquim de Campos, Notícia Necrológica" (Vasconcelos, 1909c), publicada em Lisboa, em 1909, que não foi registada por Isabel Cepeda (1960), mas que consta no *Indículo* (Amzalak, 1924). Dois dos títulos foram publicados em publicações diversas ou tiveram uma edição em separado: *Coup d'Oeil sur la Numismatique en Portugal* (Vasconcelos, 1897b) e *Da Numismática em Portugal* (Vasconcelos, 1923a). O título "Sete Medalhas da Guerra Peninsular" (Vasconcelos, 1911b) foi parcialmente reeditado no volume III dos *Opúsculos* (Vasconcelos, 1931b) e 17 títulos ficaram apenas com uma edição.

Segundo o plano de Leite de Vasconcelos, o volume de numismática seguiria a seguinte organização: I. numismática propriamente dita, integrando os artigos: "Numismática Nacional" (Vasconcelos, 1888c), "Coup d'Oeil sur la Numismatique en Portugal" (Vasconcelos, 1897b), "Achados de Moedas Romanas da República" (Vasconcelos, 1909a) e "Se Há Moedas de Miranda do Douro" (Vasconcelos, 1917f); II. Medalhística: "Sete Medalhas da Guerra Peninsular" (Vasconcelos, 1911b), "Medalha da Sociedade Económica de Ponte de Lima" (Vasconcelos, 1913b), *Gratulação ao Prof. Stückelberg* (Vasconcelos, 1917d) e III. Tesserologia: "Contos para Contar" (Vasconcelos, 1900b). Teria ainda pensado integrar o artigo "O Numismata Manuel Joaquim de Campos, Notícia Necrológica" (Vasconcelos, 1909c), mas, não surgindo no plano de organização final, poderá ter equacionado adicionar ao volume de biografias, também planeado.

5.2. B) Biografias

Sabe-se que terá planeado um volume de biografias para os *Opúsculos*, no entanto, a existir um plano para o mesmo, este ainda não foi encontrado no espólio do autor.

Leite de Vasconcelos publicou 50 títulos no campo das biografias. O primeiro, "Ricardo Pinto de Mattos" (Vasconcelos, 1882e), em 1882, e o último, "Papeis de Aquiles Estaço" (Vasconcelos, 1940c), em 1940. Publicou seis títulos durante a década de 1880, três na década de 1890, 11 na década de 1900, 14 na década de 1910, 11 na década de 1920, quatro na década de 1930 e um na década de 1940.

Destes artigos foram publicadas 12 separatas. Os títulos "Elogio de Gaston Paris" (Vasconcelos, 1902–1909) e "Monsenhor Sebastião Rodolfo Dalgado" (Vasconcelos, 1921–1924) foram publicados em publicações diversas. Ficaram com uma publicação 37 títulos, número que seria certamente menor caso tivesse conseguido

concluir o projeto dos *Opúsculos*. Não tendo sido encontrado o projeto para esta secção, não se podem, para já, tirar conclusões sobre o que teria pretendido incluir.

5.3. C) Outros trabalhos

Neste âmbito, Leite de Vasconcelos publicou 49 títulos. O primeiro, "Convento de São João de Tarouca" (Vasconcelos, 1875b), em 1875, e o último, *Ucanha e o seu Pelourinho* (Vasconcelos, 1938d), em 1938. Publicou seis títulos na década de 1870, oito na década de 1880, cinco na década de 1890, três na década de 1900, 13 na década de 1910, dez na década de 1920 e quatro na década de 1930.

Foram publicadas quatro separatas e integrados onze títulos nos *Opúsculos*: foram integrados oito títulos no volume IV (Vasconcelos, 1929b), "Memórias de Mondim da Beira" (Vasconcelos, 1922b) foi parcialmente incorporado no volume III (Vasconcelos, 1931b), "Para a História da Museografia portuguesa" (Vasconcelos, 1915b) foi integrado no volume V (Vasconcelos, 1938b) e, finalmente, foi integrado o título "O Príncipe do Mónaco e a Etnologia" (Vasconcelos, 1920a) no volume VII (Vasconcelos, 1938c). Desta secção foram também incluídos três títulos em *Mês de Sonho* (de Vasconcelos, 1926a): "Brinde Proferido num Banquete em Ponta Delgada em Honra da Missão" (Vasconcelos, 1924d), simultaneamente com os dois artigos de "Política infausta" (Vasconcelos, 1924e) e incluído o artigo "Delimitação da fronteira portuguesa" (Vasconcelos, 1918–1919a) no volume I de *Etnografia Portuguesa* (Vasconcelos, 1933c). Foram reeditados três títulos em publicações diversas: *Dicionário da Corografia de Portugal* (Vasconcelos, 1883–1884a), do qual foi também feita uma segunda edição, "Instituto de Surdos-Mudos de Lisboa" (Vasconcelos, 1889b) e "Congresso da História das Religiões" (Vasconcelos, 1900–1901a). Foram publicados somente uma vez 29 títulos.

Tabela 4

Poligrafia

	A)	B)	C)
Uma vez apenas	17	37	29
Separatas	15	12	4
Publicações várias	2	2	3
Etnografia Portuguesa	—	—	1
Mês de Sonho	—	—	3
Opúsculos	1	—	11

6. Críticas bibliográficas

No que diz respeito a críticas bibliográficas, Leite de Vasconcelos publicou 125 títulos. O primeiro, "Henriqueta" (Vasconcelos, 1876), em 1876, e o último, "As Origens da Cidade do Porto" (Vasconcelos, 1935a), em 1935. Publicou um título na década de 1870, 60 na década de 1880, 19 na década de 1890, 39 na década de 1900, dois na década de 1910, dois na década de 1920 e dois na década de 1930.

Para estes títulos foram publicadas quatro separatas e integrados 13 artigos nos *Opúsculos*. No volume I (Vasconcelos, 1928a) foram incluídos os títulos "Uma Obra Inédita do Condestável D. Pedro de Portugal" (Vasconcelos, 1900–1901b) e "*Zeitschrift für Romanische Philologie*" (Vasconcelos, 1906f); no volume III (Vasconcelos, 1931b) foram incluídos também dois títulos: "Portugália" (Vasconcelos, 1906d) e "*Bulletin de Dialectologie Romane*" (Vasconcelos, 1912a); no volume IV (Vasconcelos, 1929b) integrou quatro títulos: "Dicionário de Algumas Palavras, Frases e Sentenças Peregrinas" (Vasconcelos, 1884a), "*Historia de la Literatura Gallega*" (Vasconcelos, 1887–1889a), "*Volvoretas*" (Vasconcelos, 1889c) e "*Grammaire des*

Langues Romanes" (Vasconcelos, 1890–1892b); no volume V (Vasconcelos, 1938b) são dois os títulos integrados: "*Recherches Ethnographiques sur la Salive et le Crachat*" (Vasconcelos, 1897–1899) e "Boletim da Sociedade Arqueológica de Santos Rocha" (Vasconcelos, 1904a); finalmente, no volume VII (Vasconcelos, 1938b) foram integrados três títulos: "*Mélusine*" (Vasconcelos, 1890–1892c), "*Tausend Portugiesische Sprichwörter*" (Vasconcelos, 1906e) e "As Origens da Cidade do Porto" (Vasconcelos1935a). No que diz respeito aos *Ensaios Etnográficos* foram integrados seis títulos: um título no volume III (Vasconcelos, 1906a), "Revista de Etnologia e de Glotologia" (Vasconcelos, 1880–1881c), e cinco no volume IV (Vasconcelos, 1910): "*Blason Populaire de la France*" (Vasconcelos, 1884–1885a), "Revista de Guimarães" (Vasconcelos, 1884f), "*Quinientas Comparaciones Populares Andaluzas*" (Vasconcelos, 1885e), "Cancioneiro Popular Político" (Vasconcelos, 1891a) e "Mil Trovas Populares Portuguesas" (Vasconcelos, 1904c). Em publicações diversas reeditou seis títulos e são 99 as críticas publicadas apenas uma vez.

7. Títulos por décadas

Leite de Vasconcelos dá início às suas publicações em periódicos no ano de 1875, inicialmente na área da poesia e da literatura, não podendo deixar de se mencionar o romance que não chegou a concluir. Os anos 80 do século XIX, ainda antes de se dedicar aos temas que iriam ocupá-lo até aos últimos dias de vida, são a época mais produtiva nesta área, tendo, inclusive, compilado as suas poesias nas *Baladas do Ocidente* (Vasconcelos, 1885a) e em *Nuvens* (Vasconcelos, 1898). Não obstante, mais de metade do que escreveu nesta área foi publicado apenas uma vez.

No campo da etnografia portuguesa publicou um total de 625 títulos. A década de 80 do século XIX concentra o maior número de publicações no âmbito do Folclore (69 títulos); nas décadas de 1890 e 1900 encontra-se uma maior concentração de títulos no âmbito da etnografia lusitana (11 títulos) e da arqueologia (85 e 72 títulos, respetivamente). O Museu Etnológico foi alvo de mais publicações na década de 1910 (cinco títulos) e a área da ergologia moderna na década seguinte (73 títulos), tendo sido os seus artigos maioritariamente publicados apenas uma vez no *Boletim de Etnografia* (Vasconcelos, 1920–1937). Na secção de generalidades, da terra e do povo, a década de 30 do século XX foi a mais produtiva (quatro, doze e oito títulos, respetivamente), sendo também do final da sua vida a sua dedicação quase total à etnografia portuguesa, com os primeiros volumes a serem publicados entre 1933 e 1941. Reeditou 20 títulos na *Etnografia Portuguesa* (Vasconcelos, 1933c, 1936, 1941b), 56 títulos nos *Ensaios Etnográficos* (Vasconcelos, 1891b, 1903a, 1906a, 1910), oito na *História do Museu Etnológico Português* (Vasconcelos, 1915a) e dois nas *Religiões da Lusitânia* (Vasconcelos, 1897, 1905, 1913d). Em *De Terra em Terra* (Vasconcelos, 1927b, 1927c) foram reeditados 20 títulos e em *Mês de Sonho* (Vasconcelos, 1926a) foram incluídos cinco títulos. No que concerne aos volumes de etnografia dos *Opúsculos* (Vasconcelos, 1928a, 1928b, 1929b, 1931b, 1938b, 1938c, 1985), foram incluídos, mesmo que apenas parcialmente, 131 títulos.

No que diz respeito à filologia, publicou 281 títulos. Nesta área, a década de 80 do século XIX foi maioritariamente dedicada à glotologia geral (quatro títulos), sendo a época na qual se inclui a sua tese de licenciatura, à história e crítica (seis títulos), à dialetologia portuguesa (oito títulos) e ao galego e dialetos espanhóis (quatro e um títulos, respetivamente). Os anos de 1900 são a época mais produtiva no âmbito da literatura (12 títulos), do provençal (dois títulos) e das linguagens das raias portuguesa e espanhola (cinco títulos). É também de 1901 a sua tese de doutoramento. A década de 10 do século XX reflete uma maior publicação de títulos na área do latim (dois títulos) e dos trabalhos doutrinais (26 títulos), que se prolonga até à década seguinte (22 títulos), e durante a qual se dedica também à área da antroponímia e da toponímia (sete e 16 títulos, respetivamente). Incluiu, mesmo que parcialmente, 166 títulos nos dois volumes de filologia, nos dois volumes de dialetologia e no volume de onomatologia dos *Opúsculos*.

No que concerne à numismática, medalhística e tesserologia, são 134 os títulos publicados. As suas décadas mais produtivas no âmbito da numismática são os anos 90 do século XIX e os anos de 1900 (dez títulos, respetivamente), sendo, no entanto, da década de 20 do século XX a publicação da sua obra mais extensa, *Da*

Numismática em Portugal (Vasconcelos, 1923a). No campo das biografias e outros trabalhos, é a década de 10 do século XX a mais produtiva (14 e 13 títulos, respetivamente).

Finalmente, no que diz respeito às críticas bibliográficas, publicou um total de 125 títulos, sendo 60 da década de 1880.

Apresenta-se a tabela dos títulos publicados distribuídos por décadas.

Tabela 5

Títulos por Décadas

	1870	**1880**	**1890**	**1900**	**1910**	**1920**	**1930**	**1940**	**?**	**Global**
Poesia	8	38	12	4	2	2	1	—	—	67
Etnografia portuguesa	4	88	116	111	94	148	57	5	2	625
Filologia	2	53	35	51	50	53	31	6	—	281
Poligrafia	6	16	18	24	35	25	9	1	—	134
Críticas bibliográficas	1	60	19	39	2	2	2	—	—	125
Totais	21	255	200	229	183	230	100	12	2	1232

8. Conclusão

A análise até aqui desenvolvida evidencia que Leite de Vasconcelos não foi apenas um autor no sentido tradicional, mas também um estratega na circulação de conhecimentos e um editor de si mesmo. Para compreender o alcance e o significado das suas decisões editoriais, seja através da publicação de separatas, da integração de artigos noutras obras ou da fundação de revistas, é necessário convocar uma reflexão teórica sobre os modos de construção da autoria científica e sobre o papel das práticas editoriais na construção da identidade científica. Em Leite de Vasconcelos, os gestos de separar, reformular e reeditar não são neutros. Sendo ou não conscientes, são movimentos de autoria estratégica, que permitem a Leite de Vasconcelos construir a sua autoridade científica, redes de circulação das suas obras e uma biblioteca de autor. Estas formas de publicação não se encontram limitadas pela lógica editorial da época, mas constituem, na obra leitiana, formas estratégicas de autoria, em que se conjugam continuidade e reconhecimento. No autor Leite de Vasconcelos, autoria e edição coexistem no momento em que toma para si mesmo o papel de editor, decidindo o que deve ser perpetuado e a forma como tal deve acontecer.

Segundo Foucault, a função autor é uma "*opération complexe qui construit un certain être de raison qu'on appelle l'auteur*" (Foucault, 1994, pp. 800–801). Nesta operação, a presença do editor é fundamental, em particular no processo de transformação do texto em livro, destacando também a diversidade de egos que coexistem num mesmo texto (Foucault, 1994, pp. 803–804). Este entendimento reforça a ideia de que a autoria está profundamente ligada à edição e circulação da obra. De forma complementar, Moreira e Magalhães afirmam que "a construção histórica e a atual noção de autor, no âmbito do livro literário, pressupõem a existência de um livro editado e publicado" (Moreira & Magalhães, 2023, p. 112), e citando Chartier reforçam o facto de ser essencial para o autor "fazer circular as suas obras entre o público, por meio da impressão" (Moreira & Magalhães, 2023, p. 113) para se afirmar enquanto tal. Também Ragusa e Girotto, no editorial "Intercâmbios finisseculares: o português entre norma e tradução" reforçam que:

> Na viragem do século XIX para o XX vive-se um momento de transformação que se revelaria fundamental para os desenvolvimentos posteriores, tanto em Portugal como no Brasil, na vida cultural e artística, bem como no âmbito da mudança linguística. No cenário europeu finissecular […] a pesquisa filológica, o interesse etnográfico e, antes de mais, a contaminação e a integração linguísticas, são sintomas abrangentes

de um constante intercâmbio, marcando um momento áureo também na projeção, ou até na expansão, da cultura portuguesa […]. (Ragusa & Girotto, 2024, p. 1)

Neste sentido, a bibliografia crítica contemporânea permite situar o percurso leitiano num quadro mais vasto de práticas autorais no campo científico e humanístico europeu.

Leite de Vasconcelos inicia a publicação dos seus estudos com apenas 17 anos. Ao longo da sua vida, publicou um total de 1240 títulos, contabilizando, agora, os títulos considerados compilações, porque, na verdade, além de incluírem artigos previamente publicados, incorporam títulos inéditos e todo o processo de investigação e escrita necessários. De facto, estas compilações não são apenas um agrupamento de artigos, mas novos objetos autorais, para os quais foi necessário tomar decisões textuais, estruturais e temáticas.

Esta análise permite não só ter uma noção da produtividade de Leite de Vasconcelos, mas também uma melhor perspectiva da vontade de abranger diversas áreas científicas. Mais do que quantidade, está em causa a coerência subjacente a essa produtividade. O fio condutor das suas investigações, a origem do povo português, é feito através de investigações paralelas e complementares, como a língua, o folclore ou os vestígios arqueológicos, pelo que a sua obra é mais facilmente compreendida quando analisada na totalidade. Como bem observou Manuel Heleno, respondendo aos que viam dispersão na obra de Leite de Vasconcelos:

> Tem-se afirmado a falta de plano, a incapacidade construtiva do saudoso Mestre. Os que lidaram com ele ou conhecem os seus papéis sabem quanto as preocupações de sistematização eram nele dominantes. Simplesmente a unidade, na amplidão do conjunto, escapa a quem estiver fora do seu campo, a quem não tiver ou não quiser ter um conhecimento pleno da obra. (Heleno, 1960, p. 49).

Elaborou 233 separatas, o que corresponde a cerca de 19% do *corpus*. Foram maioritariamente publicadas em datas muito próximas da primeira impressão, sendo plausível que, mesmo as que não têm a indicação da tipografia ou casa impressora, tenham sido produzidas na tipografia onde foi impressa a publicação original, a pedido do autor. Este padrão sugere um claro objetivo de autonomia da sua obra e um desejo de circulação para lá das barreiras dos periódicos. O número é significativo e indica uma intenção de autonomia e controlo sobre a circulação dos seus textos, bem como a criação de uma rede de receção, através da captação de novos leitores. As separatas parecem, assim, assumir um papel estratégico na afirmação do autor como agente da sua própria difusão.

Também a fundação das revistas *O Pantheon* em 1880, com Mont'Alverne de Sequeira, da *Revista Lusitana* em 1887, de *O Arqueólogo Português* em 1895 e do *Boletim de Etnografia*, entre 1920 e 1937, permitiu a publicação dos seus estudos nas diversas áreas de forma autónoma. A sua participação ativa na fundação e direção das revistas reforça a ideia de autonomia e construção da sua própria identidade científica nos diversos campos do conhecimento.

Conforme foi possível verificar no capítulo anterior, Leite de Vasconcelos selecionou, reeditou e reutilizou os seus textos ao longo da vida, enquanto peças autónomas ou reintegrando-as em obras de maior envergadura, quase como estudos preliminares posteriormente desenvolvidos e complementados. A reutilização e reaproveitamento de textos publicados inicialmente de forma dispersa e posteriormente organizados e publicados de forma temática permitiu uma maior difusão dos mesmos. Para isso, os textos foram relidos, revistos e reinscritos num novo plano de leitura. O modo como efetuou a reedição de artigos publicados em periódicos demonstra uma preocupação em constituir uma obra definitiva concentrada, partindo de uma publicação inicial fragmentada. Para isso, assume o papel de editor de si mesmo, escolhendo o que deve ou não ser perpetuado e a forma como tal deveria acontecer. Ao reeditar os seus textos em obras temáticas integrais, como os *Opúsculos* ou a *Etnografia Portuguesa*, Leite de Vasconcelos solidifica o seu nome como figura de referência em diferentes áreas do saber como a etnografia, a linguística ou a arqueologia, sistematizando os conhecimentos adquiridos nas diferentes áreas e transformando textos elaborados para circulação em obras de referência. As compilações redefinem o estatuto dos artigos, tornando-os coerentes numa obra única e eliminando a dispersão que a publicação em periódicos implica.

Numa época em que a difusão massiva de informação era uma impossibilidade, a publicação de artigos em separatas assume um papel primordial na circulação do conhecimento. Este facto é salientado por Aileen Fyfe (2018): "The ways in which offprints circulated – whether requested by authors in locations where the journal was not available, or distributed strategically by the author to people s/he wanted to impress […]". As separatas permitiram-lhe divulgar o seu trabalho de forma autónoma, orientar o público alvo e o espaço de circulação do conhecimento, fazendo com que os textos escapassem ao contexto limitado das revistas ou jornais onde foram publicados. Tê-las-á, por isso, entregue a quem as solicitasse ou enviado como oferta a colegas no estrangeiro, construindo, desta forma, uma rede de conhecimentos académicos, consolidando a sua posição nesta rede e alcançando um público académico já reconhecido e do qual pretenderia ser parte integrante. Um testemunho direto da eficácia da estratégia de divulgação e circulação de Leite de Vasconcelos pode ser encontrado na correspondência com George C. Keidel (1906) e Murray Peabody Brush (1907), ambos professores na Johns Hopkins University, que agradecem o envio de um exemplar da sua separata *Fabulário Português*. Este tipo de reedição cria um novo horizonte de receção, novos potenciais leitores, e, não menos importante, funciona ainda como um objeto de internacionalização do seu trabalho. Este facto foi já salientado, apesar de num contexto diferente, por Ragusa e Girotto:

> Nova é também a maneira de olhar para a literatura em língua portuguesa por parte de escritores e filólogos (ou romanistas, como eram definidos) estrangeiros, alimentada por uma constante troca de cartas, livros, revistas, traduções e colaborações [...] criando e ampliando uma rede intelectual que atravessa duas gerações (Ragusa & Girotto, 2024, p. 1).

É possível, assim, vislumbrar um Leite de Vasconcelos editor de si próprio, com decisões a tomar quanto ao que desejaria reeditar, quanto à melhor forma de o fazer e, não menos importante, quanto ao melhor modo de fazer circular os seus textos. A integração destes dados com a análise das publicações leitianas reforça a ideia de que a transformação de textos dispersos em obras de referência e o controlo da circulação dos seus textos entre destinatários específicos o auxiliaram na criação de uma rede internacional de troca de conhecimentos.

Leite de Vasconcelos foi, por isso, exímio na divulgação da sua obra, tornando o seu nome reconhecido dentro e fora de Portugal, fazendo a sua obra circular não apenas num conjunto de textos publicados em periódicos, mas também como obra autónoma (separata), partilhando-a com quem desejava ser lido. As reedições e reimpressões dos seus textos não constituem uma simples prática recorrente, mas a construção consciente e sistemática de uma vida dedicada à ciência, com intenção de perpetuar a sua obra. Leite de Vasconcelos é o investigador, o autor e o seu próprio editor, acautelando a forma como pretendia que os seus textos fossem perpetuados. Neste enquadramento, a tríade revista, reedição e separata produziu autoridade, rede e permanência para o autor e para a obra. Esta política material do texto explica a presença de Leite de Vasconcelos no centro de redes nacionais e internacionais e clarifica por que razão a sua obra venceu a casualidade dos periódicos para se transformar num *corpus* de referência. O caso leitiano mostra, por fim, como, no início do século XX, a autoria científica se fez também através de decisões editoriais.

O presente estudo, tendo como ponto de partida o artigo de Isabel Cepeda, não o substitui, mas acrescenta uma visão transversal da forma como Leite de Vasconcelos escreveu e publicou ao longo de 66 anos de trabalho. A análise dos artigos reintegrados em novas obras e da publicação de separatas permite perceber como estas duas ações constituem pontos fulcrais para a divulgação e construção do lugar de Leite de Vasconcelos no mundo da ciência em Portugal. Permite ainda compreender que a obra impressa ao longo dos anos não é um simples gesto de autoria e edição, mas também de seleção e circulação, que permitem inserir o seu nome no campo das ciências em Portugal e no estrangeiro.

A obra de Leite de Vasconcelos, tendo cessado pela sua mão no momento da morte, foi prolongada no tempo pelos seus discípulos, a quem deu instruções para que a continuassem, tendo este desejo de continuidade

ficado inscrito no seu testamento (Coito & Coelho, 1988–1989, pp. 335–337).[2] Ao criar um plano para a edição da sua obra, Leite de Vasconcelos não apenas preserva, mas projeta uma memória científica que vai para além da sua morte e orienta futuros leitores. Das anotações existentes no espólio leitiano, percebe-se que a sua obra não se resume ao que publicou, mas que muito haverá ainda por conhecer. Um estudo mais aprofundado das obras e investigações futuras no seu espólio, em particular no campo da crítica genética, poderá trazer ainda novidades no que concerne aos trabalhos publicados, aos que ficaram por publicar ou mesmo obras planeadas que não chegaram a ser concluídas. Também o cruzamento das correspondências que se encontram já em fase de publicação[3] ampliará o mapa de circulação internacional das obras leitianas e irá certamente complementar um legado que julgávamos conhecer.

Referências

Alonso Montero, X. (2011). O soneto neolatino (Vila Nova de Famalicão, 1929–1933): Estudio dunha revista poética singular, especialmente das colaboracións galegas. *Anuario de Estudios Literarios Galegos*: 1995, 11–37. Reedição em Poesiagalega.org. https://www.poesiagalega.org/arquivo/ficha/f/1570

Amzalak, M. B. (1924). *Indículo dos trabalhos literários de J. Leite de Vasconcelos*. Oficinas Gráficas do Museu Comercial.

Bico, M. I. (Ed.). (2024). *Correspondência entre Rufino José Cuervo e Leite de Vasconcelos*. Imprensa Nacional-Casa da Moeda.

Brush, M. P. (1907, 13 janeiro). Carta para Leite de Vasconcelos. Correspondente 494, doc. 3333, Espólio de José Leite de Vasconcelos. Museu Nacional de Arqueologia, Lisboa.

Cepeda, I. V. (1960). Bibliografia de José Leite de Vasconcelos. In *José Leite de Vasconcelos: Livro do centenário (1858–1958)* (pp. 139–265). Imprensa Nacional.

Coito, L. C., & Coelho, J. T. P. (1988–1989). Nota acerca do legado do Doutor Leite de Vasconcelos no Museu Nacional de Arqueologia e Etnologia. *O Arqueólogo Português*, série IV, 6/7, 333–365.

Foucault, M. (1994). Qu'est-ce qu'un auteur? In *Dits et écrits II* : 1976–1988 (pp. 789–821). Gallimard.

Fyfe, A. (2018, 6 junho). Did authors get off-prints/separate copies? *Philosophical Transactions*. University of St Andrews. https://arts.st-andrews.ac.uk/philosophicaltransactions/did-authors-get-off-prints-separate-copies/

Heleno, M. (1960). Algumas palavras sobre Leite de Vasconcelos. In *José Leite de Vasconcelos: Livro do Centenário (1858–1958)* (pp. 45–51). Imprensa Nacional.

Keidel, G. C. (1906, 6 dezembro). Carta para Leite de Vasconcelos. Correspondente 1668, doc. 10740, Espólio de José Leite de Vasconcelos. Museu Nacional de Arqueologia, Lisboa.

Moreira, P. R., & Magalhães, F. D. (2023). O autor, o editor e a autopublicação: Construção histórica das categorias e reflexões sobre a sua prática. *ArtCultura*, *25*(46), 100–117. https://doi.org/10.14393/artc-v25-n46-2023-71188

PORBASE — Base Nacional de Dados Bibliográficos. (2024, 6 fevereiro). Catálogo bibliográfico da Biblioteca Nacional de Portugal. https://porbase.bnportugal.gov.pt/ipac20/ipac.jsp?profile=

Ragusa, A., & Girotto, A. (2024). Intercâmbios finisseculares: O português entre norma e tradução. *Cadernos de Tradução*, *44*(3), e103189. https://doi.org/10.5007/2175-7968.2024.e103189

Vasconcelos, J. L. de (s.d.a). Caixas 1, 2, 3 dos *Opúsculos* [Manuscritos]. Espólio de José Leite de Vasconcelos. Museu Nacional de Arqueologia, Lisboa.

Vasconcelos, J. L. de (s.d.b). A ideia de «fonte» na toponímia portuguesa. *Archivio Glottologico Italiano*, *XXI*, Sezione Bartoli.

[2] Para informações adicionais acerca do testamento de Leite de Vasconcelos, recomenda-se a consulta do artigo de Coito e Coelho (1988-1989)

[3] A correspondência do autor encontra-se em fase de publicação, sendo o volume mais recente a correspondência com Rufino Cuervo, editada por Maria Inês Bico (2024).

Vasconcelos, J. L. de (1875a). A naiade. *A Liberdade* (Viseu), *255*.
Vasconcelos, J. L. de (1875b). Convento de São João de Tarouca. *A Atalaia* (Viseu).
Vasconcelos, J. L. de (1876). *Henriqueta*, por Maria Peregrina de Sousa. *A Liberdade* (Viseu).
Vasconcelos, J. L. de (1878–1879). O Presbitério de Vila Cova. *O Académico* (Porto), *3*, *4*, *5*, *6*, 1878; *Aurora do Cávado*, *546*, *547*, *548*, *550*, *552*, *554*, *563*, *569*, *572*, *575*, *581*, *583*, *588*.
Vasconcelos, J. L. de (1879a). Mitologia popular portuguesa (Culto da noute). *Aurora do Cávado*, *597*.
Vasconcelos, J. L. de (1879b). Um inédito (Soneto atribuído a D. Fr. Marcelino José da Silva). *Correio do Ave*, *81*.
Vasconcelos, J. L. de (1880–1881a). Linguagem popular portuguesa. *O Pantheon*, *I*, 319–321.
Vasconcelos, J. L. de (1880–1881b). Notas de pré-história. *O Pantheon*, *I*, 301–303, 364–365.
Vasconcelos, J. L. de (1880–1881c). *Revista de etnologia e de glotologia*, dirigida por F. Adolfo Coelho (Lisboa, 1880). *O Pantheon*, *I*, 100–101.
Vasconcelos, J. L. de (1881). *Estudo etnográfico a propósito da ornamentação dos jugos e cangas dos bois nas províncias portuguesas do Douro e Minho*. *Jornal de Agricultura*.
Vasconcelos, J. L. de (1882a). As Maias (Costumes populares portugueses ― Cartas ao ilustre folclorista espanhol, o Sr. D. Rodrigues Marin). *Pero Galego*, *19*.
Vasconcelos, J. L. de (1882b). Currente calamo (Contra os poetastros). *O Crepúsculo*, *4*.
Vasconcelos, J. L. de (1882c). Duas etimologias [Chaminha e Vilarelho]. *Pero Galego*, *23*, *27*, *29*.
Vasconcelos, J. L. de (1882d). O dialecto mirandês (Notas glotológicas). *O Penafidelense*, *472*, *473*, *479*, *482*, *483*.
Vasconcelos, J. L. de (1882e). Ricardo Pinto de Mattos. *Tirocínio*, *14*.
Vasconcelos, J. L. de (1882f). Uma excursão ao Soajo. *Pero Galego*, n.º 31–32, outubro; *Tirocínio*, 1882.
Vasconcelos, J. L. de (1883a). Contribuições para o estudo da linguagem infantil. *Tirocínio*, *77*.
Vasconcelos, J. L. de (1883b). Poesia popular galega. *Anuário para o Estudo das Tradições Populares Portuguesas*, *1*, 37–38.
Vasconcelos, J. L. de (1883c). Sub-dialecto alentejano: Estudo glotológico. *Elvense*.
Vasconcelos, J. L. de (1883–1884a). *Dicionário da corografia de Portugal*. Livraria Portuense de Clavel e C.ª.
Vasconcelos, J. L. de (1883–1884b). Fonética da linguagem infantil (Ligeiras observações). *Revista de Estudos Livres*, *I*(8), 379–382.
Vasconcelos, J. L. de (1883–1884c). Literatura popular galega. *El Folk-Lore Frexnense y Bético-Extremeño* (Fregenal), 96–102.
Vasconcelos, J. L. de (1883–1884d). Tradições populares e dialecto da Estremadura espanhola. *Revista de Estudos Livres*, *1*, 88–94.
Vasconcelos, J. L. de (1884a). *Dicionário de algumas palavras, frases e sentenças peregrinas, traduzidas e explicadas* por Narciso José de Morais (Porto, 1884). A Discussão.
Vasconcelos, J. L. de (1884b). Em Trás-os-Montes. *A Voz do Douro*, *259*.
Vasconcelos, J. L. de (1884c). Linguagem popular de Barcelos (Introdução a C. A. Landolt, *Tradições Populares de Barcelos*). Tipografia da *Aurora do Cávado*.
Vasconcelos, J. L. de (1884d). Nomes gentílicos (A propósito de «paredense»). *O Paredense*, *15*.
Vasconcelos, J. L. de (1884e). Patologia da linguagem. *A Saúde Pública* (Porto).
Vasconcelos, J. L. de (1884f). *Revista de Guimarães*. *A Discussão*.
Vasconcelos, J. L. de (1884–1885a). *Blason populaire de la France*, por H. Gaidoz et Paul Sébillot (Paris, 1884). *Revista de Estudos Livres*, *2*, 413–416.
Vasconcelos, J. L. de (1884–1885b). Dialetos beirões (Contribuições para o estudo da dialetologia portuguesa. *Revista de Estudos Livres*, *2*, 81–91, 215–228.
Vasconcelos, J. L. de (1885a). *Baladas do Ocidente*. Livraria Portuense Editora.
Vasconcelos, J. L. de (1885b). Dialetos interamnenses: Contribuições para o estudo da dialetologia portuguesa. *Revista de Guimarães*, *2*.

Vasconcelos, J. L. de (1885c). Dialetos minhotos. *Revista de Guimarães*, *2*.
Vasconcelos, J. L. de (1885d). *Portugal pré-histórico* (Vol. 106). David Corazzi, Biblioteca do Povo e das Escolas.
Vasconcelos, J. L. de (1885e). *Quinientas comparaciones populares andaluzas*, por F. Rodriguez Marín (Osuna, 1884). *Revista do Minho*, *1*, 13–15.
Vasconcelos, J. L. de (1885–1886). Línguas raianas de Trás-os-Montes (Sucintas notas). *Revista de Estudos Livres*, *3*, 374–384.
Vasconcelos, J. L. de (1886). *A evolução da linguagem*. Tipografia Ocidental.
Vasconcelos, J. L. de (1887). A nacionalidade portuguesa. *A Apoteose, jornal comemorativo do septicentenário e inauguração da estátua de D. Afonso Henriques, número único*, *5*. Lisboa.
Vasconcelos, J. L. de (1887–1889a). *Historia de la literatura gallega*, por Augusto G. Besada (Corunha, 1887). *Revista Lusitana*, *I*, 183–185.
Vasconcelos, J. L. de (1887–1889b). Pronúncia do «H» latino no século V. *Revista Lusitana*, *I*, 73.
Vasconcelos, J. L. de (1887–1889c). Pronúncia galega. *Revista Lusitana*, *I*.
Vasconcelos, J. L. de (1888a). A reforma do Curso Superior de Letras e a filologia portuguesa. *O Correio do Norte*, *2483*, *2484*, *2488*, *2490*, *2493*, *2497*.
Vasconcelos, J. L. de (1888b). Em Évora. *O Repórter*, *180*, *200*.
Vasconcelos, J. L. de (1888c). *Numismática nacional (Lição inaugural do curso de Numismática da Biblioteca Nacional de Lisboa, no ano letivo de 1888–1889)*. Tipografia de *O Dia*.
Vasconcelos, J. L. de (1889a). Crítica literária: *Os Lusíadas* de Camões (Edição crítica e anotada em todos os lugares duvidosos, restituindo, quanto possível, o texto primitivo pela correção de erros que nunca se tinham expungido), por Francisco Gomes de Amorim. *O Dia*, *546*, *549*, *556*, *563*, *570*, *576*.
Vasconcelos, J. L. de (1889b). *Instituto de Surdos-Mudos de Lisboa*. *O Dia*.
Vasconcelos, J. L. de (1889c). *Volvoretas*, por Alberto Garcia Ferreiro (Orense, 1887). *O Repórter*, *493*.
Vasconcelos, J. L. de (1890). Borges de Figueiredo e a arqueologia portuguesa (Notícia necrológica lida na sessão de 24 de outubro de 1890 da secção das Ciências Étnicas da Sociedade de Geografia de Lisboa). *O Dia*, *980*, *981*.
Vasconcelos, J. L. de (1890–1892a). Dialetos transmontanos (Contribuições para o estudo da dialetologia portuguesa). *Revista Lusitana*, *II*, 97–120.
Vasconcelos, J. L. de (1890–1892b). *Grammaire des langues romanes*, por W. Meyer Lübke (Vol. 1, trad. franç. por E. Rabiet, Paris, 1890). *Revista Lusitana*, *II*, 364–376.
Vasconcelos, J. L. de (1890–1892c). *Mélusine (Revue de mythologie, littérature populaire, traditions et usages*, dirigée par H. Gaidoz et E. Roland, II, 1885). *Revista Lusitana*, *II*, 376–377.
Vasconcelos, J. L. de (1891a). *Cancioneiro popular político*, por A. Tomás Pires (com uma carta de Oliveira Martins, Elvas, 1891). *O Dia*.
Vasconcelos, J. L. de (1891b). *Ensaios etnográficos* (Vol. 1). Esposende.
Vasconcelos, J. L. de (1891c). *Lições práticas de linguagem portuguesa*, por Cândido de Figueiredo. *O Dia*, *1217*, *1218*, *1219*, *1223*, *1227*, *1228*, *1229*, *1234*.
Vasconcelos, J. L. de (1891c). *O gralho depenado (Réplica às «caturrices» filológicas do Sr. Cândido de Figueiredo)*. Tipografia do jornal *O Dia*.
Vasconcelos, J. L. de (1892). Inscrição inédita de Mercúrio em Moura e vários costumes sepulcrais da época romana em Portugal. *Stemma*, número único.
Vasconcelos, J. L. de (1894a). Notícias transmontanas do princípio do século XVII. *O Mirandês*, *8*.
Vasconcelos, J. L. de (1894b). *Quid apud Lusitanos verbum «Aedeoli» significaverit*. Imprensa Libânio da Silva.
Vasconcelos, J. L. de (1895a). Dialetos transmontanos (Contribuições para o estudo da dialetologia portuguesa). *Revista Lusitana*, *III*, 57–74.
Vasconcelos, J. L. de (1895b). Excursão Arqueológica a Alcácer-do-Sal. *O Arqueólogo Português*, *I*, 65–92.

Vasconcelos, J. L. de (1895c). Museu Etnográfico Português [Plano de organização]. *Revista Lusitana*, *III*, 193–250.
Vasconcelos, J. L. de (1896). Antiguidades da Beira (Diário de um excursionista) (Fragmentos). *Gazeta da Figueira*, *486*, *487*, *492*, *493*, *494*.
Vasconcelos, J. L. de (1897a). *Annotationes ad Geographiam Lusitanam*. *Rivista di Storia Antica e Scienze Affine*, *II*.
Vasconcelos, J. L. de (1897b). *Coup d'œil sur la numismatique en Portugal*. *Gazette Numismatique Française*, 484–497.
Vasconcelos, J. L. de (1897c). Estudos sobre Panóias. *O Arqueólogo Português*, *III*, 58–61.
Vasconcelos, J. L. de (1897d). Mapa dialectológico do continente português. In F. Deusdado, *Corografia de Portugal*. Guillard, Aillud & C.ª.
Vasconcelos, J. L. de (s. d. [1897e]). *Museu Etnológico Português*. Tipografia de *O Dia*.
Vasconcelos, J. L. de (1897f). Museu Municipal de Bragança. *O Arqueólogo Português*, *III*, 48–58.
Vasconcelos, J. L. de (1897g). *Religiões da Lusitânia* (Vol. 1). Imprensa Nacional.
Vasconcelos, J. L. de (1897–1899). *Recherches ethnographiques sur la salive et le crachat*, por Camille de Mensignac (Bordéus, 1892). *Revista Lusitana*, *V*, 236–237.
Vasconcelos, J. L. de (1898). *Nuvens*. Livraria Chardron, de Lello & C.ª.
Vasconcelos, J. L. de (1899). Antiquités du sud de Portugal. *Bulletin de la Société des Antiquaires de France*, 269–273.
Vasconcelos, J. L. de (1900a). Arqueologia do Alto-Minho (Dádiva ao Museu Etnológico Português). *O Arqueólogo Português*, *V*, 33.
Vasconcelos, J. L. de (1900b). Contos para contar (Introdução a uma lista elaborada por Júlio Meili, de Zurique). *O Arqueólogo Português*, *V*, 52–54.
Vasconcelos, J. L. de (1900c). *Les monnais de la Lusitanie portugaise*. In *Procès-verbaux et Mémoires du Congrès International de Numismatique, réuni à Paris, 1900*, 63–78.
Vasconcelos, J. L. de (1900d). *Onomasticon Lusitanien*. *Revue Celtique*, *XXII/XXI*, 307–311.
Vasconcelos, J. L. de (1900–1901a). Congresso da história das religiões. *Revista Lusitana*, *VI*, 86–87; *O Arqueólogo Português*, *V*, 123–128.
Vasconcelos, J. L. de (1900–1901b). Uma obra inédita do Condestável D. Pedro de Portugal, por Carolina Michaelis de Vasconcelos (Madrid, 1899). *Revista Lusitana*, *VI*, 93–94.
Vasconcelos, J. L. de (1901). *Esquisse d'une dialectologie portugaise*. Aillaud et C.ie.
Vasconcelos, J. L. de (1902a). Canção de «Santa Fides» de Agen (Texto com introdução, notas e glossário). *Romania*, *31*.
Vasconcelos, J. L. de (1902b). *Les Celtes de la Lusitanie portugaise*. *Revue Celtique*, *XXIII*, 74–82.
Vasconcelos, J. L. de (1902c). Linguagens fronteiriças de Portugal e Espanha. *Revista Lusitana*, *VII*, 133–145.
Vasconcelos, J. L. de (1902d). Notícia bibliográfica do poema provençal de «Santa Fé». *O Instituto*, *49*, 497–510.
Vasconcelos, J. L. de (1902e). O cantador de Setúbal. *Folha da Saudação aos 82 Anos de António Maria Calafate (O cantador de Setúbal)*, número único.
Vasconcelos, J. L. de (1902f). Poesia e etnografia (A propósito do *Alívio de Tristes*, do Sr. Correia de Oliveira). *Sociedade Futura*, *I* (3;7).
Vasconcelos, J. L. de (1902g). Sátira à linguagem de Palaçoulo. *Revista Lusitana*, *VII*, 148–149.
Vasconcelos, J. L. de (1902h). Silva Mirandesa. *Revista Lusitana*, *VII*, 282–302.
Vasconcelos, J. L. de (1902i). Uma Viagem à Beira (Fragmento). *A Folha*.
Vasconcelos, J. L. de (1902–1909). Elogio de Gaston Paris. *Boletim da Segunda Classe da Academia*, 52–53.
Vasconcelos, J. L. de (1903a). *Ensaios etnográficos* (Vol. 2). Esposende.
Vasconcelos, J. L. de (1903b). *Geografia da Lusitânia na época proto-histórica*. Imprensa Nacional.
Vasconcelos, J. L. de (1903c). *Uma crónica de 1404*. Oficina Tipográfica.

Vasconcelos, J. L. de (1903–1905a). Amaral. *Revista Lusitana*, *VIII*, 226.
Vasconcelos, J. L. de (1903–1905b). *Textos arcaicos* (Para uso da aula de Filologia Portuguesa estabelecida na Biblioteca Nacional de Lisboa, por portaria de 31 de dezembro de 1903). *Revista Lusitana*, *VIII*, 187–214.
Vasconcelos, J. L. de (1904a). Boletim da Sociedade Arqueológica de Santos Rocha (Figueira, 1904, n.º 1). *O Arqueólogo Português*, *IX*, 142–144.
Vasconcelos, J. L. de (1904b). Esopo Adelfo. *Revista Pedagógica*, *I*, 388–390.
Vasconcelos, J. L. de (1904c). *Mil trovas populares portuguesas*, por Agostinho de Campos e Alberto de Oliveira (Lisboa, 1903). *O Século*.
Vasconcelos, J. L de (1904–1905). *Notice sommaire sur le Musée Ethnologique Portugais*. Lisbonne.
Vasconcelos, J. L. de (1905a). O santuário de Terena. *O Arqueólogo Português*, *X*, 340–343.
Vasconcelos, J. L. de (1905b). *Religiões da Lusitânia* (Vol. 2). Imprensa Nacional.
Vasconcelos, J. L. de (1905c). *Signification religieuse, en Lusitanie, de quelques monnais percées d'un trou (Mémoire lu au Congrès Internacional d'Archéologie, session d'Athènes, dans la séance du 10 avril, 1905)*. *O Arqueólogo Português*, *X*, 169–176.
Vasconcelos, J. L. de (1906a). *Ensaios etnográficos* (Vol. 3). Imprensa Lucas.
Vasconcelos, J. L. de (s. d. [1906b]). *Musée Ethnologique Portugais — Belém*. Lisboa.
Vasconcelos, J. L. de (s. d. [1906c]). *Plano sumário do Museu Etnológico Português*. Lisboa.
Vasconcelos, J. L. de (1906d). *Portugália* (Materiais para o estudo do povo português), redigida por Ricardo Severo. *O Arqueólogo Português*, *XI*, 321–379.
Vasconcelos, J. L. de (1906e). *Tausend portugiesische Sprichwörter*, por Carolina Michaelis de Vasconcelos. *Revista Lusitana*, IX, 182–186.
Vasconcelos, J. L. de (1906f). *Zeitschrift für Romanische Philologie*, Halle, 28, 1904. *Revista Lusitana*, *IX*, 396–398.
Vasconcelos, J. L. de (1909a). Achados de moedas romanas da República. *O Arquéologo Português*, *XIV*, 58–59.
Vasconcelos, J. L. de (1909b). Miuçalhas galegas. *In Mélanges offerts à M. Maurice Vilmotte* (Vol. 1, 317–327). Paris.
Vasconcelos, J. L. de (1909c). *O numismata Manuel Joaquim de Campos: Notícia necrológica*. Lisboa.
Vasconcelos, J. L. de (1910). *Ensaios etnográficos* (Vol. 4). Livraria Clássica Editora.
Vasconcelos, J. L. de (1910–1911). *Visita ao Museu Etnológico Português*. Imprensa Nacional.
Vasconcelos, J. L. de (1911a). Da importância do latim (Lição inaugural na Faculdade de Letras da Universidade de Lisboa). *Revista Lusitana*, *XIV*, 113–124.
Vasconcelos, J. L. de (1911b). Sete medalhas da Guerra Peninsular (Existentes no Gabinete Numismático da Biblioteca Nacional de Lisboa). *O Arqueólogo Português*, *XVI*, 139–174.
Vasconcelos, J. L. de (1911c). Uma expressão popular portuguesa: «Mais vale um gosto que quatro vinténs». *Revue de Dialectologie Romane*, *V*, 225–227.
Vasconcelos, J. L. de (1912a). *Bulletin de Dialectologie Romane*, *4*(2), 1914 [Artigo de Paul Barbier em que se debatem as etimologias de «mocho» e «Roriça»]. *Revista Lusitana*, *XV*, 360.
Vasconcelos, J. L. de (1912b). *Significação do Museu Etnológico Português*. Lisboa.
Vasconcelos, J. L. de (1912–1913). Discussão filológica: A palavra «Momo». *Boletim da Segunda Classe da Academia*, *VII*, 222–224.
Vasconcelos, J. L. de (1913a). Geografia antiga do Algarve. *Lucta*.
Vasconcelos, J, L. de (1913b). Medalha da Sociedade Económica de Ponte de Lima (Século XVIII). *O Arqueólogo Português*, *XVIII*, 102–106.
Vasconcelos, J. L. de (1913c). «Pedro» e «Pedra». *Limiana*, 127–129.
Vasconcelos, J. L. de (1913d). *Religiões da Lusitânia* (Vol. 3). Imprensa Nacional.
Vasconcelos, J. L. de (1914–1915a). O «Dicionário da Academia». *Boletim da Segunda Classe da Academia*, *IX*, 468–471.

Vasconcelos, J. L. de (1914–1915b). Objetos paleolíticos do Casal do Monte (Oferecidos ao Museu da Academia das Ciências de Lisboa). *Boletim da Segunda Classe da Academia*, *VIII*/*IX*, 390–395.
Vasconcelos, J. L. de (1915a). *História do Museu Etnológico Português*. Imprensa Nacional.
Vasconcelos, J. L. de (1915b). Para a história da museografia portuguesa. *O Arqueólogo Português*, *XVIII*, 167–168.
Vasconcelos, J. L. de (1915c). Três anos de latim na Faculdade de Letras da Universidade de Lisboa (1911–1912, 1912–1913, 1913–1914). *Arquivos da Universidade de Lisboa*, *II*, 233–240.
Vasconcelos, J. L. de (1916a). Cacografia dialectal. *Revista Lusitana*, *XIX*, 337.
Vasconcelos, J. L. de (1916b). Entre Tejo e Odiana. *O Arqueólogo Português*, *XXI*, 152–195.
Vasconcelos, J. L. de (1917a). Coisas velhas. *O Arqueólogo Português*, *XXII*, 107–169.
Vasconcelos, J. L. de (1917b). *Excursão Interamnense (Vila do Conde e Póvoa, 1895)*. Imprensa Nacional.
Vasconcelos, J. L. de (1917c). *Excursão Extremenha*. Lisboa.
Vasconcelos, J. L. de (1917d). *Gratulação ao Prof. Stückelberg*. Tipografia do Comércio.
Vasconcelos, J. L. de (1917e). Por Trás-os-Montes. *O Arqueólogo Português*, *XXII*, 1–53.
Vasconcelos, J. L. de (1917f). Se há moedas de Miranda do Douro. *O Arqueólogo Português*, *XXII*, 105–107.
Vasconcelos, J. L. de (1917–1918). Nótulas etnográficas. *Lusa*, *I*(11, 21), 81–161.
Vasconcelos, J. L. de (1918a). Amostra toponímia portuguesa. *Revista Lusitana*, *XXI*, 58–63.
Vasconcelos, J. L. de (1918b). Enquissas. *Revista Lusitana*, *XXI*, 316–336.
Vasconcelos, J. L. de (1918–1919a). Delimitação da fronteira portuguesa (Notícia histórica). *Boletim da Segunda Classe da Academia*, *XIII*, 1275–1292.
Vasconcelos, J. L. de (1918–1919b). Safira (freguesia condal do Alentejo): Estudo etimológico. *Boletim da Segunda Classe da Academia*, XIII, 1061–1066.
Vasconcelos, J. L. de (1919a). Amostra toponímia portuguesa. *Revista Lusitana*, *XXII*, 197–199.
Vasconcelos, J. L. de (1919b). Importância da Etnografia. *Revista Lusitana*, *XXII*, 5–18.
Vasconcelos, J. L. de (1919c). *Sinopse do Museu Etnológico Português*. Tipografia Moderna.
Vasconcelos, J. L. de (1919). *Textos românicos* (Vol. 1). Tipografia do Comércio.
Vasconcelos, J. L. de (1919–1920). Discussão filológica: Perneta. *Lusa*, *II*/*III* (48), 17–18.
Vasconcelos, J. L. de (1920a). O Príncipe do Mónaco e a etnologia. *A Pátria*.
Vasconcelos, J. L. de (1920b). Pôr freio alguém. s.l.
Vasconcelos, J. L. de (1920c). *Textos românicos* (Vol. 2). Tipografia de J. Teixeira.
Vasconcelos, J. L. de (1920d). Trás-os-Montes Etnográfico (Memória apresentada ao Congresso Transmontano de 1920) [Resumo]. In *Conclusões das Teses*, 24–25.
Vasconcelos, J. L. de (1920–1921). Nomes de pessoas tornados geográficos em Portugal. *Boletim da Segunda Classe da Academia*, *XV*/*XVI*, 785–822.
Vasconcelos, J. L. de (1920–1937). *Boletim de Etnografia* (Publicação do Museu Etnológico Português). Imprensa Nacional.
Vasconcelos, J. L. de (1921a). Etimologia de um nome ilustre. *Memoriam do Juiz Pinto Osório*. Porto.
Vasconcelos, J. L. de (1921b). Geografia Tradicional da Beira (Memória apresentada ao 1.º Congresso Beirão, em Viseu). [s. l., s. d.; publicado em Sintra].
Vasconcelos, J. L. de (1921c). Publicação de textos antigos. *Anais das Bibliotecas e Arquivos*, *6*, 163–165.
Vasconcelos, J. L. de (1921d). *Textos românicos* (Vol. 3). Tipografia Lusitânia.
Vasconcelos, J. L. de (1921–1922). Riba d'Ave: Estudo toponímico. *Boletim da Segunda Classe da Academia*, *XVI*, 572–587.
Vasconcelos, J. L. de (1921–1924). Monsenhor Sebastião R. Dalgado (Discurso lido à beira da sepultura). *Lusa*, *4*, 70–71.
Vasconcelos, J. L. de (1922a). Descobridores de Monsanto. *O Arqueólogo Português*, *XXV*, 178–180.
Vasconcelos, J. L. de (1922b). Memórias de Mondim da Beira (Plano de uma obra em preparação). *Boletim do 2.º Congresso Beirão*, (3).

Vasconcelos, J. L. de (1922c). Próloquio. *Revista de Arqueologia e História*, 5–6.
Vasconcelos, J. L. de (1923a). *Da numismática em Portugal* (Vol. 9). Arquivo da Universidade de Lisboa
Vasconcelos, J. L. de (1923b). Preito filológico, prestado a um insigne orador. In *Homenagem a António Cândido, prestada pela Academia das Ciências de Lisboa*. Academia das Ciências de Lisboa.
Vasconcelos, J. L. de (1924a). À Beira-Mar. *Correio dos Açores*, *1195*.
Vasconcelos, J. L. de (1924b). Algumas particularidades da vida de Antero. *Correio dos Açores*.
Vasconcelos, J. L. de (1924c). Aspetos do Minho. *Almanaque de Ponte de Lima*, 200–201.
Vasconcelos, J. L. de (1924d). Brinde proferido num banquete em Ponta Delgada em honra da missão. *Correio dos Açores*.
Vasconcelos, J. L. de (1924e). Política infausta [Acerca da finalidade da missão de homens de letras aos Açores]. *Diário de Lisboa*.
Vasconcelos, J. L. de (1924f). «Unha na palma». *Correio dos Açores*.
Vasconcelos, J. L. de (1925a). *A figa — Estudo de etnografia comparativa, precedido de algumas palavras a respeito do "sobrenatural" na medicina popular portuguesa (Conferência lida na Faculdade de Medicina do Porto no I Centenário da Régia Escola de Cirurgia do Porto)*. Araújo e Sobrinho.
Vasconcelos, J. L. de (1925b). História da língua portuguesa (Origem e vida externa). *Revista Lusitana*, *XXV*, 5–28.
Vasconcelos, J. L. de (1925c). *Medicina dos Lusitanos (Conferência proferida em 12 de dezembro de 1925, seguida dum apêndice)*. Lisboa.
Vasconcelos, J. L. de (1926a). *Mês de sonho: Conspecto de etnografia açórica*. Academia das Ciências de Lisboa.
Vasconcelos, J. L. de (1926b). Para a história da arqueologia portuguesa. *Cultura*, *I*(1), 25–26.
Vasconcelos, J. L. de (1927a). Alguns topónimos portugueses. *Revista Filológică*, *1*, 33–34.
Vasconcelos, J. L. de (1927b). *De terra em terra* (Vol. 1). Lisboa: Imprensa Nacional.
Vasconcelos, J. L. de (1927c). *De terra em terra* (Vol. 2). Lisboa: Imprensa Nacional.
Vasconcelos, J. L. de (1928a). *Opúsculos* (Vol. 1). Imprensa da Universidade de Coimbra.
Vasconcelos, J. L. de (1928b). *Opúsculos* (Vol. 2). Imprensa da Universidade de Coimbra.
Vasconcelos, J. L. de (1929a). Campo de Coimbra. *Ação Regional*.
Vasconcelos, J. L. de (1929b). *Opúsculos* (Vol. 4). Imprensa da Universidade de Coimbra.
Vasconcelos, J. L. de (1929c). Lenda popular: A truta de Celorico da Beira. *Correio*.
Vasconcelos, J. L. de (1929–1933). Post Mortem. In *O Soneto Neo-Latino* (Vol. 2), p. 63.
Vasconcelos, J. L. de (1930a). Celorico da Beira. *Correio*.
Vasconcelos, J. L. de (1930b). Considerações gerais e sumárias acerca das fontes de investigação etnográfica. *Portucale*, *III*, 3–6.
Vasconcelos, J. L. de (1930c). Divisões geográficas. *Biblos*, *VI*, 486–489.
Vasconcelos, J. L. de (1930d). Linguagem portuguesa de Alamédilla ou «Almedilha». In *Estudios eruditos in Memoriam de Adolfo Bonilla y San Martín* (Vol. 2, pp. 627–631. Madrid.
Vasconcelos, J. L. de (1930e). O Museu Etnográfico dos Açores (Projeto). Carta aberta ao Dr. José Bruno Carreiro, diretor do Correio dos Açores. *Correio dos Açores*.
Vasconcelos, J. L. de (1930f). Um dicionário português-latino. *Revista Lusitana*, *XXVIII*, 296–297.
Vasconcelos, J. L. de (1931a). Matéria filológica. *Revista Lusitana*, *XXIX*, 287–294.
Vasconcelos, J. L. de (1931b). *Opúsculos* (Vol. 3). Imprensa da Universidade de Coimbra.
Vasconcelos, J. L. de (1933a). Castros lusitânicos. *O Arqueólogo Português*, *XXIX*, 31–49.
Vasconcelos, J. L. de (1933b). Discurso inaugural, pronunciado em Coimbra em 21 de setembro de 1930, na 1.ª sessão (conjunta) dos Congressos de Antropologia e Arqueologia Pré-Histórica e do Instituto Internacional de Antropologia [História da Arqueologia em Portugal]. *O Arqueólogo Português*, *XXIX*, 76–82.
Vasconcelos, J. L. de (1933c). *Etnografia portuguesa: Tentame de sistematização* (Vol. 1). Imprensa Nacional.

Vasconcelos, J. L. de (1933d). *Memórias de Mondim da Beira (Para a história do concelho deste nome)*. Imprensa Nacional.
Vasconcelos, J. L. de (1934a). Etnografia Portuguesa (Tentame de sistematização) [Anúncio publicado pela Imprensa Nacional]. *Revista Lusitana*, *XXXII*, 318.
Vasconcelos, J. L. de (1934b). Inscrição ibérica do Algarve. *Revista de Arqueologia*, *II*, 43–44.
Vasconcelos, J. L. de (1935a). As origens da cidade do Porto, por Mendes Correia. *Revista Lusitana*, *XXXIII*, 315–318.
Vasconcelos, J. L. de (1935b). Prefácio. In M. de Vasconcelos, *A Vila de Canavezes — Notas para a sua história* (v–viii). Imprensa Nacional.
Vasconcelos, J. L. de (1936). *Etnografia portuguesa: Tentame de sistematização* (Vol. 2). Imprensa Nacional.
Vasconcelos, J. L. de (1937a). Fontes de Investigação Etnográfica. *Boletim Etnográfico*, *5*, 5–16.
Vasconcelos, J. L. de (1937b). Lume e iluminação. *Boletim Etnográfico*, *5*, 22–26.
Vasconcelos, J. L. de (1938a). Nomenclatura numismática. *O Arqueólogo Português*, *XXX*, 126–149.
Vasconcelos, J. L. de (1938b). *Opúsculos* (Vol. 5). Imprensa Nacional de Lisboa.
Vasconcelos, J. L. de (1938c). *Opúsculos* (Vol. 7). Imprensa Nacional de Lisboa.
Vasconcelos, J. L. de (1938d). *Ucanha e o seu pelourinho* [Folheto].
Vasconcelos, J. L. de (1939). Os Saloios (Na Estremadura cistagana). *Revista Lusitana*, *XXXVII*, 271–299.
Vasconcelos, J. L. de (1940a). Da fala de Barrancos. *Boletim de Filologia*, *6*, 159–177.
Vasconcelos, J. L. de (1940b). Lápide romana de São Miguel de Odrinhas, concelho de Sintra. *Portucale*, *XIII*, 3–5.
Vasconcelos, J. L. de (1940c). Papéis de Aquiles Estaço (Informação biobibliográfica). *Petrus Nonius*, *III*, 153–170.
Vasconcelos, J. L. de (1940d). Um cognome raro na epigrafia [Presalso]. *Petrus Nonius*, *III*, 229.
Vasconcelos, J. L. de (1940–1941). Condados da Idade-Média. *Revista da Faculdade de Letras* (Lisboa), *VII*, 313–318.
Vasconcelos, J. L. de (1941a). Curiosidades etnográfico-antropológicas. *Portucale*, *XIV*, 3–7.
Vasconcelos, J. L. de (1941b). *Etnografia portuguesa: Tentame de sistematização* (Vol. 3). Imprensa Nacional.
Vasconcelos, J. L. de (1941c). Origem do povo português (Estudo atual e sucinto do problema). *Revista Lusitana*, *XXXVIII*, 196–246.
Vasconcelos, J. L. de (1942). A província de Entre-Douro-e-Minho: Seu nome geral e nomes étnicos. *Gil Vicente*, *18*, 101–110.
Vasconcelos, J. L. de (1958). *Romanceiro português* (com uma notícia preliminar de R. Menéndez Pidal). Coimbra: Por Ordem da Universidade, em comemoração do centenário do nascimento de José Leite de Vasconcelos.
Vasconcelos, J. L. de (1985). *Opúsculos* (Vol. 6; M. A. Cintra, Org.). Imprensa Nacional-Casa da Moeda.

Isto dá muito trabalho – a interação metalinguística de estudantes de educação

This is a lot of work – metalinguistic interaction of teacher students

Ana Luísa Costa[1], *Xavier Fontich*[2]

[1] Instituto Politécnico de Setúbal e Centro de Linguística da Universidade de Lisboa
[2]Universidade Autónoma de Barcelona

Abstract

The study of the interaction between students of a teacher training course during the resolution of a Grammar Lab is guided by two goals: (i) to characterize metalinguistic interactions between schooled adults in the face of grammatical problems, in collaborative work and (ii) to consolidate a model of integrated analysis of the interaction, deepening the relationship between metalinguistic argumentation and metalinguistic knowledge. Metalinguistic interaction is understood as an extension of metalinguistic activity, and materialized in dialogical verbal discourse, in our case contextualized in the resolution of grammar problems. Framed by other studies of interaction in the sociocultural-Vygotskian framework (Mercer, 2019; Edwards & Mercer, 2013), the analysis follows the integrated model of Fontich (2010a), which explores the relationship between metalinguistic argumentation and verbalized metalinguistic knowledge. The results show three different patterns of interaction and point to a dissociation between the grammar knowledge expressed as a product, in the written answers, and the ability to explain that same knowledge.

Keywords: interaction, metalinguistic activity, grammar concepts, grammar lab, teacher education.

Resumo

O estudo da interação entre estudantes de um curso de formação de professores durante a resolução de um Laboratório Gramatical é guiado por dois objetivos: (i) caraterizar interações metalinguísticas de adultos escolarizados perante problemas gramaticais em trabalho colaborativo e (ii) consolidar um modelo de análise integrada da interação que explore a relação entre argumentação metalinguística e o conhecimento metalinguístico. A *interação metalinguística* é entendida enquanto vertente da atividade metalinguística (Camps, 2014), sendo materializada em discurso verbal dialógico, neste caso contextualizado na resolução de problemas de gramática. Enquadrada por outros estudos da interação no quadro sociocultural-vygotskiano (Mercer, 2019; Edwards & Mercer, 2013), a análise segue o modelo integrado de Fontich (2010a), que explora a relação entre argumentação metalinguística e conhecimento metalinguístico verbalizado. Os resultados permitem identificar três padrões de interação e apontam para uma dissociação entre a gramática que se sabe, evidenciada nos produtos-resposta, e a capacidade para fundamentar o que se sabe.

Palavras-chave: interação, atividade metalinguística, conceitos gramaticais, laboratório gramatical, formação de professores.

1. Breve introdução ao problema

Tem sido crescente o interesse pela *interação* no âmbito da Linguística Educacional, emergente tanto em estudos sobre a relação entre desenvolvimento da competência de escrita e aprendizagem da gramática, como

em estudos sobre a aprendizagem de *conceitos gramaticais* e da capacidade de *raciocinar metalinguisticamente* para resolver problemas de gramática. Ainda assim, considerando a longa rota de investigação no campo da interação em outras áreas da educação – por exemplo, na Educação Matemática, a resolução de problemas matemáticos, ou, na Didática das Ciências, a explicação de experiências laboratoriais – pode considerar-se que são escassos os trabalhos que se dedicam a gravar e analisar interações professor-alunos ou diálogos entre alunos em trabalhos de grupo nas aulas de língua e, em particular, no domínio da gramática. E, quanto ao ensino superior, especificamente sobre cursos de formação de professores, pode afirmar-se que muito raras são as pesquisas sobre o que se passa na interação em trabalhos de grupo de estudantes de educação sobre a gramática que vão ensinar e sobre como a vão ensinar. É este o nosso ponto de partida ou, dito de outra forma, o nosso *problema* de investigação. Pretendendo contribuir para a caracterização da *interação* entre estudantes do ensino superior, futuros professores, o objeto de estudo específico é a *interação metalinguística*, entendida enquanto vertente da atividade metalinguística e materializada em discurso verbal dialógico, contextualizado na resolução de problemas de gramática. As verbalizações associadas a este discurso são metalinguísticas, algumas envolvem termos e conceitos linguísticos e podem ser sequências de natureza argumentativa, fundamentando pontos de vista.

O presente estudo de interações verbais orais durante a realização de um Laboratório Gramatical (LG) em trabalho colaborativo por estudantes de um 3.º ano de uma licenciatura na área da educação orientou-se por dois objetivos:

(i) Caraterizar interações metalinguísticas de estudantes de educação perante problemas gramaticais durante a realização de um LG em trabalho colaborativo.

(ii) Refinar um modelo de análise integrada da interação que explore a relação entre argumentação metalinguística e conhecimento metalinguístico.

Com o primeiro objetivo, procuraremos responder a duas questões de investigação: (i) quais as estratégias discursivas de argumentação dominantes no raciocínio metalinguístico destes estudantes de educação? e (ii) o que caracteriza o conhecimento metalinguístico enunciado nas verbalizações de estudantes? Relativamente ao segundo objetivo, procuraremos responder à questão (iii) quais as vantagens de um modelo de análise de interações sobre gramática que integra simultaneamente o processo de reflexão metalinguística e o processo de emergência de conceitos?

As duas primeiras questões de investigação enquadram-se nos pressupostos teóricos dos estudos sobre interação com enquadramento sociocultural-vygotskiano, explicados na secção 2. Considerando os poucos estudos empíricos de interações em aulas de gramática, como se poderá constatar na secção 3, a maioria foca-se no discurso de alunos do ensino básico e secundário, sendo raros os estudos no campo da formação de professores. Além disso, a análise de dados de interações verbais orais tem recebido tratamentos metodológicos bastante díspares, que impedem uma maior comparabilidade entre os resultados. Assim, na sequência da terceira questão, exploraremos uma abordagem metodológica integradora e, para isso, na secção 4, explicamos o modelo de análise proposto, que suporta os resultados apresentados em 5. A secção 6 inclui algumas conclusões e considerações, apontando trabalho futuro.

2. Porquê estudar a interação metalinguística nas aulas de língua?

Numa perspetiva em que a dimensão social explica o desenvolvimento humano, individual e histórico, entende-se que a interação seja motor desse desenvolvimento. Por isso, estudar a interação significa estudar a aprendizagem, procurar entender como se desenvolve e que processos lhe são inerentes.

Este estudo, que foca a interação verbal sobre a própria linguagem verbal, enquadra-se numa teoria sociocultural da educação. Esta perspetiva postula que o desenvolvimento psicocognitivo humano é

inerentemente social e que se alimenta de interações em que os participantes co-participam da criação de um sistema partilhado de conhecimento, essencial à atividade humana (Hernández-Rojas, 2008; Smagorinsky, 2011). Miller (2011), contudo, esclarece que, no pensamento vygotskiano, a atividade social começa com a interação, mas culmina no diálogo interno, integrando o "outro social" nos processos cognitivos do indivíduo. Para Vygotsky, as funções psicológicas superiores emergem através da mediação de sistemas semióticos culturalmente constituídos, nos quais a linguagem verbal é a principal ferramenta de desenvolvimento, juntamente com outras representações simbólicas (Smagorinsky, 2023). A aprendizagem ocorre quando os aprendentes passam da dependência de um "outro mais capaz" para a competência autónoma (Miller, 2011). Cazden (1981) descreve este processo como "desempenho antes da competência". Nesta linha de ideias, os alunos devem envolver-se em atividades estruturadas que conduzam gradualmente ao domínio da competência. Esta transição do intermental para o intramental promove capacidades mais profundas de compreensão do conhecimento e a autonomia no raciocínio (van der Veer, 2020). Assim, na educação, o quadro sociocultural-vygotskiano desloca-se de ambientes de aprendizagem tradicionais para ambientes participativos e colaborativos.

Para entender os princípios deste quadro teórico, importa ter igualmente presente que a visão sociocultural se alinha com o dialogismo bakhtiniano, que perspetiva a linguagem como inerentemente reactiva e dialógica (Riestra, 2010). Bakhtin afirma que o discurso é moldado por outras vozes, além da do próprio, formatando processos cognitivos isolados (Barnes & Todd, 2001; Wegerif & Major, 2023). Esta perspetiva apela a que se repense o papel da escola, superando a mera instrução direta e promovendo ambientes de aprendizagem interactivos e enriquecedores, nos quais os alunos tenham oportunidade de construir conhecimento de forma colaborativa (Mercer, 2013). As escolas devem proporcionar espaços para a pesquisa, o diálogo e a exploração, permitindo que os alunos interajam de forma significativa com ferramentas culturais (Wells & Mejia, 2005; Smagorinsky, 2023).

A aprendizagem vygotskiana, contudo, não consiste simplesmente em interagir com um parceiro conhecedor, mas em atividades estruturadas e mediadas pelo professor, que mergulham os alunos em experiências de aprendizagem significativas (van der Veer, 2020). O ensino é visto como um processo mediado e orientado para a autorregulação do aluno. Inicialmente regulada por outros (Miller, 2011), a aprendizagem transita para a zona de desenvolvimento proximal, na qual os alunos progridem por meio de apoio orientador intencional ou, por outras palavras, progridem "andaimados".

Ao contrário dos contextos informais de aprendizagem, como as práticas familiares ou profissionais, a aprendizagem escolar distingue-se pela estruturação explícita de cenários de resolução de problemas e pela sua orientação estratégica (Taber et al., 2016). O desafio do professor reside na gestão de diversos alunos em diferentes fases de desenvolvimento, o que requer interações dinâmicas na sala de aula que promovam a metacognição e espaços semióticos partilhados (Mercer, 2019). Uma vez mais, saliente-se que o envolvimento sustentado em atividades colaborativas aumenta a capacidade de reflexão autónoma e individual dos alunos, uma vez que o "outro social" interiorizado continua a moldar os seus processos de pensamento (Miller, 2011).

Alguns conceitos-chave neovygotskianos, como os *andaimes* e a *fala exploratória*, sustentam a conceptualização da aprendizagem como um processo co-construído (Lave & Wenger, 1991). Esta visão da aprendizagem, no quadro teórico socioconstrutivista, integra diversas perspetivas, como o construtivismo piagetiano e a teoria sociocultural-vygotskiana, as quais valorizam uma agência ativa na aprendizagem (Cubero, 2005). A mediação semiótica, particularmente através da linguagem verbal, é facilitadora do desenvolvimento cognitivo, reforçando a ideia de que os processos psicológicos humanos são inerentemente sociais e enraizados em atividades culturais (Wertsch, 1985).

A metáfora do *andaime* descreve o apoio prestado por um professor para ajudar um aluno a adquirir conhecimentos. Inicialmente aplicada ao ensino de línguas, foi alargada a processos de ensino e aprendizagem em diversas disciplinas. Os *andaimes* implicam uma redução da liberdade do aluno, pela necessidade de orientação da sua atenção e pelo fornecimento de modelos de compreensão do conhecimento. Destacam-se três tipos de andaimes: os ecológicos (no contexto natural de aprendizagem), os sociais (como as interações com o

professor ou entre pares) e os individuais (como os apoios ao desenvolvimento cognitivo). Alguns centram-se no ambiente da sala de aula, enquanto outros se focam nas estratégias de aprendizagem colaborativa, na internalização dos saberes ou nos mecanismos discursivos. A investigação educacional tem reconhecido os *andaimes* (o *scaffolding*) como um processo dialético em que os alunos desenvolvem autonomia superando a instrução direta do professor. Por exemplo, o discurso exploratório, no qual os alunos interagem de forma crítica com as ideias dos seus pares, é um mecanismo crucial para o desenvolvimento cognitivo. A aprendizagem é reforçada quando os alunos abordam abertamente os erros e as suas conceções erróneas, em vez de se limitarem a procurar as respostas corretas esperadas pelo professor.

O conceito de *andaime*, portanto, também se aplica às interações entre pares, exigindo uma reconsideração da conceção original de *zona de desenvolvimento proximal* (ZDP) na aprendizagem colaborativa. Do ponto de vista do ensino, torna-se necessário que o professor seja flexível e incentive interações que conduzam a uma transformação conceptual, para desenvolver aprendizagens efetivas. Sublinhe-se que uma negociação significativa exige esforço e tempo, não se podendo restringir à transmissão de respostas corretas pré-existentes. Implica orientar os alunos na construção da aprendizagem através do questionamento, da pesquisa e da reflexão. Em suma, *andaimar* é um processo colaborativo e responsivo que promove a autonomia na aprendizagem.

Na investigação sobre interação, nas dinâmicas de sala de aula, Barnes e Todd (2001) foram pioneiros nas abordagens socioculturais, destacando o discurso como ferramenta para a construção do conhecimento. O seu trabalho mostrou que os conceitos emergem da interação, não sendo entidades fixas e pré-existentes. Paralelamente, também desde o final do século XX, os estudos de análise do discurso foram adotando uma perspetiva social de aprendizagem, considerando a educação como um processo de comunicação que constrói contextos mentais compartilhados (Cubero et al., 2008; Mercer, 2019). A análise do discurso passou do estudo da própria linguagem para o estudo do seu papel na construção da aprendizagem. Esta perspetiva assenta em três pressupostos: (i) o conhecimento é inerentemente social e cultural, e só existe quando é partilhado; (ii) a análise do discurso deve focar-se mais nos conteúdos e nas funções da linguagem do que nas estruturas linguísticas e (iii) o conhecimento depende do contexto e é moldado por atividades conjuntas e normas de comunicação (Wells et al., 2024).

O contexto, enquanto espaço intersubjetivo, desempenha um papel crucial na aprendizagem (Edwards & Mercer, 2013). Por essa razão, vários autores têm desenvolvido recentemente investigação sobre interação em contexto escolar associada à aprendizagem colaborativa (Mercer, 2019). Nesta linha, destaca-se o trabalho de Van Lier (2004), que propõe uma perspetiva semiótica ecológica, reconhecendo múltiplas fontes de andaimes na aprendizagem cooperativa e analisando o discurso interacional em diferentes contextos educativos.

A investigação sobre interação em sala de aula revela, contudo, um paradoxo. Reconhece-se que a interação é crucial para a aprendizagem, no entanto, a qualidade da interação é muitas vezes negligenciada, como se se esperasse que os alunos infiram por si, sem apoio, as estratégias e regras para uma interação com sucesso (Edwards & Mercer, 2013). A consciência desta lacuna aponta para a necessidade de instrução explícita que apoie a emergência da fala exploratória, sendo ainda urgente ultrapassar contextos de aprendizagem em que o diálogo é dominado pelo professor e os alunos procuram adivinhar as respostas corretas, em vez de se envolver em discussões significativas (Mercer, 2019). A instrução explícita para a co-construção de discurso exploratório ajuda os alunos a envolverem-se em discussões de qualidade e a desenvolverem capacidades de raciocínio autónomas. Por outras palavras, os trabalhos de grupo e a interação não garantem só por si a aprendizagem, podendo o confronto de perspetivas conduzir a disputas pessoais, em vez de trocas construtivas (Schwarz et al., 2009). Cabe ao professor promover uma reflexão intencional sobre uma comunicação eficaz, baseada em discurso exploratório. Também com este propósito, o papel do professor muda de instrutor para facilitador, orientando uma comunidade de aprendizagem colaborativa (Garth-James & Hollis, 2014). Assim, uma sala de aula dialógica baseia-se na negociação de regras de discussão e reflexão, criando um espaço intersubjetivo onde as ideias competem, mas a coesão do grupo é mantida (Vandenberg et al., 2021).

3. Como estudar a interação metalinguística?

Na secção anterior, sublinhámos o interesse pela interação em situações de aprendizagem colaborativa num quadro em que a atividade verbal dialógica é o cerne da aprendizagem em qualquer área do saber. Muitos dos estudos mencionados na secção anterior investigam em contextos de aulas de diferentes disciplinas. Nesta secção, focamos estudos no âmbito da didática da L1. Nem todos os autores perspetivam a interação como motor de aprendizagem: alguns veem a interação como mais uma estratégia de ensino-aprendizagem (reconhecendo as potencialidades gerais da aprendizagem colaborativa), outros como janela de acesso a processos psicolinguísticos de aprendizagem ou, no caso das línguas, de reflexão metalinguística. Observa-se ainda um desconhecimento mútuo entre autores de origens linguísticas diferentes, como se depreende, a título de exemplo, da ideia de que "[o]ne aspect is that most research conducted in this area has almost exclusively been done in Anglophone context" (Strandberg & Van Rijt, 2024, s/p). Esta dispersão de visões e de objetivos de investigação conduz a uma considerável dispersão de metodologias de recolha, tratamento e análise de dados, o que limita bastante a comparabilidade de resultados.

Para ilustrar este estado da arte, apresentaremos exemplos de alguns estudos representativos de diferentes origens linguísticas (associados ao ensino do português, do catalão, do espanhol, do francês ou do inglês como L1), os quais, tal como o nosso, têm como foco a interação metalinguística. Organizamos esta sistematização de abordagens e categorias de análise disponíveis na literatura em dois eixos que permitem caracterizar diferentes respostas à pergunta "como estudar a interação metalinguística?". O primeiro eixo inclui estudos cujos dados são (i) interações durante o processo de escrita, em situações de escrita colaborativa, a pares ou em grupo. No segundo eixo, os dados são (ii) interações durante o processo colaborativo de resolução de problemas de gramática por pares ou por grupos.

Começando pelo primeiro eixo, são pioneiros trabalhos enquadrados por protocolos típicos da investigação em psicologia da aprendizagem, como os de Camps (1991, 1994) e de Barbeiro (1994, 1999), nos quais se analisam interações verbais durante o processo de escrita.

Em Camps (1991, 1994), apresenta-se um estudo descritivo/interpretativo que analisa interações orais de um grupo de 3 alunos (2 raparigas e um rapaz) a frequentar o 7.º ano de escolaridade (com cerca de 12 anos), falantes do catalão. Os alunos receberam instruções explícitas sobre géneros textuais argumentativos e desenvolveram uma tarefa de escrita colaborativa de um texto argumentativo sobre o racismo. Os enunciados e episódios foram codificados segundo categorias como o tipo de operação do processo de escrita (controlo da tarefa, elaboração, planificação, textualização e revisão) e verbalizações que evidenciam conhecimento metalinguístico, metatextual e metadiscursivo, com recurso a termos e conceitos metalinguísticos ou com expressões de fala quotidiana. No estudo qualitativo e quantitativo, com validação estatística, desenvolvido em Barbeiro (1994, 1999), analisam-se as interações de grupos de 3 alunos, oito por cada ano de escolaridade (2.º, 4.º, 6.º e 8.º anos), num total de 96 participantes com idades entre os 7;10 e os 14;00. Tal como em Camps (1991, 1994), analisam-se "atividades do processo de escrita", segundo os núcleos *planificação, redação e revisão* (Barbeiro, 1999, p. 119) e são catergorias de análise operações de construção textual, de conceção textual e de realização textual. No discurso dos alunos que acompanha as diferentes operações, são também objeto de análise argumentos de fundamentação linguística, que correspondem a manifestações verbais que explicam, questionam, contestam propostas ou que aceitam ou recusam essas propostas. Em síntese, este trabalho integra a análise de sequências argumentativas que fundamentam o raciocínio metalinguístico.

Outros estudos em sala de aula, associados ao ensino de línguas românicas e anglossaxónicas, exemplificam este eixo de análise de interação metalinguística durante o processo de escrita. Tendo por base metodológica para recolha de dados o sistema Ramos (Calil, 2020), Eduardo Calil tem interagido com diversos colegas e desenvolvido análises que se focam nas interações durante o processo de escrita por crianças portuguesas, brasileiras e francesas no 1.º ciclo/ensino primário. O sistema Ramos permite alinhar não só a gravação das falas, mas também os movimentos da caneta e a imagem do manuscrito, o que favorece a perspetiva de análise genética do manuscrito adotada pelo autor. Nesta linha de estudos, o foco são as interações que geram comentários que antecedem, que acompanham ou que são posteriores a decisões marcadas no

manuscrito. Entre estes comentários, emergem verbalizações ou ações que evidenciam conhecimento metalinguístico em construção. Por exemplo, em Barbeiro et al. (2022), a análise assume como categorias termos metalinguísticos que surgem na interação, nos domínios "texto, género e discurso", "gramática", "ortografia", "pontuação" e "grafia", e ainda operações e dimensões linguístico-discursivas associadas a opções de construção do texto. Outro exemplo é o artigo de Calil e Myhill (2020), no qual os autores analisam comentários e intervenções textuais, como rasuras, de uma díade durante as operações de revisão textual. Nas interações das duas alunas, são predominantes verbalizações sobre a dimensão gráfico-espacial, mesmo quando o reconhecimento de problemas no manuscrito é de natureza lexical ou sintática. Várias análises de interação metalinguística ou, usando o conceito prevalente, de *metatalk* relacionam desenvolvimento da escrita e *metalinguistic understanding*, fundamentalmente em contextos de aprendizagem do inglês L1 (Myhill et al., 2013; Myhill & Newman, 2019: Myhill, 2021, e.o.).

Um conjunto representativo de estudos que analisam interação numa perspetiva de atividade metalinguística encontra-se associado ao trabalho do grupo catalão GREAL (Camps & Fontich, 2020; Fontich & García-Folgado, 2019; Fontich et al., 2022). Entre estes estudos, tal como nos casos anteriormente apresentados, vários analisam interações em tarefas de escrita colaborativa (Camps et al., 1997; Ribas et al., 2020, e.o.). Neste quadro, surgem como categorias dicotomias como conhecimento processual / conhecimento declarativo ou verbalizações com recurso a metalinguagem / verbalizações com linguagem do dia a dia (Camps, 2014), presentes nas interações de alunos falantes do catalão e do espanhol de diferentes idades. No mesmo alinhamento teórico da atividade metalinguística, salientam-se os estudos de interação de Fontich (2010a, 2010b, 2021, 2014), cujo modelo de análise seguiremos e que ilustram o segundo eixo de análise, sobre interações durante o processo de resolução de problemas de gramática em trabalho colaborativo.

No que concerne este segundo eixo, ainda mais escassa é a investigação sobre interação especificamente em torno de tarefas de gramática. Alguns estudos sobre a construção de conceções gramaticais e de argumentação metalinguística analisam argumentos metalinguísticos expressos em respostas escritas a questões sobre problemas gramaticais. Disso é exemplo o artigo de Jouili e Elalouf (2021), no qual se estuda a conceptualização da noção da função sintática *sujeito* entre grupos de alunos de ensino secundário com tradições gramaticais diferentes (a francesa e a espanhola). Igualmente, em Van Rijt et al. (2021), um artigo sobre capacidades de reflexão metalinguística, os dados são respostas a testes realizados por escrito, nos quais estudantes do ensino secundário e estudantes do ensino superior de cursos de formação de professores explicam os seus raciocínios ao resolver problemas que convocam conhecimento sintático explícito. Ainda que o objetivo destes estudos seja descrever o raciocínio metalinguístico e o conhecimento associado, os dados não são interações orais, como acontece nos estudos que referimos no primeiro eixo e nos que destacaremos em seguida.

Numa abordagem naturalista de recolha de dados de aula, em Gauvin e Boivin (2013), analisam-se cerca de 9 horas de gravação de interações de um total de 54 alunos com idades entre 12 e 13 anos, a trabalhar em pares, resolvendo exercícios de que exploram a noção de *verbo*. Uma vez que o objetivo é descrever a natureza dos critérios que a que os alunos recorrem para identificar *verbos*, são categorias de análise a oposição conhecimento declarativo / conhecimento processual e tipos de conhecimento face aos conhecimentos instituidos (decisivos, não decisivos, marginais). As autoras concluem que o recurso a conhecimentos marginais diminui após a institucionalização dos saberes, mas há conhecimentos prévios que são resistentes à institucionalização de novos conhecimentos, o que é uma pista importante para a transposição didática de conhecimentos gramaticais.

Num outro contexto linguístico, numa sala de aula sueca, o estudo igualmente naturalista de Strandberg e Van Rijt (2024) analisa 2h51' de interações de estudantes de ensino secundário, com idades entre os 15 e os 17, organizados em 3 grupos de 4. A tarefa induz uma reflexão em torno de propriedades linguísticas a partir de texto literário, numa abordagem didática que os autores consideram "contextualized grammar teaching" (Strandberg & Van Rijt, 2024, p. 3). Com o objetivo de analisar quer os termos e conceitos de emergem no discurso dos alunos, quer a qualidade da capacidade de reflexão metalinguística, os dados são analisados qualitativamente e quantitativamente, com validação estatística. Como categorias de análise para o estudo da

argumentação metalinguística, os autores recorrem aos tipos de fala (disputativa, cumulativa e exploratória) propostos, por exemplo, em Mercer (2019) e distinguem o recurso a conceções erróneas e "rules of thumb" da mobilização intencional de manipulações linguísticas correspondentes a testes sintáticos validados pela teoria linguística. Em síntese, a qualidade do raciocínio metalinguístico é considerada dependente do recurso adequado a testes linguísticos. Numa linha de análise muito semelhante, importa ainda referir a comunicação de Cardoso et al. (2019) à ARLE 2019. Neste trabalho, analisam-se interações de 20 estudantes de uma licenciatura em Educação Básica, organizados em pares, durante a resolução de tarefas que envolvem conhecimento sintático. Tal como no estudo de Strandberg e Van Rijt (2024), as categorias de análise são os tipos de fala de Mercer (2019), sendo unidades de análise episódios de fala exploratória. Nestes episódios, analisam-se verbalizações com e sem recurso a metalinguagem sobre classes de palavras. As autoras enquadram a análise destas verbalizações na perspetiva da atividade metalinguística, tendo por referência as categorias antes mencionadas de Camps (2014).

Neste quadro teórico de Atividade Metalinguística, a metodologia de análise originalmente proposta em Fontich (2010a) adota uma conceptualização coerente com o enquadramento sociocultural a que nos referimos na secção 1. Em Fontich (2010), analisam-se interações orais de um total de 24 estudantes do início do ensino secundário (com cerca de 12 anos), organizados em três grupos com perfis de desempenho diferentes (elevado, médio e baixo). A tarefa consistiu na realização, em pequenos grupos, de uma Sequência Didática (SD) sobre complementos verbais e o conceito de transitividade. O modelo proposto tem a vantagem de apresentar, de forma integrada, evidenciando as inter-relações, uma análise do discurso argumentativo metalinguístico, que corporiza o raciocínio que fundamenta pontos de vista e opções dos alunos, e do conhecimento metalinguístico verbalizado.

Neste modelo, distinguem-se cinco unidades de análise do discurso dialógico. Uma primeira unidade é o *diálogo*, que pode ser entendido como a resposta para uma questão de aula ou tarefa. Os limites do diálogo são variáveis e podem ir desde o início de uma discussão até ao momento em que o grupo decide parar. Um diálogo divide-se em *sequências discursivas*, organizadas em torno de um tópico específico, associado a uma tarefa (pode envolver o convite aos colegas a seguir a instrução dada, por exemplo a identificação de complementos verbais, a incitação à participação na discussão, a expressão de dúvidas, a tomada de turnos sobre tópicos fora da tarefa, etc.). Entre as diversas sequências discursivas, algumas são de tipo metalinguístico, ou seja, focam especificamente tópicos associados aos conteúdos gramaticais. Estas *sequências metalinguísticas* também têm uma extensão variada, podendo ser bastante longas durante a realização de tarefas de gramática. Por essa razão, para a análise, podem ser divididas em subsequências. Uma *subsequência metalinguística* explora um mesmo tópico ou uma mesma noção, de duas ou mais perspetivas (por exemplo, a função acusativa de um complemento verbal pode ser referida em relação à forma do pronome por que pode ser substituído e em relação a propriedades de subcategorização verbal).

Relativamente ao fluxo do raciocínio metalinguístico, as sequências e subsequências metalinguísticas integram *sequências argumentativas*, as quais incluem *episódios argumentativos aditivos* (como *explicar*, *expandir*, *adicionar*, *aceitar*, *concluir*) ou *reativos* (*divergir*, *contradizer*, *esclarecer*, *desafiar*), em conformidade com as propostas de Bee (2000) e Kumpulainen e Wray (2002). Assim, para a análise desta dimensão de discurso argumentativo, os quadros aditivos e reativos são categorias que permitem manter uma visão integradora das relações entre as ideias, explicando a organização argumentativa do fluxo discursivo. Ao contrário de outras categorias de análise que isolam ideias e argumentos, este modelo de análise mantém a perspetiva geral do fluxo discursivo em contexto. Quanto aos conceitos metalinguísticos que emergem no discurso, especificamente nas sequências argumentativas, a unidade básica deste modelo de análise é o *enunciado metalinguístico*, que corresponde a uma expressão metalinguística proferida de uma só vez ou em mais de um turno de fala. As categorias de análise dos enunciados metalinguísticos atendem à classificação da natureza do conhecimento que as sustenta, que pode ser de base pragmática, semântica ou formal (por exemplo, sintática). Estas categorias enquadram-se nos trabalhos propostos em Larsen-Freeman (2003) e Van Lier (2004).

Na secção seguinte, a explicação da abordagem metodológica permitirá ilustrar estes instrumentos que apoiam o modelo de análise integradora de Fontich (2010).

4. Abordagem metodológica

Neste estudo, adotamos uma abordagem naturalista de recolha de dados e, em consonância com alguns dos estudos congéneres resenhados nos pontos 2 e 3, os dados serão analisados numa perspetiva qualitativa/interpretativa. Entendemos que o modelo a que recorremos, proposto originalmente em Fontich (2010a), partilha de características de um modelo de análise de conteúdo, com unidades (os diálogos) e categorias previamente estabelecidas, fundamentadas quer em outros estudos sobre interação em contextos educativos não exclusivamente de ensino da língua, quer em estudos sobre atividade metalinguística.

A recolha de dados foi feita em duas turmas de um 3.º ano de uma licenciatura em Educação Básica, numa disciplina da área da didática. Antes de iniciarem o módulo sobre aprendizagem da gramática, como estratégia de introdução ao Laboratório Gramatical numa perspetiva de isomorfismo pedagógico, os estudantes realizaram em grupos o LG "Que tipos de subordinadas existem?" (Costa & Santos, 2007). Foram previamente informados da dupla finalidade formativa e de investigação da tarefa e assinaram um termo de consentimento informado. Os grupos realizaram o LG como trabalho autónomo para a aula, gravando as suas próprias interações, em áudio ou vídeo, conforme se sentissem mais à vontade.

Para o presente artigo, focamo-nos na análise das interações de três grupos, com um total de 11 participantes. Os seus nomes foram anonimizados, sendo substituídos por nomes de código. A seleção destes grupos prendeu-se com a qualidade do som recolhido (e, em dois casos, com a disponibilização do vídeo). A Tabela 1 sistematiza a informação sobre participantes e dados.

Tabela 1

Caracterização dos Dados

GRUPOS	N.º de PARTICIPANTES	N.º de PALAVRAS	DURAÇÃO	GRAVAÇÃO
1	3	7199	74'32ss	áudio
2	4	5587	52'28ss	vídeo
3	4	4766	38'16ss	vídeo
Total	**11**	**17552**	**2h73'16ss**	

De entre as sete etapas do LG - Pergunta/Problema, Observar, Formular hipóteses, Testar hipóteses, Validar hipóteses, Exercitar e Avaliar (Duarte, 2008), foram selecionados episódios da segunda e da quinta etapas.

A análise das sequências argumentativas teve como categorias de análise dos enunciados argumentativos os movimentos reativos e aditivos usados em Fontich (2010a, 2010b, 2023), fundamentados nos trabalhos de Bee (2000) e Kumpulainen e Wray (2002). A Tabela 2 sistematiza estas categorias.

Tabela 2

Tipos de Movimentos nos Episódios Argumentativos

EPISÓDIOS ARGUMENTATIVOS	
REATIVOS	**ADITIVOS**
Divergir (para mudar o tópico)	**Explicar** (para introduzir um argumento)
Contradizer (para mostrar total desacordo)	**Expandir** (para ampliar a perspetiva sobre o que está a ser observado)
Esclarecer (para mostrar desacordo parcial)	**Adicionar** (para repetir ou reiterar)
Desafiar (para mostrar discordância, destacando uma contradição em relação ao que foi dito).	**Aceitar** (aprovar uma ideia)
	Concluir (para fechar uma argumentação)

As categorias que distinguem a natureza do conhecimento metalinguístico verbalizado foram definidas em função das propostas de Larsen-Freeman (2003) e Van Lier (2004) e são explicadas a partir de Fontich (2010a, 2023), com exemplos deste estudo. Uma vez que os termos adotados não são usados nas suas aceções mais comuns nos estudos linguísticos, recorre-se a exemplos ilustrativos dos critérios seguidos

Tabela 3

Natureza do Conhecimento nos Enunciados Metalinguísticos

CONHECIMENTO METALINGUÍSTICO	Explicação	Exemplos
PRAGMÁTICO	Argumento fundamentado em conhecimento pragmático, ou seja, em como se usa (como se diz, como se escreve).	"Porque tens sempre de ter o início da… Ai, da oração, que no caso é *Quando*." (por associação à ideia de que no início da oração se usam palavras como "quando"). (Tabela 4)
SEMÂNTICO	Argumento fundamentado em conhecimento de natureza semântica, por exemplos, em valores semânticos.	"É, geralmente é, o que faz a ligação da frase". (Tabela 4)
		"porque ela está a restringir." (Tabela 7)
		"Não está a completar? 'Tá!" (Tabela 10)
FORMAL	Argumento fundamentado em conhecimento metalinguístico explícito, que, no caso em estudo, é principalmente relativo a categorias sintáticas ou processos sintáticos (como testes sintáticos). O conhecimento formal expressa-se com recurso a termos e conceitos metalinguísticos.	"essa é complemento de verbo" (Tabela 8)
		"Dá para trocar" (Tabela 11) - teste mobilidade de um adjunto.

A categorias apresentadas nas Tabelas 2 e 3 constituem a estrutura das tabelas que são o instrumento, traduzido para o português a partir de Fontich (2010a, 2010b), que apoia a análise apresentada na secção 5.

5. Análise da interação metalinguística e discussão dos resultados

Apresentamos os resultados da análise organizados em duas secções correspondentes às duas etapas do LG, Observar (5.1.) e Validar uma hipótese (5.2.).

5.1. Episódios da etapa 2 do LG – Observar

Na etapa 2 do LG, que consiste em observar dados para identificar regularidades, a instrução, transcrita em (1), tinha como verbo de comando *sublinhar*, correspondendo a uma operação de identificação de uma oração subordinada, capacidade que os estudantes precisavam de mobilizar como conhecimento prévio.

(1) Considera as frases seguintes e sublinha a oração subordinada que cada uma delas contém.

Na Tabela 4, analisa-se um momento de interação do grupo 1 durante esta tarefa.

Tabela 4

Subsequência Metalinguística: Sublinha a Oração Subordinada em "Quando te vi no cinema, não te conheci", "Apesar de estar mau tempo, foi passear de barco" e "Já que fazes bolos bons, encomendo-te um bolo de chocolate".

SA	EA		EM			Diálogo A (Grupo 1)
	Reativo	Aditivo	Natureza			
			P	S	F	
I	Divergir	-	-	-	-	T1 Marta E1 – *Quando te encontrei no cinema…*
	-	Expandir	-	-	-	T1 Marta E2 – *Não te conheci*, não é?
	Contradizer	-	-	-	-	T2 Micá E3 – Não, não, não…
	-	Explicar	X	-	-	T2 Micá E4 – Porque tens sempre de ter o início da… Ai, da oração, que no caso é *Quando*.
II	-	-	-	-	-	T2 Micá E5 – Agora é *Apesar de estar mau tempo*.
	-	Explicar	X	-	-	T2 Micá E6 – Porque *Apesar* também faz parte do início da oração.
	Esclarecer	-	-	-	-	T3 Marta E7 – Mas, por exemplo, o *não* também fazia parte?
	Contradizer	-	-	-	-	T4 Micá E8 – Não.
	-	Explicar	-	X	-	T4 Micá E9 – É, geralmente é, o que faz a ligação da frase. Por exemplo, tens o *e*.
	-	Explicar	X	-	-	T5 Miriam E10 – É sempre o *que, quando, apesar, embora, cujo*.
	-	Explicar	-	X	-	T6 Micá E11 – Geralmente, é o que te liga às frases.
	-	Expandir	-	X	-	T6 Micá E12 – Por exemplo, o *não* geralmente não tem uma ligação tão forte.
	-	Explicar	-	X	-	T6 Micá E13 – Quando tu metes *quando*, o *quando* já tem tanto como o *e* ou o *que*… Fazem ligações de frases.
	-	Aceitar	-	-	-	T7 Marta E14 – OK, já percebi.
III	-	-	-	-	-	T7 Marta E15 – *Já que fazes bolos bons, encomendo-te um bolo de chocolate.*
	-	Expandir	-	-	-	T8 Micá E16 – O que é que achas que é? *[silêncio]*
	-	Explicar	-	X	-	T9 Micá E17 – É o que faz a ligação. Pensa assim…
	Esclarecer	-	-	X	-	T9 Micá E18 – Não é bem a ligação nestas… Não há bem uma forma…
	-	Adicionar	X	-	-	T10 Marta E19 – É o *já que*?
	-	Concluir	-	-	X	T11 Micá E20 – Exatamente. Prontos. É quando é uma palavra… Não é nenhum verbo, não é nenhum nome.

Nota. SA: sequência argumentativa; EA: episódio argumentativo; EM: enunciado metalinguístico; P: pragmática; S: semântica; F: formal.

No diálogo A do grupo 1, relativamente ao episódio argumentativo, ocorrem mais movimentos aditivos, do que reativos. Ainda assim, trata-se de um enunciado com movimentos diversificados, aparentemente exploratório. Marta evidencia dificuldades em identificar a subordinada e em mobilizar o conceito de conjunção como pista para a identificação de subordinadas nos turnos iniciais (T1, E2 e T3). Com a colaboração de Micá e Miriam, numa aproximação à sua zona de desenvolvimento proximal, no final da sequência, parece ter ativado esta pista para a identificação das subordinadas ("É o *já que*?", T10, E19).

Focando agora a natureza do conhecimento metalinguístico mobilizado pelo grupo, verifica-se que se trata de um conhecimento baseado em conceções pouco formais, dominantemente de natureza semântica e pragmática, apoiado no que alguns autores designam como "rules of thumb", que persistem no raciocínio metalinguístico de futuros professores (Van Rijt et al., 2021), numa fase em que se espera um conhecimento mais científico. Disso é exemplo o recurso ao conceito de "palavra de ligação", sem que nunca tenha ocorrido o termo "conjunção". Por outro lado, a marginalidade do conhecimento formal mobilizado, apesar do carácter exploratório da interação, parece ecoar na conclusão de que "exploratory talk is not necessarily connected to the quality of metalinguistic activity", no estudo com estudantes com o mesmo perfil de Cardoso et al. (2019, s/p).

Considerando, contudo, as respostas finais, ou seja, o produto da tarefa, o grupo registou uma identificação correta das três subordinadas em questão.

As Tabelas 5 e 6 apresentam duas subsequências metalinguísticas de um mesmo diálogo, durante a realização da tarefa pelo grupo 2.

Tabela 5

Subsequência Metalinguística: Sublinha a Oração Subordinada em "A - A rapariga que eu vi no teatro não era a tua amiga.", "B - Ele pediu-te que não fumasses tanto.", "C - Quando te encontrei no cinema, não te conheci."

SA	EA		EM			Diálogo A (Grupo 2)
	Reativo	Aditivo	Natureza			
			P	S	F	
I	Divergir	-	-	-	-	T1 Manuela E1 - A primeira...
	-	-	-	-	-	T2 Margarida E2 – Quem é que lê, então? Alguém lê?
	-	-	-	-	-	T3 Manuela E3 - Sim, pode ser.
	-	Aceitar	-	-	-	T4 Margarida E4 – Ok.
	-	-	-	-	-	T4 Margarida E5 - Considera as frases seguintes e sublinha a oração subordinada que cada uma delas contém: *A rapariga que eu vi no teatro não era a tua amiga*.
	-	Adicionar	-	-	-	T5 Manuela E6 - *Que eu vi no teatro*?
	-	Aceitar	-	-	-	T6 Margarida E7 - Sim
	-	Repetir	-	-	-	T7 Mafalda E8 – hum hum.
II	-	-	-	-	-	T8 Margarida E9 - *Ele pediu-te que não fumasses tanto*.
	-	Adicionar	-	-	-	T8 Margarida E10 - *Que não fumasses tanto*.
	-	Repetir	-	-	-	T9 Mafalda E9 - *Que não fumasses tanto*.
III	-	-	-	-	-	T10 Margarida E10 - *Quando te encontrei no cinema, não te conheci.*
	-	Adicionar	-	-	-	T11 Mariana E11 - *Quando te encontrei no cinema.*
	-	Aceitar	-	-	-	T12 Manuela E12 – hum hum.
	-	Repetir	-	-	-	T13 Margarida E13 – hum hum.

Nota. SA: sequência argumentativa; EA: episódio argumentativo; EM: enunciado metalinguístico; P: pragmática; S: semântica; F: formal.

O enunciado integra três subsequências com aparência de discurso cumulativo (Mercer, 2019). Contudo, a ideia de acumulação acrítica de perspetivas é contrariada pelo facto de as propostas de resposta se integrarem no discurso sujeitas a escrutínio do grupo e validação. Esta validação ativa é manifestada por movimentos quase totalmente aditivos, de aceitação, adição/acrescentamento de ideias e repetição. Ainda assim, considera-se que se trata de um discurso dominantemente cumulativo, em que não se evidencia construção de conhecimento, havendo mobilização de conhecimento internalizado e estabilizado. Face à inexistência de problemas a resolver, não há verbalizações que permitam inferir a natureza do conhecimento metalinguístico que fundamenta as opções. Relativamente ao produto das respostas, todas são corretas.

A Tabela 6 ilustra a emergência de um problema que gera uma rutura neste fluxo discursivo cumulativo.

Tabela 6

Subsequência Metalinguística: Sublinha a Oração Subordinada em "A Eva disse que queria uma boneca no seu aniversário."

SA	EA		EM			Continuação do diálogo A (Grupo 2)
	Reativo	Aditivo	Natureza			
			P	S	F	
I	Divergir	-	-	-	-	T1 – Manuela E1 - *A Eva disse que queria uma boneca no seu aniversário*. [20ss de silêncio]
	-	-	-	-	-	T1 Manuela E2 – [□□□]
	-	-	-	-	-	T2 Margarida E3 – Ah! [□.□w]
	-	Expandir	X	-	-	T3 Manuela E4 – Deve ser a partir do *que*, não?
	-	Aceitar	-	-	-	T4 Mafalda E5 – *Que queria uma boneca.*
	-	Aceitar	-	-	-	T4 Mafalda E6 – hum hum.
	Esclarecer	-	-	-	-	T5 Margarida E7 – A parte toda?
	Esclarecer	-	-	-	-	T6 Madalena E8 – Mas é tudo?
	-	Explicar	X	-	-	T7 Manuela E9 – Não sei… Ou é até *boneca* ou deve ser até *aniversário*.
	-	Explicar	-	-	X	T8 Manuela E10 – *A Eva disse que no seu aniversário* não dá, não é?
	-	Aceitar	-	-	X	T9 Mafalda E11 – Ya… Eu acho que assim está correto. *[sublinham tudo até aniversário]*
	-	Concluir	-	-	X	T10 Margarida E12 – Então vá… Mete tudo.

Nota. SA: sequência argumentativa; EA: episódio argumentativo; EM: enunciado metalinguístico; P: pragmática; S: semântica; F: formal.

A dúvida sobre as fronteiras da subordinada substantiva, primeiro relativamente ao elemento que a encabeça ("Deve ser a partir do *que*, não? T3, E4"), depois, fundamentalmente sobre a integração ou não do Grupo Preposicional no predicado da subordinada ("Ou é até boneca ou deve ser até aniversário." T7, E9) gera um raciocínio suportado por movimentos argumentativos mais diversificados, ainda que bastante cumulativos. Contudo, o diálogo continua a evidenciar colaboração, havendo integração das ideias e a solução encontrada pelo grupo sustenta-se no teste da hipótese de "no seu aniversário" não integrar a subordinada através da sua eliminação. Ainda que este raciocínio seja formalmente marginal e que não haja recurso à substituição da subordinada pelo pronome *-o*, a resposta final é, uma vez mais, a correta (T9 e T10). Note-se ainda que não há qualquer recurso a metalinguagem.

Em contraste, na Tabela 7, o discurso do grupo 3 pauta-se pelo uso de vários termos metalinguísticos.

Tabela 7

Subsequência Metalinguística: Sublinha a Oração Subordinada em "A rapariga que eu vi no teatro não era tua amiga."

SA	EA		EM			Diálogo A (Grupo 3)
	Reativo	Aditivo	Natureza			
			P	S	F	
I	Divergir	-	-	-	-	T1 Maria E1 - "Considera as frases seguintes e sublinha a oração subordinada que cada uma delas contém".
	-	Explicar	-	X	-	T1 Maria E2 - Eu já tinha visto esta e eu acho que é "que eu vi no teatro" porque ela está a restringir.
	-	Expandir	-	-	X	T1 Maria E3 - É uma subordinada restritiva. Eu acho.
	-	Aceitar	X	-	-	T4 Manuel E4 - É, é… Tem o "que".
	-	Expandir	-	-	X	T5 Manuel E5 - Por norma quando tem o "que" é sempre assim aquelas orações subordinadas adjetivas restritivas…
	-	Explicar	X	-	-	T6 Manuel E6 - E eu decorei por causa de quando não tem vírgula é restritiva…
	-	Expandir	-	-	X	T7 Matilde E7 - Ah… *Ya*, isso é como as funções sintáticas.
	-	Aceitar	-	-	-	T8 Maria E8 - Pois…
	-	Aceitar	-	-	-	T9 Matilde E9 - *Ya*, já me lembro desta parte.

Nota. SA: sequência argumentativa; EA: episódio argumentativo; EM: enunciado metalinguístico; P: pragmática; S: semântica; F: formal.

Tal como o padrão discursivo encontrado na primeira parte do episódio A do grupo 2 (Tabela 5), o diálogo do grupo 3 apresenta um discurso prototipicamente cumulativo, com uma série de movimentos aditivos. Não se trata, porém, de fala cumulativa acrítica, há colaboração na validação do raciocínio que fundamenta a identificação da subordinada relativa. Além disso, o enunciado metalinguístico é constituído por verbalizações que permitem inferir conhecimento metalinguístico de natureza complementar (a pista do valor semântico restritivo associado ao conceito sintático de subordinada adjetiva, a pista pragmática do uso de "que"). Salienta-se que, uma vez mais, o conhecimento metalinguístico evidencia algumas conceções expressas de forma marginal, como a relação com "funções sintáticas" (T7) ou a "rule of thumb" que relaciona o uso de vírgulas com as relativas restritivas (T6). Esta ideia do que se "sabe de cor" é ainda reforçada pelas expressões "por norma" (T5), "eu decorei" (T6) ou até "já me lembro" (T9), que permitem inferir uma conceção transmissiva das aprendizagens gramaticais.

A existência de discurso de tipo cumulativo, mas com evidência de progressão colaborativa na resolução do problema coincide com um padrões de tipo [+ aditivos, -reativos], em que a acumulação de episódios aditivos resulta de integração de argumentos validados pelo grupo, tal como se descreve em Fontich (2010a). O predomínio de fala cumulativa (tabelas 5 e 7) pode também ser explicado pela característica da tarefa, que consiste apenas numa operação de identificação. Na secção seguinte, estará em foco uma tarefa que requer operações mais complexas.

5.2. **Episódios da etapa 5 do LG – Validar hipóteses**

Em (2), transcreve-se a instrução da etapa 5 do LG, que requer a validação de hipóteses e o seu registo, tendo em vista a enunciação de generalizações (Duarte, 2008).

(2) De seguida, com as subordinadas que identificaste, constitui os seguintes grupos (Tabela 8):

GRUPO A Subordinadas que são complementos ou sujeitos (são pedidas pelo verbo da oração principal)	GRUPO B Subordinadas que não são sujeitos nem complementos e não podem ser deslocadas	GRUPO C Subordinadas que não são sujeitos nem complementos e podem ser deslocadas
A. *Ele pediu-te que não fumasses tanto.*	B. *A rapariga que eu vi no teatro não era a tua amiga*.	C. *Quando te encontrei no cinema, não te conheci.*

Para resolver o problema colocado por esta tarefa, é preciso realizar várias operações: (i) sintetizar as hipóteses parciais das etapas anteriores, (ii) comparar as propriedades das subordinadas relativamente a função sintática desempenhada e a mobilidade e (iii) categorizar as subordinadas.

A Tabela 8 apresenta o momento da realização da tarefa em que o grupo se foca na análise da primeira frase que não aparece dada como exemplo no enunciado, que é a frase D, a qual integra a subordinada adverbial "Apesar de estar mau tempo".

Tabela 8

Subsequência Metalinguística: Complemento ou Modificador em "D - Apesar de estar mau tempo, o Pedro foi passear de barco."

SA	EA		EM			Diálogo B (Grupo 1)
	Reativo	Aditivo	Natureza			
			P	S	F	
I	Divergir	-	-	-	-	T1 Marta E1 – Agora, a D? *[silêncio]*
	Contradizer	-	-	-	X	T2 Miriam E2 – Hum… Eu acho que, eu ao meu ver, essas... Eu metia todas que eram complemento do verbos.
	Contradizer	-	-	-	X	T3 Micá E3 – Mas se a C está no quadro que não é nada[ii]....
	Esclarecer	-	-	-	-	T4 Miriam E4 – Pois então não sei.
	Esclarecer	-	-	-	-	T5 Marta E5 - Então e a D?
	-	Expandir	-	-	X	T6 Micá E6 - A primeira não é nada.
	Contradizer	-	-	-	-	T7 Miriam E7 - Eu deixava assim, mas tanto faz.
	Esclarecer	-	-	-	-	T8 Micá E8 - Eu já não estou a perceber bem quando não é nada, portanto…
	-	Explicar	-	X	-	T8 Micá E9 – Porque eu acho que isso se refere ao passear, talvez.
	-	Adicionar	-	X	-	T8 Micá E10 – Pois, eu acho que isso se refere ao passear.
	-	Expandir	-	-	X	T8 Micá E11 – Passear o quê? *Apesar de estar mal tempo*.
	Esclarecer	-	-	-	-	T8 Micá E12 – Mas não sei…
	-	Aceitar	-	-	-	T9 Marta E13 – Ok. Então fica assim?
	Esclarecer	-	-	-	-	T10 Micá E14 – Eu não sei.
	Divergir	-	-	-	-	T11 Marta E15 – E a E?
	Esclarecer	-	-	-	-	T12 Micá E16 – Madalena, o que é que achas?
	Esclarecer	-	-	-	-	T13 Madalena E17 – Eu não faço ideia *[risos]*.
	-	Adicionar	-	-	-	T14 Micá E18 – Era só para irmos por maioria, já que também estamos todos no meio da dúvida.
	Esclarecer	-	-	-	-	T14 Micá E19 – Miriam, o que é que achas?
	Esclarecer	-	-	-	-	T15 Marta E20 – O que é que vocês acham?
	Esclarecer	-	-	-	-	T16 Miriam E21 – Eu já nem estou a perceber.
	Esclarecer	-	-	-	-	T17 Marta E22 – Pois nem eu…
	Esclarecer	-	-	-	-	T18 Madalena E23 – Pois, eu também não.
	-	Explicar	-	X	-	T18 Madalena E24 – Eu pensei que como 'tava, tinha o verbo, e o verbo no caso é "passear", isso se deve... Ou "ele fez isso apesar de estar mal tempo". "Ele passeou, apesar de estar mau tempo".
	Esclarecer	-	-	-	-	T18 Madalena E25 – Mas também não sei.
	Contradizer	-	-	-	X	T19 Micá E26 – Mas podia ser que só *O Pedro foi passear*.
	-	Aceitar			X	T20 Miriam E27 – Sim, sim…
	Esclarecer	-	-	-	X	T20 Miriam E28 – Sim, mas estamos a falar do complemento do verbo… Se é complemento do verbo ou se não é nada.
	-	Adicionar	-	-	X	T21 Micá E29 – Sim. Então e se fosse *O Pedro foi passear*?
	-	Aceitar	-	-	X	T22 Miriam E30 – Sim, pode ser.
	Esclarecer	-	-	-	X	T22 Miriam E31 – Mas temos de dizer se é complemento do verbo ou se não é nada.
	Esclarecer	-	-	-	-	T23 Marta E32 – O que é que tu achas, Madalena e Micá?
	-	Concluir	-	-	X	T23 Madalena E33 – Eu acho que não é nada.
	Esclarecer	-	-	-	-	T23 Marta E34 – Apago[iv]?
	-	Concluir	-	-	-	T24 Micá E35 – Acho que é melhor.

Notas (i) SA: sequência argumentativa; EA: episódio argumentativo; EM: enunciado metalinguístico; P: pragmática; S: semântica; F: formal; (ii) No contexto, por "não é nada", entenda-se "não são sujeito nem complemento"; (iii) Marta apaga a frase D do espaço de resposta para os complementos do verbo (Grupo A).

Encontramos neste episódio um novo padrão no discurso dialogal. Trata-se de uma sequência extensa com vários turnos curtos, com movimentos argumentativos dominantemente divergentes, padrão que se aproxima ao identificado como [-aditivo, +reativo] em Fontich (2010a), e que parece ser proporcionalmente inverso ao encontrado para o mesmo grupo na Tabela 4. Neste episódio, embora o grupo se foque num mesmo tópico, que é a resolução do problema, observam-se movimentos de deriva, enunciados por verbalizações como "Eu não

sei", "Eu não faço ideia", evidenciando pouca segurança na apresentação de argumentos. Quanto ao conhecimento metalinguístico, apesar das verbalizações incluírem recorrentemente a termos formais, a dúvida sobre a classificação da subordinada, que reflete uma conceção não estabilizada da diferença entre complementos e modificadores, evidencia um conhecimento marginal, tal como é reportado em outros estudos em cursos de formação de professores (Cardoso et al., 2019; Van Rijt et al., 2021). Ainda assim, o raciocínio sobre a categoria da subordinada evolui ou, por outras palavras, o conhecimento altera-se no fluxo discursivo, passando de (i) um argumento de natureza semântica para validar a hipótese de ser complemento, provavelmente associado à ideia de "completar o sentido" por se "referir a estar mau tempo" (T8, E9-10, retomado por Madalena em T19, E24), para (ii) um argumento de natureza formal, o teste com a interrogativa "o quê", que põe em dúvida a hipótese de ser complemento (T9, E11), e depois para (iii) o teste formal da possibilidade de omissão de modificadores (T19, E26 e 29), que finalmente conduz à conclusão adequada e à exclusão da subordinada adverbial concessiva da lista do Grupo A, onde o grupo a tinha registado. E, mais uma vez, apesar de uma argumentação dominantemente reativa, com muitos movimentos de tipo "esclarecer", o resultado a que o grupo chega é o correto.

Para a mesma tarefa, o grupo 2 só se foca no problema depois de uma longa divergência relativamente ao procedimento a seguir para dar conta da instrução, acabando por pintar em cores diferentes as frases, segundo a sua classificação. No episódio transcrito na Tabela 9, ainda há marcas dessa discussão paralela (por exemplo, no T3, E4).

Tabela 9

Subsequência Metalinguística. Complemento ou Modificador em "A - A rapariga que eu vi no teatro não era a tua amiga."; "F- Os escuteiros que são bons rapazes ajudam velhinhas."; "G - Os escuteiros, que são bons rapazes, ajudam velhinhas." e "H - A criança cujo pai está doente falta hoje à escola."

SA	EA		EM			Diálogo B (Grupo 2)
	Reativo	Aditivo	Natureza			
			P	S	F	
I	Divergir	-	-	-	-	T1 Mafalda E1 - Deixa-me só ler outra vez, porque eu não vi os três grupos. *[silêncio]*
	-	-	-	-	-	T1 Mafalda E2 – *Subordinadas que não são sujeitos nem complementos e não podem ser deslocadas.*
	-	Adicionar	-	-	X	T2 Manuela E3 – Então, os que não são complementos e não podem ser deslocadas…
	Esclarecer	-	-	-	-	T3 Mafalda E4 – Vemos uma de cada vez ou uma a uma?
	-	Explicar	-	X	-	T4 Manuela E5 – … Desculpem. É precisamente aquelas que eu estava a dizer, são as que são [□□□] os modificadores restritivos.
	-	Expandir	-	X	-	T4 Manuela E6 – São restritivos ou apositivos, exatamente. Os modificadores do nome.
	-	Aceitar	-	X	-	T5 Mafalda E7 – hum hum.
	Esclarecer	-	-	-	-	T6 Margarida E8 - Neste caso não é a efe e a guê?
	-	Explicar	-	-	-	T7 Manuela E8 - É o *que eu vi no teatro*, *que ajudam as velhinhas*, *que ajudam as velhinhas* e aquele *cujo pai está doente*.
	-	Aceitar	-	-	-	T8 Margarida E9 – Pois.
	-	Expandir	-	-	X	T9 Manuela E10 - São os modificadores do nome.
	-	Concluir	-	-	X	T10 Margarida E11 – É a A, a efe, a guê e a agá.
	-	-	-	-	X	T11 Mariana E12 – Qual é o grupo? É o amarelinho?
	-	Concluir	-	-	X	T12 Manuela E13 - É o amarelinho.
		Aceitar	-	-	-	T13 Mafalda E14 – Pois é… Eu também não estava a perceber, foi muito rápido a ideia.
	Esclarecer	-	-	-	-	T14 Manuela E15 - Desculpa, percebeste?
		Concluir	-	-	-	T13 Mafalda E14 – Não, agora percebi.

Nota. SA: sequência argumentativa; EA: episódio argumentativo; EM: enunciado metalinguístico; P: pragmática; S: semântica; F: formal

O episódio B do grupo 2 inclui 13 turnos, maioritariamente aditivos, com alguns movimentos reativos, principalmente de esclarecimento, sendo um padrão semelhante ao da Tabela 6. Os elementos do grupo evidenciam um igual envolvimento na resolução do problema, estando a acompanhar o raciocínio de Manuela (T8, E9; T10, E11; T11, E12). Salienta-se desde o início o envolvimento ativo de Mafalda, que só valida a proposta depois de perceber ("Não, agora percebi"). A acompanhar este padrão, verifica-se a emergência de argumentos de natureza variada, um semântico, a pista do valor restritivo ou explicativo das relativas (T4, E5-E6), e outro sintático, o conceito "modificador do nome", introduzido no nono turno e aceite coletivamente. Note-se, porém, que o grupo chega à categorização correta das quatro frases (A, F, G e H) quase sem necessidade de uma reflexão exploratória, aproximando-se mais de uma fala cumulativa, por se tratar de conhecimento metalinguístico estável entre todos os elementos do grupo.

O mesmo não acontece no episódio do grupo 3 analisado na Tabela 4.

Tabela 10

Subsequência Metalinguística: Complemento ou Modificador em "Se estudasse mais, o José teria melhores resultados."

SA	EA		EM			Diálogo B (Grupo 3)
	Reativo	Aditivo	Natureza			
			P	S	F	
I	Divergir	-	-	-	-	T1 Manuel E1 – E eu acho que o próximo também é C.
	-	Aceitar	-	-	-	T2 Maria E2 – OK.
	Contradizer	-	-	-	-	T3 Matilde E3 – *Well*… Não, se estudasses mais… Como é que é o C?
	-	Aceitar	-	-	-	T3 Matilde E4 – Quer dizer… 'Tá bem… É.
	-	Explicar	-	-	X	T3 Matilde E5 – Sim… Dá para trocar…
	-	Acrescentar	-	-	X	T4 Manuel E6 – *O José teria melhores resultados…*
	-	Acrescentar	-	-	X	T5 Maria E7 - *se estudasse mais.*
	-	Adicionar	-	-	X	T6 Manuel E8 – *se estudasse mais.* Pronto.
	Contradizer	-	-	-	X	T7 Maria E9 – Mas é um complemento…
	Esclarecer	-	-	X	-	T7 Maria E10 – Não está a completar? 'Tá!
	-	Aceitar	-	X	-	T8 Matilde E11 – Sim, também está, ya. Ou *se estudasse mais* [xxx].
	-	Explicar	-	X	-	T9 Maria E12 – *O José teria melhores resultados, se estudasse mais.* Ou seja, eu acho que pode modificar. Eu já não sei. Isto dá 'muita' trabalho!
	-	Adicionar	-	-	-	T10 Manuel E13 – Mete onde eu pus! Grupo C. Confia em mim!
	-	Aceita	-	-	-	T11 Maria E14 – *Ya*…[xxx].
	-	Concluir	-	-	-	T12 Manuel E15 – Eu assumo a responsabilidade!

Nota. SA: sequência argumentativa; EA: episódio argumentativo; EM: enunciado metalinguístico; P: pragmática; S: semântica; F: formal.

O padrão dos movimentos argumentativos parece semelhante ao encontrado na tabela anterior, sendo [+aditivo, -reativo]. Verifica-se, contudo, uma alternância mais equilibrada entre movimentos dos dois tipos e argumentos de natureza diferente vão sendo integrados. Matilde, no terceiro turno, introduz a dúvida e, no mesmo turno, recorre ao teste sintático de deslocação (T3, E5), apoiada pelos seus colegas (T4 e T5). Mais uma vez, a emergência de um conhecimento fossilizado (Gauvin & Boivin, 2013), de tipo semântico, correspondente à ideia de que "um complemento completa", surge na argumentação (T7, T8 e T9). Esta ideia é substituída pelo recurso ao conceito enunciado a partir da noção igualmente semântica de que "pode modificar" (T9, E12). E assim, a resposta do grupo acaba por ser a inicialmente proposta, que é, novamente, a opção correta, acompanhada pela ideia de que "Isto dá 'muita' trabalho!"

Todas as respostas finais dos episódios analisados são as corretas. A análise das interações dos grupos, contudo, deu-nos acesso a informação sobre a forma como se desenrola o raciocínio metalinguístico durante a resolução de problemas de gramática e sobre a natureza do conhecimento que fundamenta as soluções encontradas. Na secção seguinte, apresentam-se conclusões e algumas considerações sobre o trabalho que dá aprender e ensinar gramática.

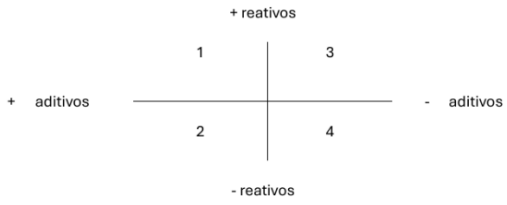
6. Considerações finais

Com o presente estudo de cariz qualitativo/interpretativo sobre interações entre estudantes de um curso de formação de professores durante a realização de um Laboratório Gramatical, procurámos contribuir para aprofundar o conhecimento sobre raciocínio metalinguístico e a natureza das conceções metalinguísticas, tema que tem ganhado o interesse da investigação recente em Linguística Educacional.

Em resposta à nossa primeira questão de investigação, (i) quais as estratégias discursivas de argumentação dominantes no raciocínio metalinguístico destes estudantes de educação?, adotámos o modelo de análise integrada da atividade metalinguística, tendo sido identificados padrões de argumentação no discurso metalinguístico dialogal (Fontich, 2010a, 2010b, 2023). Estes padrões não constituem categorias estanques ou macrocategorias prototípicas, mas antes pontos situados em três dos quatro gradientes representados na Figura 1.

Figura 1

Gradientes em Enunciados Argumentativos



Ao contrário do que se encontra na interação entre alunos do ensino secundário (Fontich, 2010a, 2010b), na interação dos grupos analisados não se identifica um padrão [-reativo, -aditivo], que correponde a turnos muito breves, com frequentes mudanças de tópico e sem progressão coletiva discurso dialogal. A maioria dos episódios analisados (Tabelas 4, 6, 9 e 10) integra um padrão de tipo 1, [+aditivo, +reativo], com uma maioria de movimentos aditivos e alguns movimentos reativos, como "esclarecer", que sustentam uma argumentação pertinente, que integra argumentos de diferentes pontos de vista, com base em conhecimento de natureza diversa. Nestes episódios, mesmo num caso como o da Tabela 4, em que há pouca mobilização de conhecimento formal, é possível observar mudanças no conhecimento verbalizado e aprendizagem, tratando-se de um discurso de natureza exploratória (Mercer, 2003, 2019).

Dois dos episódios, apesar de resultarem em resoluções bem-sucedidas, podem ser situados num padrão típico do discurso cumulativo, num gradiente de tipo 2 segundo o nosso esquema. São os episódios das tabelas 5 e 7, nos quais, após um único movimento divergente de introdução do tópico, há movimentos aditivos subsequentes. Na interação de alunos mais novos, a fala cumulativa é frequentemente uma aceitação desinteressada das propostas dos pares (Mercer, 2019). Porém, encontram-se também casos como estes, nos quais, em sequências aparentemente cumulativas, se vão integrando argumentos metalinguísticos que passam pelo escrutínio coletivo (Fontich, 2010a). Finalmente, na Tabela 8, reconhece-se um padrão de argumentação [+reativo, - aditivo], em que, à semelhança do que se encontra para alunos do secundário (Fontich, 2010a), o grupo mantém o foco, mas acumula movimentos reativos, sendo incapazes de desenvolver conjuntamente a argumentação. Acrescente-se que, neste caso, é mobilizado pouco conhecimento formal, o que contrasta com a tendência geral, observada nos outros dois grupos, de recurso a conceitos formais que acabam por conduzir a respostas corretas. Esta observação é uma primeira resposta à segunda questão de investigação: (ii) o que caracteriza o conhecimento metalinguístico enunciado nas verbalizações de estudantes? Em contraste com outros estudos (Fontich, 2010a; Gauvin & Boivin, 2013), incluindo estudos com estudantes de cursos de

educação (Van Rijt et al., 2021), encontramos nas verbalizações metalinguísticas destes grupos o recurso dominante a termos e conceitos de natureza formal, havendo também a integração de argumentos de natureza semântica. Em alguns casos, durante a argumentação, emergem ideias que podem ser consideradas "rules of thumb" (Van Rijt et al., 2021), ecoando em conhecimento fossilizado, difícil de alterar (Gauvin & Boivin, 2013). Importa ainda destacar a quase total ausência de argumentos de natureza pragmática, que pode refletir uma tradição pedagógica de ensino da gramática na qual não se promove o recurso a dados do próprio uso de língua para fundamentar a reflexão metalinguística.

Passando para a resposta à terceira questão, (iii) quais as vantagens de um modelo de análise de interações sobre gramática que integra simultaneamente o processo de reflexão metalinguística e o processo de emergência de conceitos?, a análise de dados em português e com estudantes do ensino superior contribuiu para refinar o modelo de análise integrada da interação proposto em Fontich (2010a), em primeiro lugar, na medida em que foi aplicado a um novo contexto de interação. Por outro lado, a análise permitiu consolidar a possibilidade de descrever os diferentes tipos de fala em termos de gradientes, evidenciando variáveis de discurso exploratório e cumulativo. Por outras palavras, este modelo faculta uma descrição menos prototipada de discurso cumulativo e exploratório, possibilitando um entendimento mais detalhado dos mecanismos variáveis da fala exploratória, a qual é incubadora de aprendizagem e deve ser intencionalmente promovida em aula. Em trabalho futuro, uma análise quantitativa dos padrões identificados poderá robustecer o modelo.

Para terminar, salientamos que, apesar das fragilidades existentes, não se encontra entre as nossas conclusões o facto de que de que os estudantes "não sabem e deviam saber gramática". Há muito por saber sobre a história da construção do conhecimento metalinguístico. Afinal, como sabemos se sabem gramática? Se não tivéssemos invadido a privacidade dos diálogos dos grupos, o que nos teria chegado eram apenas as respostas corretas. Mesmo que nos tivessem chegado respostas incorretas, o raciocínio podia ser adequado, como na resolução de problemas matemáticos e, nesse caso, teria interessado fazer a história do raciocínio em processo, para compreender como se aprende. A dissociação entre ser capaz de mostrar que se sabe, em respostas-produto, mas não saber fundamentar o que se sabe é, talvez, a principal descoberta e motivo maior para continuar a investigar possíveis etapas da construção do conhecimento e do raciocínio metalinguístico.

Agradecimentos / Financiamento

Ana Luísa Costa é investigadora em Universidade de Lisboa, Faculdade de Letras, Centro de Linguística. Esta investigação recebeu apoio financeiro da FCT (UIDB/00214/2020).

Referências

Barbeiro, L. (1994). *Consciência metalinguística e expressão escrita* [Dissertação de doutoramento não publicada]. Universidade do Minho.

Barbeiro, L. (1999). *Os alunos e a expressão escrita – Consciência metalinguística e a expressão escrita*. Fundação Calouste Gulbenkian.

Barbeiro, L., Pereira, L., Calil, E., & Cardoso, I. (2022). Termos metalinguísticos e operações de natureza gramatical na escrita colaborativa dos alunos do ensino básico / Metalinguistic terms and grammatical choices in the collaborative writing of primary school students. (2022). *TEJUELO. Didáctica De La Lengua Y La Literatura. Educación*, *35*(2), 45–76. https://doi.org/10.17398/1988-8430.35.2.45

Barnes, D., & Todd, F. (2001). *Communication and learning revisited: Making meaning through talk.* Routledge. https://doi.org/10.4324/9781003140832

Bee, T. (2000). Multi-dimensionality of idea framing in group work in academic settings. *Language and education*, *14*(4), 223–249. https://doi.org/10.1080/09500780008666791

Camps, A. (1991). *L'ensenyament de la composició esenta en situació escolar. Desenvolupament y anàlisi de dues seqüències didàctiques d'ensenyament de l'argumentació escrita* [Dissertação de doutoramento não publicada]. Universidad de Barcelona.
Camps, A. (1994). *L'ensenyament de la composició escrita*. Editorial Barcanova.
Camps, A., Guasch, O., Milian, M., & Ribas, T. (1997). Dialogue d'élèves et production textuelle. Activité métalinguistique pendant le processus de production d'un texte argumentatif. *Recherches,* 27, 133–156.
Camps, A. (2014). Metalinguistic activity in language learning. In T. Ribas, X. Fontich, & O. Guasch (Eds.), *Grammar at School. Research on metalinguistic activity in language education* (pp. 25–41). Peter Lang.
Camps, A., & Fontich, X. (eds.) (2020*). Research and teaching at the intersection. Navigating the territory of grammar and writing in the context of metalinguistic activity.* Peter Lang.
Calil, E. (2020). Sistema Ramos: Método para captura multimodal de processos de escritura a dois no tempo e no espaço real da sala de aula. *Alfa*, *64*, e11705.
Calil, E., & Myhill, D. (2020) Dialogue, erasure and spontaneous comments during textual composition: What students' metalinguistic talk reveals about newly-literate writers' understanding of revision. *Linguistics and Education*, *60*, 1–15. https://doi.org/10.1016/j.linged.2020.100875
Cardoso, A., Pereira, S., & Leite, T. (2019). Metalinguistic activity in higher education: thinking about words. [Comunicação oral]. Conference of the International Association for Research in L1 Education. ARLE 2019. Universidade Nova de Lisboa, 26–28 de junho.
Cazden, C.B. (1981). Performance before competence: Assistance to child discourse in the zone of proximal development. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 3*(1), pp. 5–8.
Costa, A.L., & Santos, A.L. (2007). *Orações subordinadas: Que tipo de subordinadas existem?* [material didático disponibilizado em GramáTIC[a].pt] https://area.dge.mec.pt/gramatica/5_matdidactic.htm
Cubero, R. (2005). *Perspectivas constructivistas. La intersección entre el significado, la interacción y el discurso*. Graó.
Cubero, R., Cubero, M., Santigosa, A. S., Benítez, M. D., Carmona, M. J., & Gallardo, M. D. (2008). La educación a través de su discurso. *Revista de Educación*, *336*, 71–104.
Duarte, I. (1992). Oficina gramatical: Contextos de uso obrigatório do conjuntivo. In M. R. Delgado Martins, D. R. Pereira, A. I. Mata, A. Costa, L. Prista, & I. Duarte (Eds.), *Para a didáctica do português. Seis estudos de linguística* (pp. 165–177). Edições Colibri.
Duarte, I. (2008). *O conhecimento da língua: Desenvolver a consciência linguística.* Direção Geral de Inovação e Desenvolvimento Curricular. Ministério da Educação.
Edwards, D., & Mercer, N. (2013). *Common knowledge: The development of understanding in the classroom*. Routledge. https://doi.org/10.4324/9780203095287
Fontich, X. (2010a). *La construcció del saber metalingüístic: estudi sobre l'aprenentatge de la gramàtica d'escolars de secundària en el marc d'una seqüència didàctica* [Dissertação de doutoramento, Universitat Autònoma de Barcelona]. TDX (Theses and Dissertations Online). https://www.tdx.cat/handle/10803/4684#page=1
Fontich, X. (2010b). La construcció del saber metalingüistic en la interacció en petit grup. In O. Guasch & M. Milian (Eds.), *L'educació lingüística i literària en entorns multilingües.* Universitat Autònoma de Barcelona. Servei de Publicacions, LLeRA.
Fontich, X. (2014). Grammar and language reflection at school. Checking out the whats and the hows of gramar instruction. In T. Ribas, X. Fontich, & O. Guasch (Eds.), *Grammar at School. Research on metalinguistic activity in language education* (pp. 244–283). Peter Lang.
Fontich, X. (2021). El verbo, centro de la oración: Los alumnos clasifican los verbos considerando los complementos verbales. In A. Camps y X. Fontich (Eds.), *La actividad metalingüística como espacio de encuentro de la escritura y la gramática: Un itinerario de enseñanza e investigación en educación lingüística* (pp. 312–323). Universidad Nacional de San Juan.

Fontich, X., Troncoso, J., & Pérez-Badenes, J. (2022). Actividad metalingüística en clase de lengua: consideraciones sobre el lugar de la interacción en el aprendizaje de la gramática. *Tavira, 27*, 244–283. https://doi.org/10.25267/Tavira.2022.i27.1105

Fontich, X. (2023). *Grammar in the classroom: A model for analyzing metalinguistic dialogues* [comunicação oral]. 5th Eduling Seminar "Grammar teaching and learning: challenges and opportunities". Universitat Autònoma de Barcelona, Barcelona. 29–30 June 2023.

Fontich, X., & Garcia-Folgado, M. J. (2018). Grammar instruction in the Hispanic area: The case of Spain with attention to empirical studies on metalinguistic activity. *L1 – Educational Studies in Language and Literature, 18, 1–39*. https://doi.org/10.17239/L1ESLL-2018.18.04.02

Garth-James, K., & Hollis, B. (2014). Connecting global learners using elearning and the community of inquiry model. *American Journal of Educational Research, 2*(8), 663–668. https://doi.org/10.12691/education-2-8-15

Gauvin, I, & Boivin, M.-C. (2013). Identifier le verbe : Elaboration des connaissances pas les élèves en classe. *Revue des Sciences de l'éducation, 39*(3), 547–569. https://doi.org/10.7202/1026312ar

Jouili, E., & Elalouf, M.-L. (2021). La conceptualization de la notion de sujet : Étude comparative au cycle 4. *Le français aujourd'hui, 214*(3), 51–67.

Hernández-Rojas, G. (2008). Los constructivismos y sus implicaciones para la educación. *Perfiles Educativos, 30*(122), 38–77.

Kumpulainen, K. & Wray, D. (Eds.). (2002). *Classroom interaction and social learning: From theory to practice.* Routledge-Falmer

Larsen-Freeman, D. (2003). *Teaching language: From grammar to grammaring*. Thomson/Heinle.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Mercer, N. (2019). *Language and the joint creation of knowledge*. Routledge.

Mercer, N. (2013). The social brain, language, and goal-directed collective thinking: a social conception of cognition and its implications for understanding how we think, teach, and learn. *Educational Psychologist*, *48*(3), 148–168. https://doi.org/10.1080/00461520.2013.804394

Mercer, N. (2003). Helping children to talk and think together more effectively. *Polifonia, 7*, 1–26.

Miller, R. (2011). *Vygotsky in perspective*. Cambridge University Press. https://doi.org/10.1017/CBO9780511736582

Myhill, D. (2021). Grammar re-imagined: Foregrounding understanding of language choice in writing. *English in Education, 55*(3), 265–278. https://doi.org/10.1080/ 04250494.2021.1885975

Myhill, D., Jones, S., & Watson, A. (2013). Grammar matters: How teachers' grammatical knowledge impacts on the teaching of writing. *Teaching and Teacher Education*, *36,* 77–91.

Myhill, D., & Newman, R. (2019). Writing talk: Developing metalinguistic understanding through dialogic teaching. In N. Mercer, R. Wegerif, & L. Major (Eds.), *The Routledge international handbook of research on dialogic education* (pp. 255–268). Routledge.

Ribas, T., Farrera, N., & Camps, A. (2020). Metalinguistic Activity in group writing situations: The influence of pedagogic situations and group characteristics. In A. Camps & X. Fontich (Eds.), *Research and teaching at the intersection. Navigating the territory of grammar and writing in the context of metalinguistic activity* (pp. 213–232). Peter Lang. https://doi.org/10.3726/b17237

Riestra, D. (Ed.). (2010). *Saussure, Voloshinov y Bajtin revisitados*. Miño y Dávila.

Schwartz, B., Dreyfus, T., & Hershkowitz, R. (2009). The nested epistemic actions model for abstraction in context. In B. Schwartz, T. Dreyfus & R. Hershkowitz (Eds.), *Transformation of knowledge through classroom interaction* (pp. 11–41). Routledge

Smagorinsky, P. (2011). *Vygotsky and literacy research: A methodological framework*. Brill.

Smagorinsky, P. (2023). *L.S. Vygotsky and English in education and the language arts*. Routledge. https://doi.org/10.4324/9781003374848

Strandberg, A., & van Rijt, J. (2024). Grammatical and rhetorical reasoning in upper secondary students' collaborative talk about a literary text. *Linguistics and Education*, 84, https://doi.org/10.1016/j.linged.2024.101359

Taber, K. S., Ruthven, K., Mercer, N., Hofmann, R., Berry, F. R., & Luthman, S. (2016). Developing teaching with an explicit focus on scientific thinking. *School Science Review, 97*(361), 75–85.

Vandenberg, J., Zakaria, Z., Tsan, J., Iwanski, A., Lynch, C., Boyer, K., & Wiebe, E. (2021). Prompting collaborative and exploratory discourse: An epistemic network analysis study. *International Journal of Computer-Supported Collaborative Learning*, *16*, 339–366. https://doi.org/10.1007/s11412-021-09349-3

Van der Veer, R. (2020). Vygotsky's theory. In M. Harris & G. Westermann (Eds.), The encyclopedia of child and adolescent development (Vol. 3). Wiley. https://doi.org/10.1002/9781119171492

Van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-7912-5

Van Rijt, J., Hulshof, H., & Coppen, P.-A. (2021). 'X is the odd one out, because the other two are more about the farmland'—Dutch L1 student teachers' struggles to reason about grammar in odd one out tasks. *Studies In Educational Evaluation, 70*, 1–13. https://doi.org/10.1016/j.stueduc.2021.101007

Wegerif, R., & Major, L. (2023). *The theory of educational technology: Towards a dialogic foundation for design*. Routledge. https://doi.org/10.4324/9781003198499

Wells, G., & Mejía, R. (2005). Toward dialogue in the classroom: Learning and teaching through inquiry. *Working Papers on Culture, Education and Human Development, 1*(4), 1–45.

Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Harvard University Press.

Anexo

Frases das Etapas 2 e 5 do Laboratório Gramatical "Que tipos de subordinadas existem?" (Costa & Santos, 2007)

A. A rapariga que eu vi no teatro não era a tua amiga.

B. Ele pediu-te que não fumasses tanto.

C. Quando te encontrei no cinema, não te conheci.

D. Apesar de estar mau tempo, o Pedro foi passear de barco.

E. A Manuela foi à praia, embora estivesse frio.

F. Os escuteiros que ajudam velhinhas são bons rapazes.

G. Os escuteiros, que ajudam velhinhas, são bons rapazes.

H. A criança cujo pai está doente falta hoje à escola.

I. Já que fazes bolos bons, encomendo-te um bolo de chocolate.

J. A Eva disse que queria uma boneca no seu aniversário.

K. O professor falou alto para que os alunos o ouvissem.

L. Se estudasse mais, o José teria melhores resultados.

M. Quem não treina o suficiente perde certamente o jogo.

N. Que o Luís tenha reprovado o ano surpreendeu os seus pais.

A importância dos dicionários bilingues de línguas angolanas de origem bantu no sucesso escolar em Angola

The importance of bilingual dictionaries of Angolan languages of Bantu origin for school success in Angola

Toivo Evaristo Haipinge[1], *Raquel Fonseca Amaro*[1]

[1]Centro de Linguística da Universidade NOVA de Lisboa

Abstract

This paper discusses the importance of bilingual dictionaries of Angolan Bantu languages, focusing on their relevance for improving the quality of education and promoting students' linguistic and communicative skills. It primarily aims to enhance communicative and written proficiency in Portuguese—the official language—enabling students to better understand content taught exclusively in Portuguese and addressing school failure within an education system conducted in the official language. However, this system should align with Angolan Bantu languages, such as Oshikwanyama. The main objective of the study is to evaluate existing Portuguese-Oshikwanyama/Oshikwanyama-Portuguese bilingual dictionaries, understand their role in educational strategies in rural areas of Angola, and assess their potential and limitations in improving primary education quality and combating school failure in these regions.

Keywords: bilingual dictionaries, Bantu languages of Angola, pedagogic dictionaries, teaching and dictionaries, Oshikwanyama lexicography.

Resumo

Neste artigo, aborda-se a importância dos dicionários bilingues de línguas angolanas de origem bantu, com foco para a sua relevância na melhoria da qualidade do ensino e na promoção de competências linguísticas e comunicativas dos alunos. Visa-se, sobretudo, a melhoria da proficiência comunicativa e escrita em português, língua oficial, permitindo aos alunos descodificar as matérias lecionadas unicamente em língua portuguesa e o combate ao insucesso escolar, num sistema educacional leccionado na língua oficial, mas que deve estar em harmonia com as línguas angolanas de origem bantu, como é o caso do oshikwanyama. O estudo tem como objetivo principal avaliar os dicionários bilingues português-oshikwanyama/oshikwanyama-português existentes, perceber o seu papel nas estratégias educativas do ensino nas zonas rurais de Angola e a avaliar o seu potencial e limitações para a melhoria da qualidade do ensino primário e para o combate ao insucesso escolar nas zonas rurais de Angola.

Palavras-chave: dicionários bilingues, línguas bantu de Angola, dicionários pedagógicos, ensino e dicionários, lexicografia do oshikwanyama.

1. Introdução

A lexicografia pedagógica é um campo vasto que se ocupa da elaboração, crítica e utilização de dicionários gerais, infantis, escolares, universitários e técnico-especializados para fins didáticos. Os dicionários, para além de ferramentas para a planificação e políticas de línguas, são também instrumentos que contribuem

para políticas educacionais e ensino de línguas, em todos os níveis, e é este o objetivo geral da lexicografia pedagógica enquanto ramo da lexicografia geral (Kirkness, 2004).

Com auxílio de dicionários, sobretudo em contextos multilingues, os professores e alunos preocupam-se em dar resposta aos vários problemas relacionados com o ensino da primeira (L1) e da segunda língua (L2) ou de língua estrangeira. No entanto, principalmente nos casos em que a língua do sistema educativo não é a L1 dos alunos, são também usados para dar resposta a problemas relacionados com o ensino de outras matérias e conhecimentos. E, neste contexto, é muito relevante que os dicionários bilingues sejam úteis não só para a comunicação, mas também para o acesso ao conhecimento, principalmente, no caso de disciplinas em que a língua não é o objeto de estudo.

Considerando a utilização de dicionários bilingues no ensino angolano, o presente estudo concentra-se especificamente na utilização e na importância do dicionário de línguas angolanas de origem bantu no ensino e aprendizagem do português L2 e no ensino de outras matérias lecionadas unicamente na língua portuguesa, como por exemplo em contextos em que os alunos têm o oshikwanyama como língua materna e dominante na sua comunidade. O estudo é pertinente na medida em que fornece uma visão sobre o uso atual dos dicionários bilingues de línguas bantu e portuguesa. Permite ainda identificar lacunas e novas estratégias na utilização desses instrumentos lexicográficos em sala de aula, bem como contributos para a adaptação dos dicionários bilingues com vista a melhorar o ensino em comunidades rurais multilingues de Angola.

Este artigo foca principalmente o uso dos dicionários do oshikwanyama-português/português-oshikwanyama no contexto angolano, bem como as várias estratégias de uso aplicadas em sala de aula pelos professores das escolas do ensino primário nas zonas rurais da província do Cunene. Atenta, também, na diversidade linguística angolana e no modo como o ensino em português ocorre em contexto social multilingue. Os professores de diferentes escolas das zonas rurais naquela província precisam de lidar com falta de equivalentes e de encontrar equivalentes substitutos, funcionando muitas vezes como mediadores entre a língua de ensino e a língua materna dos alunos. Os professores não falantes nativos ou não proficientes em oshikwanyama têm de recorrer ao conhecimento intuitivo dos seus alunos, ou ainda, a estratégias visuais (ex.: figuras dos manuais) e a exemplos concretos do dia a dia para conseguirem passar o conhecimento. Esta situação é agravada ainda pelo facto de os dicionários bilingues a que os professores recorrem não serem adequados. Os poucos dicionários existentes apresentam vários problemas que vão desde a falta palavras de especialidade usadas no ensino, à falta de sistematização e adequação dos conteúdos. Neste artigo, apresenta-se uma análise preliminar deste contexto, necessária à elaboração de um instrumento lexicográfico bilingue focado nas necessidades dos alunos, que considere o léxico usado no ensino primário angolano, mas que esteja também em conformidade com os princípios das funções cognitiva e comunicativa necessárias para a produção e transmissão de conhecimento em sala de aula.

2. Enquadramento

2.1. Contexto multilingue angolano

De modo geral, a África caracteriza-se por uma grande diversidade étnica, o que permitiu que houvesse, em muitos Estados africanos, uma diversidade linguística na qual as línguas de origem bantu coabitam com outras línguas europeias e asiáticas. Esta diversidade deriva de um complexo processo de colonização e de contacto com outros povos do globo.

Segundo Petter (2015, p. 60), as línguas africanas faladas em Angola fazem parte da família bantu "benué-congo". Esta família de línguas tinha no século XX mais de 500 milhões de falantes, numa área bastante extensa que compreende aproximadamente dois terços do sul da Nigéria e da República dos Camarões, a parte Sul da República Centro-Africana e do Congo, grande parte da República Democrática do Congo, Tanzânia, Uganda, Quênia, Ilhas Comores, Moçambique, Angola, Ruanda, Burundi, Namíbia, Zâmbia, Maláui, Zimbábue,

Botsuana, Suazilândia, África do Sul, Lesoto, Guiné Equatorial e Gabão, assim como uma pequena parte da Somália (Petter, 2015, p. 60).

É no prisma do pluralismo etnolinguístico africano defendido por Petter (2015) que buscamos trazer uma breve contextualização das línguas faladas em Angola. Como sabemos, Angola é um país multiétnico e multilingue. Há em Angola várias famílias de línguas: a europeia; a asiática; a khoisan, dividida em nkhu e khwe; e a bantu.

As três grandes zonas linguísticas angolanas – H, K e R – indicadas na classificação de Guthrie (1967) e Andrade (2007) compõem-se das seguintes línguas:

i. zona H, a que pertencem os grupos linguísticos kikongo, kimbundu e kiyaka;

ii. zona K, a que corresponde o grupo linguístico tchokwe-luchaze;

iii. zona R, onde temos os grupos linguísticos umbundu, olunyaneka-nkumbi, oshierero e oshiwambo, que integra a língua oshikwanyama, em análise neste estudo, junto de outras línguas africanas de origem bantu.

Andrade (2007, p. 59) concluiu que são nove as principais línguas angolanas de origem bantu, a saber: o tchokwe (cokwe) do grupo Tucokwe, o kimbundu do grupo ambundu, o kikongo do grupo Bakongo, o nganguela do grupo vanganguela, o olunyaneka-nkumbi do grupo ovanyaneka-nkumbi, o herero do grupo ovahelelo, o oshikwanyama dos ovakwanyama, o oshindonga do grupo ovandonga e o umbundu do ovimbundu respetivamente. Estas, por conseguinte, são as mais descritas na literatura angolana, como ilustra a distribuição linguística de Angola no mapa da figura 1, que se segue:

Figura 1

Mapa Etnolinguístico de Angola.



Nota. Fonte: Fernandes e Ntondo (2002)

O oshikwanyama é uma língua falada no sudoeste de Angola, administrativamente na província do Cunene, e no norte da Namíbia. Guebe (2008), Wheeler e Pelissier (2013) e Haipinge (2017) consideram que o oshikwanyama é a língua dos Ovakwanyama, povo situado nas regiões Sul, Centro e Este da província do Cunene. Esta região situa-se mais a Sul de Angola, e o grupo estende-se até à zona Norte da República da Namíbia. De acordo com os dados do *Ethnologue* (Eberhard et al., 2024), os Ovakwanyama perfazem um total de 1.481.000 habitantes nos dois países, e [kua] é o seu código linguístico internacional.

Assim como as restantes oito línguas bantu angolanas acima, o oshikwanyama também foi introduzido no sistema educativo angolano, nos meios de comunicação social, bem em alguns documentos das instituições do Estado angolano. Por exemplo, a Constituição da República de Angola, a cartinha dos 11 compromissos da criança e o hino nacional foram traduzidos nesta língua. É considerando a presença e o peso do oshikwanyama no contexto angolano que procuramos analisar os dicionários bilingues de língua oshikwanyama e portuguesa, e vice-versa, com foco na sua utilização no ensino primário das áreas rurais da província do Cunene, no contexto da diversidade linguística maior do país.

Apesar de em Angola existir uma vasta gama de línguas bantu e não bantu, a língua portuguesa ocupa o estatuto de língua oficial. Embora o português tenha surgido no contexto colonial, é hoje a língua mais falada em Angola, por "71% da população, com a maior incidência nos centros urbanos onde 85% da população é falante do português" (Instituto Nacional de Estatística, 2016, p. 50). O português tem o estatuto de língua de Angola de origem não bantu (europeia). Além dos factos acima listados, a língua portuguesa é também a língua de unidade nacional e constitui um património cultural imaterial que todos os angolanos devem preservar e promover.

2.2. **Multilinguismo e o ensino no meio rural**

As crianças das zonas rurais, falantes nativas de línguas angolanas de origem bantu, sobretudo as falantes de oshikwanyama, ao ingressarem no ensino primário nas suas localidades enfrentam enormes dificuldades na compreensão dos conteúdos transmitidos na língua portuguesa. Apesar de língua oficial, é na escola que estas crianças têm o primeiro e principal contacto com o português. No entanto, os conteúdos da disciplina de Português, e o ensino da língua portuguesa em Angola, são lecionados como se tratasse de ensino de L1, o que nem sempre é o caso. É com base nesta situação linguística que Severo et al. (2019) constataram que, em Angola, "o professor da escola rural se vê diante de um universo linguístico no qual os alunos se tornam estrangeiros em sua própria terra, uma vez que a língua e a cultura difundidas pela escola não os representam. O professor muitas vezes busca adotar uma pedagogia baseada na negociação entre a língua portuguesa, a língua de ensino e a língua nacional" (Severo et al., 2019, p. 300).

Por sua vez, Ndombele e Timbane (2020, p. 309) constataram que, no processo de ensino-aprendizagem do português em Angola, se observa que os alunos de todas as classes e em todas as disciplinas do ensino primário têm dificuldades na leitura, na escrita e na expressão/compreensão oral. Diante desta situação, e para amenizar as dificuldades dos alunos, os professores do ensino primário angolano das zonas rurais veem-se obrigados a recorrer a novos instrumentos didáticos não necessariamente contemplados no contexto ou currículos escolares, como por exemplo, dicionários bilingues de línguas angolanas de origem bantu.

Esta medida, adotada pelos professores nestas zonas, visa permitir desenvolver competências comunicativas que incluem "a capacidade de os alunos adequarem os seus discursos às situações em que se encontram e ao meio de comunicação utilizado na escola" (Ndombele & Timbane, 2020, p. 309).

De igual modo, Scantamburlo (2002) constatou na Guiné-Bissau que a língua portuguesa é uma língua desconhecida pelos alunos das escolas das zonas rurais e que, por esta razão, impacta negativamente no processo de ensino e na aquisição de novos conhecimentos. Este é o dilema que também se observa em Angola, onde o estudante do meio rural é obrigado a aprender novos conhecimentos por meio de uma segunda língua que não conhece. Em Angola, a língua portuguesa em ensinada como L1 ou língua materna dos alunos, não havendo distinção entre os alunos que têm o português como L1 os que têm o português como L2 aprendida em idade

escolar. Por esta razão, os alunos das escolas das zonas rurais da província do Cunene têm maiores dificuldades em comunicar, compreender e aprender os conhecimentos lecionados.

Figura 2

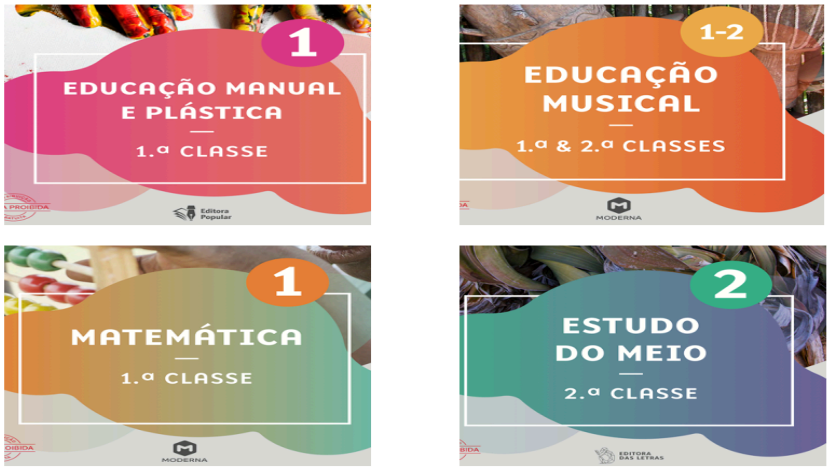
Manuais de Língua Portuguesa do Ensino Primário Angolano.



Nota. Fonte: Ministério da Educação de Angola (2021f, 2021g)

Figura 3

Manuais do Ensino Primário Angolano de Outras Disciplinas.



Nota. Fonte: Ministério da Educação de Angola (2021a, 2021c, 2021h, 2021d)

As observações de Scantamburlo (2002), Severo et al. (2019) e Nombele e Timbane (2020), acima, fizeram-nos despertar o interesse pela avaliação do papel que os dicionários bilingues de línguas angolanas de origem bantu, sobretudo os de oshikwanyama–português e português-oshikwanyama, têm tido no ensino e no

combate ao insucesso escolar, com realce para as escolas do ensino primário das áreas rurais da província angolana do Cunene.

No que respeita ao ingresso dos profissionais no ensino, em Angola não há nenhuma obrigatoriedade de que o professor da escola rural tenha a mesma língua materna que a dos alunos ou que seja proficiente na língua dominante na região em que ensina. Por esta razão, o professor pode vir de outra zona do território nacional e ser falante de outras línguas. Este facto coloca os dicionários bilingues de línguas angolanas de origem bantu em situação de instrumentos facilitadores de comunicação e transmissão de conhecimentos entre os alunos e os professores de línguas bantu diferentes ou que são nativos ou proficientes em português, língua que não é a língua materna dos alunos das escolas rurais do Cunene.

2.3. **Lexicografia bilingue e pedagógica**

Inicialmente, como disciplina geral, a lexicografia criou e promoveu um conjunto de métodos que visam responder às necessidades relacionadas com as funções de produção e recepção dos falantes através de dicionários, tipicamente monolingues (Lara, 2004, p. 148; Bergenholtz & Gows, 2012).

Da lexicografia geral nasceu a lexicografia bilingue, mais destinada a dar resposta a necessidades comunicativas considerando dois sistemas linguísticos diferentes. A lexicografia bilingue possui uma longa tradição de elaboração de dicionário bilingues, e continua a desenvolver-se cada vez mais, com a descoberta de novos métodos e técnicas. O surgimento da lexicografia bilingue deveu-se a lacunas nos dicionários monolingues, com destaque para as necessidades pedagógicas e comunicativas entre indivíduos (Holnselaar, 2003, p. 323; Carvalho, 2001).

Os dicionários bilingues podem ser classificados de acordo com várias tipologias baseadas em caraterísticas composicionais do dicionário, perspectiva de apresentação, funções, usos contextuais, disponibilização de informações, público-alvo, suporte (digital ou eletrónico), número de línguas contempladas, número de verbetes, descrição e percurso (Hartmann & James, 1998, pp. 6–7; Correia, 2009, pp. 32–46; Krienger, 2011, pp. 147–148).

Os dicionários bilingues para fins pedagógicos visam preencher as lacunas que existem entre os conhecimentos dos seus leitores e o de toda uma comunidade, seja este o conhecimento relacionado com a língua ou com a ciência. O dicionário tem naturalmente uma dimensão lexical e cultural na aprendizagem de uma língua segunda ou estrangeira, mas os dicionários também se dedicam à exploração e esclarecimento dos elementos linguísticos e extralinguísticos associados ao léxico (Dubois & Dubois, 1971, p. 7).

O uso de dicionários pedagógicos e de outras obras de informação lexicográfica de referência devem ser vistos como uma fonte de enriquecimento para o aluno (Grandi & Nadin, 2020). Os dicionários são importantes porque, além de permitirem a construção do conhecimento, também são instrumentos normativos das línguas não padronizadas ou em reestruturação da norma. De acordo com Verdelho & Silvestre (2011), a finalidade dos dicionários, sobretudo bilingues, é codificar e descodificar as aprendizagens, apoiando o autodidatismo e melhorando a escrita, a leitura e a comunicação entre indivíduos nas sociedades.

Segundo Ferreira e Vieira (2013), os dicionários bilingues são importantes no ensino de línguas e de outras disciplinas, pois "quanto mais diversificado é o vocabulário de um aprendente, maior será a sua capacidade de se expressar, de compreender e de utilizar corretamente a língua" (Ferreira & Viera, 2013, p. 31).

Em suma, os dicionários bilingues têm como objetivo favorecer a compreensão dos conteúdos, a descodificação das palavras, a pronúncia, a escrita correta das palavras, a melhoria da qualidade ensino/aprendizagem da língua, e os dicionários pedagógicos podem ser essenciais no combate ao insucesso escolar (Verdelho & Silvestre, 2011; Vilela, 1979). No entanto, a lexicografia das línguas africanas depara-se com grandes desafios no ensino, como veremos adiante.

2.3.1. A lexicografia das línguas africanas

A lexicografia das línguas africanas, em que se inclui o oshikwanyama, não pode e nem deve existir isoladamente. Ela está sujeita às mesmas questões e tendências dos utilizadores dos diferentes tipos de dicionários, mudanças e desenvolvimentos, tal como ocorre com a restante lexicografia mundial.

Os primeiros dicionários das línguas africanas foram feitos por europeus sob influência da corrente lexicográfica eurocêntrica (Prinsloo & Zondi, 2020). No entanto, a lexicografia das línguas africanas tem os seus próprios desafios, principalmente, pelo facto de tratar de línguas com poucas produções científicas e recursos linguísticos e com menos prestígio/poder económico.

As línguas de países com poucos recursos económicos geralmente carecem de dicionários de alto padrão lexicográfico, resultantes de uma combinação de critérios lexicográficos, linguísticos e extralinguísticos. As línguas africanas são "línguas com número limitado de falantes, línguas faladas por comunidades subdesenvolvidas, submetidas à opressão política e sem ortografia padronizada, com falta de recursos financeiros, de *corpora* textuais e orais, falta de uma cultura de dicionário, falta de ferramentas e tecnologias de processamento de língua natural" (Prinsloo, 2017, p. 20, tradução nossa).

Assim como outras línguas bantu, o oshikwanyama também partilha estas condicionantes. Segundo o *Ethnologue* (Ehberhard et al., 2024), tem um total de 1 481,000 de falantes nos dois países, isto é, em Angola e na Namíbia. As línguas bantu da família Níger-Congo, da qual o oshikwanyama é parte, na sua maioria têm sistemas gramaticais bastante complexos, com realce para a classificação de nomes de diferentes classes com reflexo no sistema de flexão, um sistema pronominal complexo, questões de tradição de lematização e problemas com a normalização dos seus sistemas ortográficos.

A ortografia da língua Oshikwanyama passou durante séculos por várias transformações. A autora defende-se ainda que esta língua tem uma utilidade em várias áreas, na República da Namíbia e na República de Angla, como por exemplo no ensino, no comercio e na religião só para citar (Halme, 2004).

No alfabeto do Oshikwanyama não figura o grafema /R/ pelo fato de não existir esse fonema nessa língua; quanto ao /S/ foi substituído pelo dígrafo /sh/ resultante do acordo ortográfico de 2006.

Quanto à ortografia, um dos grandes problemas que oshikwanyama enfrenta enquanto língua transfronteiriça é a definição e adoção do sistema de escrita. A República da Namíbia usa o sistema de escrita disjuntiva e em Angola recorre-se à escrita conjuntiva, sobretudo quando se trata dos verbos (Shikesho, 2020, p. 1). Pode-se verificar os dois sistemas nos exemplos abaixo:

(1) *Ohatu **i** ku lipi* = Para onde vamos? (escrita disjuntiva, Namíbia)

(2) *Ohatui kulipi* = Para onde vamos? (escrita conjuntiva, Angola)

Esta dicotomia de sistemas escrita afetou também, em certo momento, o changana, língua moçambicana da mesma família que o oshikwanyama. Por exemplo, a escrita disjuntiva dos verbos e dos nomes é problemática para o changana porque retira as marcas gramaticais à forma verbal, tornando difícil distinguir entre verbos e nomes. Por esta razão, Ngunga & Simbine (2012, pp. 74–75) sugeriram a escrita conjuntiva para os nomes e verbos do changana e a escrita disjuntiva para outras palavras. Esta proposta também poderia aplicar-se ao oshikwanyama, considerando o facto de que, nas línguas bantu, os verbos fazem-se sempre acompanhar dos seus elementos flexionais de concordância e de sujeito gramatical.

No entanto, haver diferentes propostas de escrita para a mesma língua tem influência direta nos recursos linguísticos disponíveis, incluindo a usabilidade de ferramentas de tratamento automático e semiautomático das línguas, que são hoje de grande importância na lexicografia moderna.

De acordo com Prinsloo (2015), os dicionários de línguas bantu tendem a dar apenas informações básicas para o consulente. É notório que os dicionários de línguas africanas apresentam um número reduzido de informações para os utilizadores, estando estas lacunas de informação tipicamente relacionadas com a origem dos dicionários. Ou seja, os dicionários de oshikwanyama foram produzidos por pessoas que tinham poucos

conhecimentos sobre a língua-alvo e que, por isso, disponibilizam informações básicas e, por vezes, incongruentes ou incorretas. Estas lacunas de informação nos dicionários de oshikwanyama não são um caso isolado. Como referido acima, é um problema que afeta a lexicografia africana no geral (Prinsloo 2015, 2017; Prinsloo & Zondi, 2020).

A falta de informações de tipo morfossintático e de contexto de uso fizeram com que a importância de muitos dicionários de línguas bantu fosse desconsiderada no ensino das línguas e de outras matérias pela comunidade académica em África.

Finalmente, é ainda necessário considerar questões de caráter legal e de políticas de língua que afetam o oshikwanyama, e demais línguas angolanas de origem bantu, tais como:

i. Reconhecimento oficial e estatuto legal
Nem todos os países reconhecem o estatuto legal do oshikwanyama e este fator limita o seu uso na administração pública, justiça, educação e média nacionais e internacionais. A ausência de reconhecimento por parte de governos pode, ainda, constituir uma violação dos direitos linguísticos de comunidades falantes, como o direito ao uso da língua materna no acesso a serviços básicos.

ii. Educação bilingue ou multilingue
O reconhecimento legal do ensino bilingue e multilingue, a par da distinção entre falantes L1 e L2 da língua de ensino oficial, o português, é uma questão crítica. Sem políticas claras, as crianças têm dificuldades de aprendizagem.

iii. Normalização ortográfica e padronização e adoção como língua de ciência de conhecimento
É importante que o oshikwanyama e demais línguas bantu sejam usadas pela administração pública e pela comunidade científica, em documentos, manuais, ensino. Mas para tal é preciso que a ortografia do oshikwanyama esteja normalizada e padronizada e tudo isto despende do governo ou de instituições académicas com valências em linguística vocacionadas para a criação de normas e dicionários oficiais.

iv. Direitos linguísticos e antidiscriminação
O governo angolano precisa de promover leis contra a discriminação local das línguas bantu e equiparar essas línguas à língua portuguesa como línguas oficiais do Estado. O governo deve igualmente proteger os falantes de línguas locais contra a marginalização ou exclusão social. O respeito pelos direitos linguísticos das populações passa por permitir o uso do oshikwanyama em documentos oficiais, como certidões, contratos ou pleitos eleitorais, promover intérpretes em tribunais ou hospitais e promover condições de ensino iguais para todos.

3. Objetivos de investigação e metodologia de trabalho

Considerando o enquadramento do oshikwanyama no contexto multilingue de Angola, bem como os recursos lexicográficos disponíveis para as línguas africanas de origem bantu, as questões de base no desenvolvimento da lexicografia e na descrição destas línguas para efeitos pedagógicos e de comunicação bilingue, são várias as questões de investigação que se levantam. No presente trabalho, enquadrado num projeto maior que visa a elaboração de uma proposta de dicionário bilingue escolar português-oshikwanyama que dê conta do léxico usado no ensino primário angolano, pretende-se contribuir para responder às seguintes questões:

i. De que forma é que os dicionários bilingues de oshikwanyama-português/português-oshikwanyama são utilizados pelos professores e alunos em sala de aula, no ensino primário nas áreas rurais da província angolana do Cunene?

ii. Que estratégias são aplicadas pelos professores para que estes instrumentos sejam úteis ou eficazes na fluidez da comunicação entre professores e alunos na escola e comunidade?

iii. De que modo é que os dicionários bilingues oshikwanyama-português/português-oshikwanyama têm sido úteis no combate ao insucesso escolar nas áreas rurais da província do Cunene?

A metodologia deste estudo contempla dois passos essenciais:

i. A análise de recursos lexicográficos existentes para o par de línguas português/oshikwanyama e outros comparáveis (línguas europeias - línguas bantu). Sempre que possível, outros trabalhos semelhantes sobre dicionários bilingues de línguas bantu e portuguesa foram convocados para referência e comparação.

ii. A análise do uso didático dos dicionários bilingues oshikwanyama-português/português-oshikwanyama em aulas gravadas em vídeo.

As aulas analisadas serviram como ensaio prático para observar a utilização destes recursos lexicográficos em sala de aula enquanto recurso didático-pedagógico usado recorrentemente pelos professores das zonas rurais no Cunene, no processo de ensino-aprendizagem do português e de outras matérias lecionadas unicamente na língua portuguesa. A análise qualitativa das interações em sala de aula, dos momentos de consulta aos dicionários e do desempenho linguístico dos alunos e dos professores revelou-se mais adequada para atingir os objetivos deste estudo do que a aplicação de instrumentos tradicionais de coleta de dados, como é o caso dos questionários. Assim, os vídeos forneceram informações ricas e contextualizadas para uma compreensão mais profunda da prática didática e do papel da utilização dos referidos dicionários no ensino.

A gravação das referidas aulas contemplou aulas de 4 professores falantes nativos de oshikwanyama, 3 professores falantes não nativos e 3 professores não falantes de oshikwanyama. As 10 escolas e os 10 professores selecionados para este estudo, foram escolhidos aleatoriamente pela Direcção Municipal da Educação de Namacunde, num total de 26 escolas com 156 professores colocados no meio rural. As gravações das referidas aulas foram disponibilizadas via WhatsApp.[1] É, ainda, importante neste estudo salientar que as gravações de aulas, a qual recaiu a nossa a análise, foram consentidas pelos participantes, isto é, os professores falantes nativos de oshikwanyama, falantes não nativos e os não falantes da língua oshikwanyama. Apesar dos participantes terem aceitado que as suas aulas fossem gravadas pelos membros da Direcção Municipal da Educação de Namacunde, instituição que vela pelo controlo e pelo normal funcionamento das escolas do ensino primário naquele município da província do Cunene. O certo, é que os membros da referida Direcção que efetuaram tais gravações, decidiram não anunciar de antemão o proposito real das gravações para se evitar os viés de comportamento por parte dos professores durante as aulas.

Espera-se que deste trabalho resulte a caraterização do panorama atual bem como a identificação de estratégias de utilização de dicionários bilingues em sala de aula, que possam contribuir para a qualidade do ensino primário em Angola, melhorando a comunicação entre professores e alunos, e a assimilação, a acomodação e a reprodução de conhecimentos por parte dos alunos destes níveis de escolaridade.

4. Análise crítica de dicionários português-oshikwanyama/oshikwanyama-português

Esta secção apresenta uma abordagem crítica aos dicionários bilingues português-oshikwanyama/oshikwanyama-português, focando principalmente o contexto histórico da sua elaboração e os procedimentos metodológicos utilizados. Apresenta-se, igualmente, uma classificação destas obras baseada nos

[1] Vídeos disponibilizados via WhatsApp pela Direcção Municipal da Educação de Namacunde, em 2024.

princípios da lexicografia africana propostos por Prinsloo e Zondi (2020) e avaliam-se as suas limitações. A secção enfatiza a falta de participação de especialistas em lexicografia e ensino, fator que compromete em muito a qualidade técnica dos dicionários. Além disso, aponta-se para as insuficiências dessas obras em atender plenamente às necessidades dos utilizadores, destacando a urgência de um dicionário moderno, equilibrado e pedagógico, voltado especialmente para as necessidades da comunidade educativa do ensino primário em Angola.

4.1. **Correntes lexicográficas eurocêntrica e afrocêntrica**

À semelhança de muitos dicionários de outras línguas bantu, também os dicionários da língua oshikwanyama têm origem na lexicografia eurocêntrica. Desde a sua génese, que a lexicografia do oshikwanyama produziu poucos dicionários, quer monolingues quer bilingues. Os trabalhos lexicográficos sobre o oshikwanyama analisados permitem-nos concluir que a lexicografia do oshikwanyama se pode também dividir em duas correntes: a lexicografia colonial ou eurocêntrica e a lexicografia pós-colonial ou afrocêntrica (Prinsloo & Zondi, 2020).

A lexicografia pré-colonial ou eurocêntrica do oshikwanyama é marcada pelo aparecimento das primeiras obras lexicográficas que foram elaboradas por europeus, sobretudo pelos missionários que muito se dedicaram ao estudo e à elaboração de dicionários bilingues oshikwanyama e línguas europeias. Nesta, incluem-se os seguintes dicionários:

- o dicionário bilingue *Worterbuche der Vambo-Spprache: Osikuanjama – Deutsch de Tonjes* (1910), que é a primeira obra lexicográfica bilingue unidirecional, publicada em Berlim, da era colonial, no Sudoeste africano (atual Namíbia);
- o *Dicionário Português-Oshikwanyama - Pequeno Método de Aprender Português para o Uso dos Povos do Cuanhama* (Sul de Angola), de Leconte (1949), publicado em Angola em Sá da Bandeira: Missão da Huíla;
- o *English-Oshikwanyama Dictionary* de Tobias & Turvery (1965), dicionário unidirecional que apresenta as categorias morfológicas de palavras inglesas e os seus equivalentes em Oshikwanyama.

Os dicionários acima listados foram elaborados no contexto colonial, ou seja, no âmbito da lexicografia eurocêntrica que se destinava a facilitar a comunicação e a evangelização entre os colonizadores e os colonizados. Os autores destas obras, não sendo especialistas em lexicografia ou linguística, ainda que manifestassem um nível alto de conhecimento sobre as línguas estudadas, serviram-se dos parâmetros das línguas europeias e tomaram como exemplos dicionários e tradições lexicográficas europeus. Por esta razão, apesar de se poderem configurar fontes de conhecimento, não constituem recursos lexicográficos de qualidade ou apropriados para os nossos objetivos, razão que motiva a elaboração de uma proposta de dicionário bilingue escolar português-oshikwanyama que dê conta do léxico usado no ensino primário angolano, projeto maior em que se enquadra o estudo aqui apresentado.

A lexicografia pós-colonial ou afrocêntrica do oshikwanyama manifesta-se nos dicionários bilingues português-oshikwanyama/oshikwanyama-português elaborados por falantes nativos do oshikwanyama, dos quais se espera um conhecimento mais rigoroso dos princípios e estratégias linguísticas, quer da língua portuguesa, quer do sistema linguístico do oshikwanyama, e da adequada representação desse conhecimento. A lexicografia afrocêntrica do oshikwanyama inclui os seguintes dicionários:

- O *Dicionário de Nomes em Oshikwanyama*, de Ndanyakukwa (2016), uma obra lexicográfica bilingue antroponímica, unidirecional (Oshikwanyama-Português), publicada em Angola pela Mayamba;

- o dicionário *Onyika Yetu yOshikwanyama mOshiputo* (*Pequeno dicionário Oshikwanyama – Português*), de Lidinikeni (2017).
- O *Oshikwanyama – English, English – Oshikwanyama Dictionary*, de Zimmerman, Taapopi & Hasheela (2019), publicado na República da Namibia pela Macmillan Educaction é uma obra bidirecional que resultou da continuação do projeto dicionário de: *English-Oshikwanyama Dictionary* de Tobias & Turvery (1965).

No nosso prisma, estabelece-se assim uma divisão para a lexicografia do oshikwanyama, ancorada na divisão da lexicografia africana feita por Prinsloo & Zondi (2020). Dentro das divisões da lexicografia do oshikwanyama acima apresentadas, no presente trabalho, focaremos apenas os dicionários bilingues do oshikwanyama-português e português-oshikwanyama, nomeadamente:

- o "*Dicionário" Português-Oshikwanyama - Pequeno Método de Aprender Português para o Uso dos Povos do Cuanhama* (Sul de Angola) de Leconte (1949), publicado em Sá da Bandeira: Missão da Huíla. Este é um dicionário elaborado no contexto colonial, cujo objetivo era facilitar a comunicação entre os padres e os alunos que ingressavam nas escolas das missões católicas nas regiões da Huíla e do Cunene; e,
- o dicionário *Onyika Yetu yOshikwanyama mOshiputo* [Pequeno dicionário Oshikwanyama – Português] de Lidinikeni (2017). Este é um dicionário bilingue de língua comum, elaborado para pessoas interessadas em aprender o português e o oshikwanyama. Apresenta equivalências linguísticas entre as duas línguas e paráfrases de significados para algumas entradas. Contudo, o dicionário não dá indicação das categorias morfossintáticas das palavras.

Os dois instrumentos lexicográficos bilingues em análise são os utilizados nas escolas das áreas rurais do ensino primário na província do Cunene.

Como brevemente descrito acima, o ensino primário angolano ocorre num contexto de livre circulação nacional, em que todos os cidadãos angolanos se deslocam e fixam livremente no território nacional o que contribui, de certo modo, para ampliar o multilinguismo que se verifica nas diferentes zonas da província do Cunene. É, pois, nesta perspectiva que procuramos analisar os dicionários bilingues do oshikwanyama-português/português- oshikwanyama, da autoria de Leconte (1949) e de Lidinikeni (2017), e a adequação e eficiência da sua utilização em contexto escolar na província do Cunene.

4.2. **Descrição e análise de dicionários existentes**

Nesta secção, faz-se uma abordagem crítica dos dois dicionários bilingues do oshikwanyama-português/português-oshikwanyama, com foco na identificação e na sistematização das lacunas existentes para um uso eficaz destes instrumentos lexicográficos bilingues enquanto instrumentos que visam a melhoria da qualidade do ensino nas áreas rurais e, consequentemente, destinados ao combate do insucesso escolar naquelas áreas da província do Cunene.

Os dicionários são ferramentas linguísticas essenciais para o armazenamento e organização do léxico de uma língua natural. Eles fornecem ao consulente informações fundamentais para o seu desenvolvimento científico e comunicativo na sociedade. No entanto, a sua estrutura e função podem variar conforme o propósito e o público-alvo.

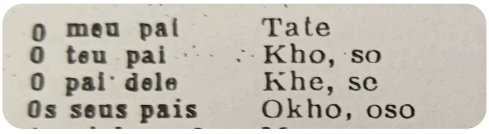
Os dicionários de português-oshikwanyama/oshikwanyama-português analisados neste estudo apresentam algumas inconsistências, especialmente no que se refere a lacunas ao nível do léxico registado, com destaque para palavras usadas no ensino primário angolano. Além disso, observa-se a falta de informações gramaticais,

como flexão verbal e nominal (género e número), bem como a ausência de exemplos de uso para os equivalentes em oshikwanyama, enquanto língua de chegada no dicionário português-oshikwanyama.

Também não há imagens a ilustrar os significados das palavras, já que não é condição obrigatória os dicionários bilingues terem o suporte de imagens que auxiliem na explicitação/explicação de palavras sem equivalente total ou parcial no oshikwanyama e no português. Ademais, as informações dos dicionários nem sempre são apresentadas da mesma forma ou pela mesma ordem, e ortografia do oshikwanyama encontra-se desatualizada em relação à ortografia oficial assinada entre Angola e a Namíbia em 2006, conforme ilustrado nas Figuras 4, 5 e 6; ao invés de propostas nossas, Tabelas 1, 2 e 3, respetivameente..

Figura 4

Falta de Informação ou de Sistematização da Informação Dada no Dicionário



Nota. Fonte: Dicionário Português-Oshikwanyama (Leconte, 1949).

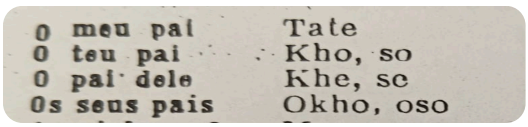
Tabela 1

Proposta Nossa

Singular	Plural
(1) o meu pai *tate wange* *tate*: 1.ª pessoa do singular *wamge*: determinante possessivo 1.ª pessoa do singular tradução literal: pai meu = o meu pai "tate wange"	(4) o nosso pai tate wetu *tate*:1.ª pessoa do singular *wetu*: determinante possessivo da 1.ª plural Tradução literal: pai nosso = o nosso pai *"tate wetu"*
(2) o teu pai *xoo woye* *xoo*: 1.ª pessoa do singular *woye*: determinante possessivo da 2.ª pessoa do singular tradução literal: pai teu = o teu pai *"tate woye"*	(5) o vosso pai *tate weni* tate: 1.ª pessoa do singular *weni*: determinante possessivo da 3.ª pessoa do plural tradução literal: pai vosso = o vosso pai *"tate weni"*
(3) O seu pai/o pai dele *tate woye/xee woye* *tate*: 1.ª pessoa do singular *weni*: determinante possessivo da 2.ª pessoa do/ *xee*: 1.ª pessoa do singular *woye*: determinante possessivo da 2.ª pessoa do singular tradução literal: Pai seu/pai dele = o seu pai/o pai dele *"tate woye/xee woye"*	(6) os seus pais/os pais deles *Ootate vavo/Ooxe vavo* *ootate*: prefixo nominal de classe + 2.ª pessoa do plural *vavo*: determinante possessivo da 3.ª pessoa do plural *Ooxe*: prefixo nominal de classe + 2.ª pessoa do plural *vavo*: determinante possessivo da 3.ª pessoa do plural tradução literal: Pais seus/pais deles = os seus pais/os pais deles *"Ootate vavo/Ooxe vavo"*

Figura 5

Ortografia Desatualizada



Nota. Fonte: Dicionário Português-Oshikwanyama (Leconte, 1949).

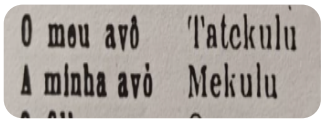
Tabela 2

Proposta Nossa

teu pai = xoo woye
o pai dele = xee waye
os seus pais = ooxe vavo

Figura 6

Equivalente ou Tradução Desviante



Nota. Fonte: Dicionário Português-Oshikwanyama (Leconte, 1949).

Tabela 3

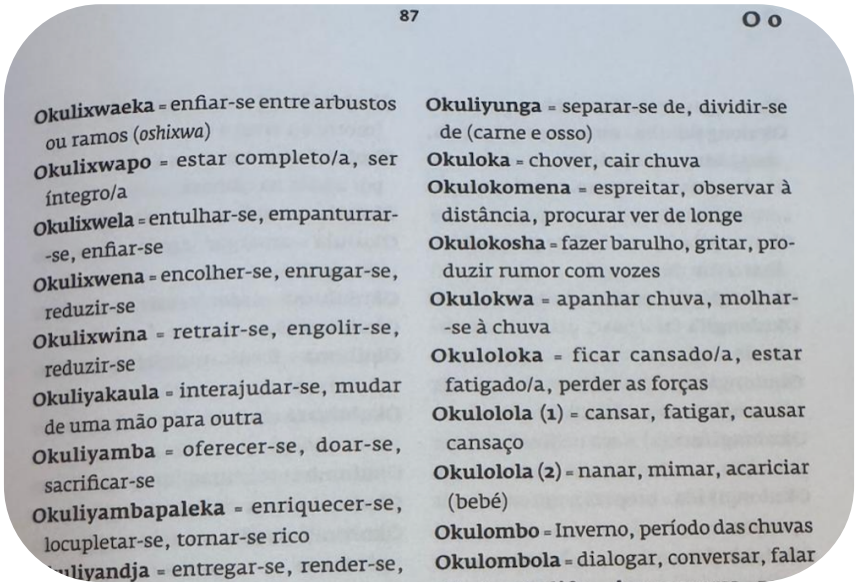
Proposta Nossa

avô materno ou paterno = Tatekulu wange
avó materna ou paterna = Mekulu wange

Tal como dissemos, acima, que os dicionários das línguas africanas tendem a apresentar informações básicas aos utilizadores, também constatamos que os dicionários do oshikwanyama em que recai o nosso estudo apresentam algumas lacunas relacionadas com a falta de informações. Por exemplo, os artigos lexicográficos presentes nas Figuras 7 e 8, retirados de páginas selecionadas de forma aleatória dos dicionários de Lidinikeni (2017) e de Leconte (1949), respetivamente, ilustram claramente a falta de informações úteis para os consulentes.

Figura 7

Entradas do Dicionário de Oshikwanyama-Português



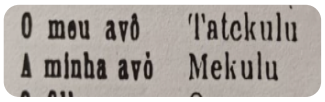
87 Oo

Okulixwaeka = enfiar-se entre arbustos ou ramos (*oshixwa*)
Okulixwapo = estar completo/a, ser íntegro/a
Okulixwela = entulhar-se, empanturrar-se, enfiar-se
Okulixwena = encolher-se, enrugar-se, reduzir-se
Okulixwina = retrair-se, engolir-se, reduzir-se
Okuliyakaula = interajudar-se, mudar de uma mão para outra
Okuliyamba = oferecer-se, doar-se, sacrificar-se
Okuliyambapaleka = enriquecer-se, locupletar-se, tornar-se rico
Okuliyandja = entregar-se, render-se,
Okuliyunga = separar-se de, dividir-se de (carne e osso)
Okuloka = chover, cair chuva
Okulokomena = espreitar, observar à distância, procurar ver de longe
Okulokosha = fazer barulho, gritar, produzir rumor com vozes
Okulokwa = apanhar chuva, molhar-se à chuva
Okulololoka = ficar cansado/a, estar fatigado/a, perder as forças
Okulololola (1) = cansar, fatigar, causar cansaço
Okulololola (2) = nanar, mimar, acariciar (bebé)
Okulombo = Inverno, período das chuvas
Okulombola = dialogar, conversar, falar

Nota. Fonte: *Pequeno Dicionário Oshikwanyama Português* (Lindinikeni, 2017, p. 87).

Figura 8

Entradas do Dicionário Português-Oshikwanyama: Pequeno Método de Aprender Português para o Uso dos Povos do Cuanhama.



O meu avô T'atekulu
A minha avó Mekulu

Nota. Fonte: Dicionário Português-Oshikwanyama (Leconte, 1949, p. 4).

Estes dicionários apresentam informações muito básicas para os consulentes. Nas entradas apresentadas, não se inclui:

- transcrição fonética para utilização secundária (professores);
- informação sobre flexão nominal e verbal;
- definições;
- informação sobre variação ortográfica, considerando que o oshikwanyama pode ter grafias diferentes em Angola e na Namíbia;

- relações entre as palavras, como por exemplo, relações derivacionais e/ou famílias de palavras (*sapato*, *sapateiro*, *sapataria* etc.);
- exemplos de uso;
- paráfrases ou imagens para os casos em que não há equivalente na língua de chegada; e
- sistematização das informações disponibilizadas.

Há, ainda, que notar que os dicionários bilingues que são usados pelos professores neste nível de escolaridade não foram desenvolvidos para o ensino primário. São instrumentos lexicográficos bilingues de língua geral que são usados como auxiliares pelos professores do ensino primário numa tentativa de transmitir conhecimentos e garantir a comunicação na sala de aula.

4.3. **Estratégias de utilização do dicionário em sala de aula**

De acordo Ferreira e Vieira (2013), as atividades produtivas de aprendizagem do léxico em sala de aula, que contribuem para a compreensão, para a produção textual e para o aperfeiçoamento de competências linguístico-discursivas, podem depender da aplicação de diversas estratégias de utilização de dicionários bilingues (Ferreira & Viera, 2013, p. 32).

O processo de ensino e aprendizagem requer a combinação de métodos e técnicas. Para tal, o professor deve aplicar em sala de aulas estratégias que lhe permitam alcançar os objetivos de ensino e do perfil de saída dos seus alunos.

Neste estudo, analisamos 10 vídeos de aulas gravadas em diferentes escolas do ensino primário angolano na província do Cunene, município de Namacunde. Por intermédio das aulas gravadas, verificámos que os professores do ensino primário de zonas recônditas, nas escolas de Onamaimbi, Oikundu, Eeyoka, Ounonge, Enyombwi, Ohaipeto, Omuti-wa-Nangala, Okayolwa, Ofenda e Okuandjaomudi da província angolana do Cunene, para amenizar as dificuldades de leitura, escrita e comunicação dos seus alunos, têm recorrido a diferentes estratégias de utilização de dicionários bilingues oshikwanyama-português/português-oshikwanyama, ainda que tais instrumentos lexicográficos não se encontrem contemplados no sistema de ensino angolano. A análise das aulas permitiu-nos observar que a utilização de tais dicionários se baseiam essencialmente em duas estratégias principais: i) mediação, com recurso a materiais lexicográficos, e ii) produção, pelo professor falante nativo de oshikwanyama, de conhecimento equivalente em oshikwanyama, com solução de lacunas lexicais e/ou conceptuais.

4.3.1. **Mediação**

Esta estratégia, que recorre a materiais lexicográficos, aplica-se tipicamente ao ensino da disciplina de língua portuguesa, que cobre léxico e conhecimento geral e não especializado.

Os professores falantes nativos do oshikwanyama servem como mediadores entre a língua de ensino (o português) e a língua nativa dos alunos, o oshikwanyama. Os exemplos (8) e (9), que se seguem, ilustram os recursos convocados neste caso:

Figura 9

Entradas do Dicionário de Lidinikeni (2017)

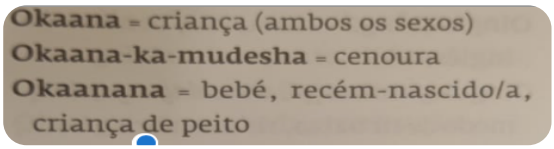
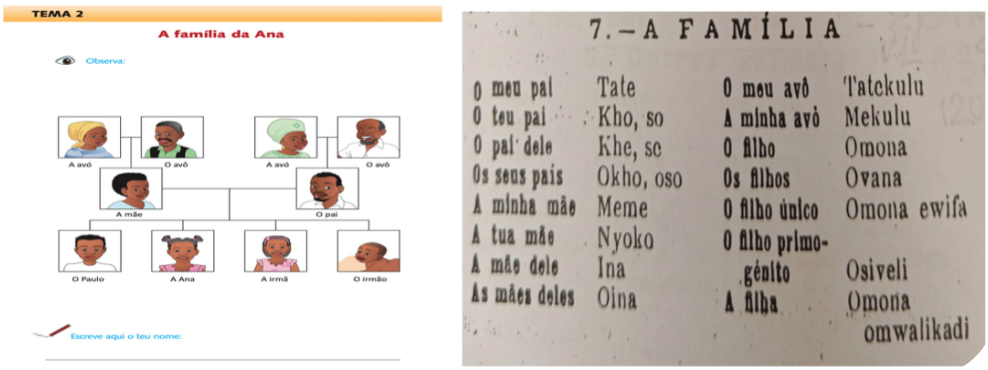


Figura 10

Manual de Ensino e Entradas Extraídos do Dicionário de Leconte (1949)



Nota. Fontes: Ministério da Educação de Angola (2021f, p. 16), Dicionário Português-Oshikwanyama (Leconte, 1949).

Considerando os recursos usados no ensino da disciplina de língua portuguesa, sobre o tema família, ilustrados em (8) e (9) acima, os professores falantes nativos são chamados a mediar as aprendizagens, quer do léxico dicionarizado, quer do léxico nos materiais de ensino que não se encontra nos dicionários existentes, recorrendo aos seus conhecimentos de português e à sua língua materna, partilhada com os alunos nas escolas rurais, para colmatar as lacunas. Sem esta mediação, e considerando que os conteúdos da disciplina de português e o ensino da língua portuguesa em Angola são lecionados como se tratasse de ensino de L1, é óbvio que os alunos têm dificuldades de comunicação e de aprendizagem dos conteúdos lecionados na língua portuguesa.

Noutra situação, estão os professores falantes não nativos ou não falantes de oshikwanyama que recorrem necessariamente aos manuais e aos recursos lexicográficos do português-oshikwanyama/oshikwanyama-português disponíveis – Leconte (1949) e Lidinikeni (2017) – para transmitir conhecimentos, manter a comunicação na sala de aulas, contando para isso também com o auxílio intuitivo dos seus alunos, que identificam os sons das palavras usadas como equivalentes da língua materna mal pronunciadas pelos seus professores durante o ensino.

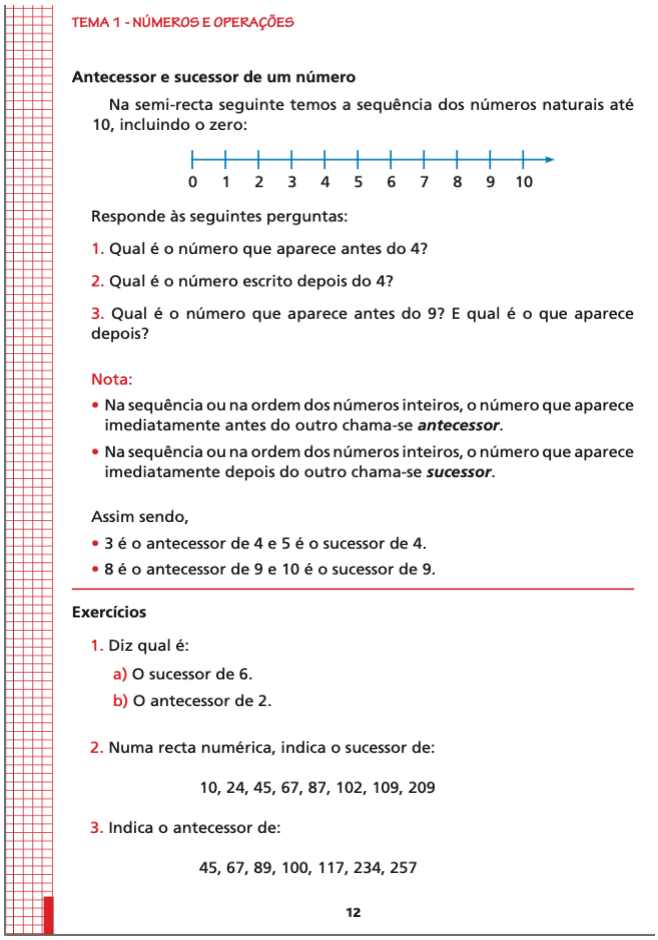
Apesar de os professores falantes não nativos e os não falantes de oshikwanyama conseguirem ensinar a língua portuguesa recorrendo aos dicionários bilingues existentes, os níveis de dificuldade de aprendizagem dos alunos aumentam com o aparecimento de novos termos técnicos, científicos e especializados de cada uma das disciplinas, na medida em que há ausência de determinadas palavras e dos seus equivalentes nos dicionários usados. Muitas vezes, para ultrapassar estas dificuldades, os professores falantes não nativos e os não falantes de oshikwanyama recorrem também ao uso de imagens que estão nos livros como equivalentes, estratégia eficaz para questões menos complexas, como as ilustradas na Figura 10.

4.3.2. **Produção**

A produção e os empréstimos lexicais são estratégias usadas pelos professores do ensino primário nas zonas recônditas e aplicam-se tipicamente no ensino de outras matérias ou disciplinas, focando-se na transmissão de conhecimentos com base nas produções dos professores, e não em recursos lexicográficos disponíveis. Os professores falantes nativos exploram os temas no quadro e, novamente, servem eles próprios como intermediários entre o português e o oshikwanyama. O exemplo do ensino de matemática (Figura 11), e as lacunas identificadas nos recursos em uso – dicionário onde as palavras em português que compõem o nome do tema, número e operação, deveriam surgir – (Figura 12) ilustram esta situação.

Figura 11

Tema "Números e operações" Matemática, 2.ª classe



TEMA 1 - NÚMEROS E OPERAÇÕES

Antecessor e sucessor de um número

Na semi-recta seguinte temos a sequência dos números naturais até 10, incluindo o zero:

0 1 2 3 4 5 6 7 8 9 10

Responde às seguintes perguntas:

1. Qual é o número que aparece antes do 4?
2. Qual é o número escrito depois do 4?
3. Qual é o número que aparece antes do 9? E qual é o que aparece depois?

Nota:

- Na sequência ou na ordem dos números inteiros, o número que aparece imediatamente antes do outro chama-se ***antecessor***.
- Na sequência ou na ordem dos números inteiros, o número que aparece imediatamente depois do outro chama-se ***sucessor***.

Assim sendo,

- 3 é o antecessor de 4 e 5 é o sucessor de 4.
- 8 é o antecessor de 9 e 10 é o sucessor de 9.

Exercícios

1. Diz qual é:
   a) O sucessor de 6.
   b) O antecessor de 2.
2. Numa recta numérica, indica o sucessor de:

   10, 24, 45, 67, 87, 102, 109, 209
3. Indica o antecessor de:

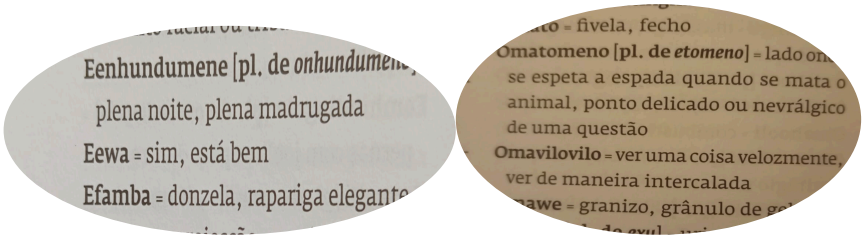
   45, 67, 89, 100, 117, 234, 257

12

Nota. Fonte: Manual de Matemática, 2.ª classe (Ministério da Educação de Angola, 2021i, p. 12).

Figura 12

Imagem de páginas do dicionário de Lidinikeni (2017)



Nota. Fonte: *Pequeno Dicionário Oshikwanyama Português* (Lindinikeni, 2017, pp. 26, 124).

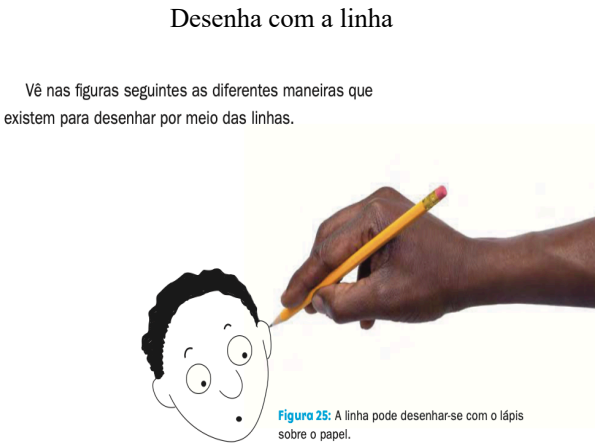
Para o caso do exemplo da Figura 12, no qual as palavras usadas no manual do ensino primário não têm equivalentes nos dicionários disponíveis, os professores falantes nativos de oshikwanyama atuam como mediadores, usando os equivalentes quando os conceitos estão lexicalizados, ou produzindo paráfrases explicativas ou recorrendo a empréstimo de léxico do português, quando não há léxico disponível na língua portuguesa. O professor falante nativo de oshikwanyama não tem problema em descodificar o tema para os alunos, valendo-se do conhecimento da sua língua materna, o oshikwanyama, e usando os equivalentes das palavras números e operações em oshikwanyama, *Eenomola* = números e *ilo Mavalulo* = operações/contagem, respetivamente.

Para a mesma situação, verificou-se que os professores falantes não nativos tentaram transmitir os conhecimentos forçando a adoção de léxico do português, contando com o conhecimento que os alunos têm sobre a sua língua materna e com a sua intuição e capacidade de dedução e aquisição de uma língua segunda/estrangeira.

No entanto, por intermédio dos vídeos não é fácil avaliar de que forma é que os professores não falantes de oshikwanyama conseguem efetivamente transmitir conhecimentos a crianças falantes nativas de oshikwanyama, principalmente quando estas crianças não conseguem comunicar na língua de ensino, como por exemplo aquelas crianças que ingressam pela primeira vez no ensino primário nas escolas do meio rural da província do Cunene. Constatamos, ainda, que os professores não falantes de oshikwanyama das 3.ª, 4.ª e 5.ª classes, ao abordarem temas mais complexos nos diversos domínios deste subsistema de ensino, têm mais dificuldades de articular ou estabelecer a intermediação entre o português e a língua materna dos alunos, devido à complexidade conceptual dos domínios e palavras em causa, o que aumenta a sua dificuldade em encontrar equivalentes em oshikwanyama. Neste contexto, os alunos também não conseguem comunicar facilmente com os professores visto que os conceitos a que se referem as palavras podem não ser conhecidos dos alunos e não fazer parte do seu sistema ou universo de referência. São as palavras técnico-cientificas de alguns domínios como matemática, ciências da natureza ou geografia que criam mais dificuldades. Como se ilustra na Figura 13 e Figura 14, verifica-se que as palavras encontradas no dicionário são diferentes das palavras quando em uso. Por esta razão, o professor falante não nativo ou não falante de oshikwanyama tem ainda mais dificuldades do que o professor falante nativo em recorrer aos dicionários para o ensino de outras matérias.

Figura 13

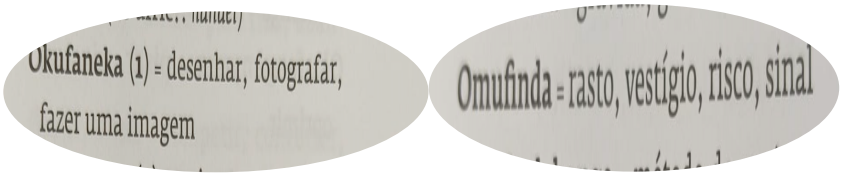
Atividade "Desenha com a linha", Educação Manual e Plástica, 2.ª classe.



Nota. Fonte: Manual de Educação Manual e Plástica, 2.ª classe (Ministério da Educação de Angola, 2021b, p. 14).

Figura 14

Entradas do dicionário de Lidinikeni (2017)



Nota. Fonte: *Pequeno Dicionário Oshikwanyama Português* (Lindinikeni, 2017, pp. 63–130).

O professor, depois de localizar os equivalentes na língua oshikwanyama no dicionário, conforme ilustrado no exemplo da Figura 14 – onde constam as entrads para as palavras em português que compõem o tema – , terá de perceber que as palavras dicionarizadas podem não corresponder às palavras em uso. De modo a transmitir o conhecimento de forma clara e sistemática, ele terá de reconstruir em oshikwanyama as frases sintaticamente corretas e compreensíveis para os alunos, combinando preposições e suprimindo prefixos de classe nominal ou verbal do oshikwanyama. No exemplo analisado acima, é evidente que a tradução literal poderia levar a interpretações equivocadas, como por exemplo: *fotografa o sinal*, que veicula o significado errado e leva ao desvio do conhecimento que se pretendia transmitir.

Além disso, o professor precisa de ter a competência para produzir as expressões adequadas para que o significado seja corretamente transmitido. Para o exemplo apresentado na Figura 13, os equivalentes adequados seriam *okufaneka* e *omufinda* e a expressão gramatical resultante *faneka nomufinda* (≅ desenhar com a linha).

Este exemplo ilustra ainda as lacunas de informação no dicionário quanto aos elementos flexionais ou preposicionais que se aglutinam sob a forma de prefixos:

i. Para se obter o equivalente de *desenha* em oshikwanyama há supressão do prefixo **oku-** do verbo *okufaneka* que se encontra dicionarizado para obter *faneka* (desenha), e

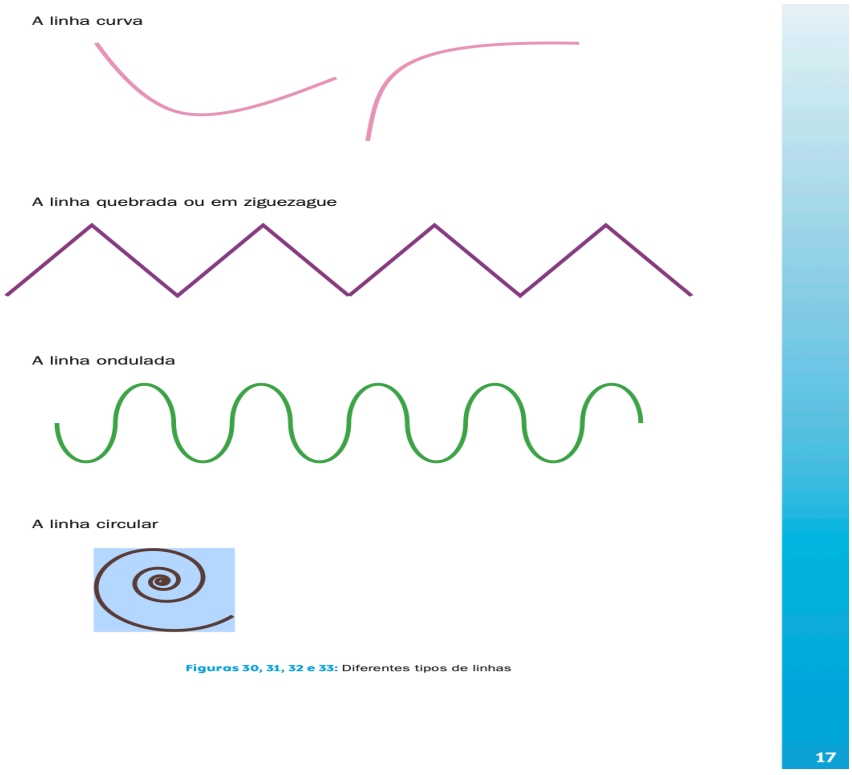
ii. há acréscimo de prefixos, como neste caso em que há acréscimo do prefixo **n-** (= *com*/*a*) ao substantivo *omufinda* (=linha) para formar "com a linha".

Como nem sempre há equivalência total, os professores de ensino primário recorrem a paráfrases, que reforçam com o auxílio de imagens. O exemplo (3) e a Figura 15 ilustram esta combinação de estratégia de ensino dos professores.

(3) Mediação do professor, com recurso a paráfrase:
faneka nomufinda + *Ngashi tashulikwa memafano talandulako*
Desenha *com as linhas* + *conforme está nas figuras que se seguem*

Figura 15

Atividade "Desenha com a linha" (≠ okufabeka omufinda*)*

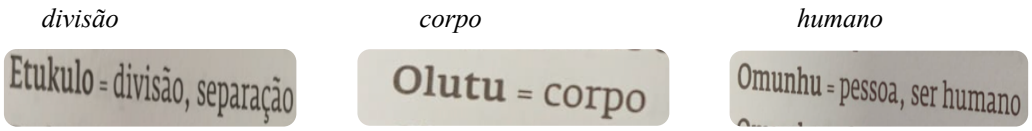


Nota. Fonte: Manual de Educação Manual e Plástica, 2.ª classe (Ministério da Educação de Angola, 2021b, p. 17).

Constatou-se também que os professores não falantes de oshikwnayama têm dificuldades em transmitir conhecimentos aos alunos durante a realização de atividades básicas em todas as disciplinas, como por exemplo na disciplina de Estudo do Meio. O exemplo da Figura 16 procura ilustrar a situação.

Figura 16

Tema "A Divisão do Corpo Humano", Estudo do Meio, 3.ª Classe



Nota. Fonte: Dicionário de Lidinikeni (2017, pp. 41–141).

O caso da Figura 15 ilustra como nem todas as formas usadas no ensino primário em Angola podem ser encontradas nos dicionários existentes de português-oshikwanyama/oshikwanyama-português, ou como há casos em que as palavras estão nos dicionários, mas estão dicionarizadas de forma diferente das formas em uso. Esta diferença entre as palavras dicionarizadas e as palavras em uso constitui um grande problema para os professores não falantes de oshikwanyama, porque tais palavras precisam de partículas de ligação, tipicamente sob a forma de prefixos, caraterísticos do oshikwanyama e das demais línguas bantu, que operam como artigos ou como preposições.

Se o professor tivesse de fazer uma tradução literal, é evidente que poderia chegar a uma forma errada. Para chegar a uma forma gramatical, o professor precisa de adequar os elementos linguísticos do oshikwanyama. O exemplo (4) ilustra outro destes casos, com os elementos não indicados no dicionário a negrito:

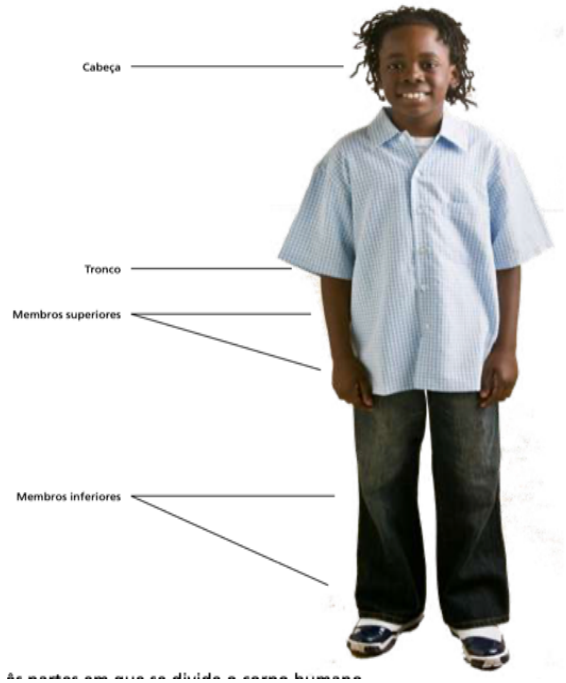
(5) Tradução para oshikwanyama do tema A Divisão do Corpo Humano, Estudo do Meio, 3.ª classe:

divisão	corpo	humano
etukulo	*olutu*	*omunhu*
etukulo ***lo****lutu* ***lo****munhu*	≅ divisão do corpo humano	

O professor recorre ainda ao auxílio de imagens, como as ilustradas na Figura 6, abaixo, que os alunos ajudam a descodificar recorrendo ao seu conhecimento implícito da língua.

Figura 17

Imagem do Manual de Estudo do Meio, 3.ª Classe



Nota. Fonte: Manual de Estudo do Meio, 3.ª classe (Ministério da Educação de Angola, 2021e, p. 7).

Os casos aqui ilustrados demonstram de forma simples como quanto mais técnico-científicas são as palavras, menor é a probabilidade de serem encontradas nos dicionários existentes - de caráter geral e com grandes fragilidades. O exemplo da Figura 11 demonstra ainda que as palavras técnico-científicas e não dicionarizadas oferecem maior grau de complexidade, em particular para os professores do ensino primário não falantes de oshikwanyama. Para contribuir para a solução deste problema, estamos a projetar um instrumento lexicográfico bilingue pedagógico para o ensino básico de Angola, com base no léxico do português usado no ensino primário angolano.

5. Considerações finais

Este trabalho apresenta uma caraterização do panorama atual dos recursos lexicográficos bilingues oshikwanyama-português/português-oshikwanyama, analisados na perspectiva da sua utilização em sala de aula no contexto do ensino multilingue nas províncias rurais do Cunene, bem como a identificação de estratégias de resolução de problemas de comunicação com recurso à utilização de dicionários bilingues em sala de aula, incluindo algumas limitações e desafios. Pretende-se que esta análise possa contribuir com informações úteis para o desenho de um novo recurso lexicográfico que responda às necessidades específicas deste(s) público(s)-alvo, com vista à contribuição para a melhoria da qualidade do ensino primário em Angola que, a par dos desafios que se colocam ao ensino em ambientes multilingues, segue ainda uma metodologia de ensino baseada

na tradução de e para a língua portuguesa, sem uma verdadeira integração de ensino bilingue com materiais pedagógico-científicos nas várias línguas ou com a distinção entre ensino de português L1 e L2.

Na província do Cunene, os professores do ensino primário utilizam ainda o *Dicionário Oshikwanyama-Português* do Padre Ernesto Leconte, de 1949, e o *Pequeno Dicionário Oshikwanyama-Português*, da autoria do Padre Virgínio Lidinikeni, de 2017. Estes dicionários, embora pouco atualizados e muito incompletos, têm-se, ainda assim, demonstrado úteis na transmissão de conhecimentos e na promoção da comunicação em sala de aula. No entanto, são dicionários muito pouco atentos às necessidades específicas do contexto educacional em que são usados, quer em termos de cobertura, quer em termos da informação dada.

O estudo desses dicionários e das estratégias usadas pelos professores revela a relevância e a importância de investir em novos recursos lexicográficos com objetivos pedagógicos com vista à melhoria da qualidade educacional, facilitando a aprendizagem da língua portuguesa e de outras matérias lecionadas em português, e contribuindo para o sucesso escolar dos alunos. A opção pelo investimento num instrumento lexicográfico deste tipo, ao invés de glossários específicos para o ensino básico, justifica-se pelo facto de não haver ainda um dicionário mais completo que promova um verdadeiro multilinguismo no ensino em Angola e que possa, também, funcionar como instrumento de normalização, ajudando a lidar com os desafios do multilinguismo nas políticas linguísticas nesta região de Angola. Um dicionário mais abrangente e consistente, que cubra os conceitos em uso no ensino, mas que contemple também a diversidade de significados associados a cada forma e que sistematize informação gramatical que ajude a compreender as diferentes estratégias das duas línguas, potenciará uma utilização mais abrangente que pode contemplar a comunidade educativa – professores e alunos –, mas também encarregados de educação, famílias e restante comunidade.

O investimento em recursos pedagógicos que respeitam a diversidade linguística existente e a natureza multilingue do contexto angolano promove uma educação mais inclusiva e eficaz. Enquanto não há reforço de recursos linguísticos e terminológicos das várias línguas angolanas, com vista a um equilíbrio entre elas, este investimento deveria ainda ser acompanhado da criação de um instrumento jurídico-legal que estabelecesse requisitos linguísticos para a comunidade escolar, obrigando o professor de escolas rurais a ser proficiente na língua da sua comunidade escolar, de modo a permitir que este haja como mediador entre a língua de ensino e a língua da comunidade escolar.

Agradecimentos / Financiamento

Agradecemos muito e sinceramente a disponibilidade e os comentários e sugestões dos revisores deste artigo.

Financiamento

Parte deste trabalho foi financiado por fundos nacionais através da FCT – Fundação para a Ciência e Tecnologia, I.P., no âmbito do projeto UID/03213 – Centro de Linguística da Universidade NOVA de Lisboa (CLUNL).

Referências

Andrade, E. D. (2007). *Línguas africanas. Breve introdução à fonologia e à morfologia*. Santos-FLUL/CLUL.

Bergenholtz, H., & Gouws, R. H. (2012). What is lexicography. *Lexikos*, *22*(1), 31–42. https://doi.org/10.5788/22-1-996

Carvalho, O. L. (2001). *Lexicografia bilngue português/alemão: Teoria e aplicação à categoria das preposições.* Thesaurus.

Correia, M. (2009). *Os dicionários portugueses.* Caminho.

Dubois, J., & Dubois, C. (1971). *Introduction à la lexigrophie le dictionnaire.* Librairie Larousse.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2024). Oshiwambo language (KUA). In *Ethnologue: Languages of the world* (27th ed.) [versão online]. SIL International. https://www.ethnologue.com/language/kua/

Fernandes, J., & Ntondo, Z. (2002). *Angola: Povos e línguas*. Nzila.

Ferreira, H. M., & Vieira, M. S. (2013). O trabalho com o léxico em sala de aula: Desafios para o ensino de língua materna. *Revista Letras Raras*, *2*(2), 19–33.

Grandi, L. D., & Nadin, O. L. (2020). O dicionário em sala de aula: Orientações para a formação lexicográfica de professores de línguas à luz da lexicografia pedagógica. *Forum Linguístico*, *17*(3), 5054–5072. https://doi.org/10.5007/1984-8412.2020.71084

Guebe, A. (2008). *Resistência à ocupação colonial do Sul de Angola - Região dos Va-Nyaneka-va-Nkumbi e dos Va-Ambo (1850-1917).* Arte Viva.

Guthrie, M. (1967). *The classification of the Bantu languages*. International African Institute.

Haipinge, T. E. (2017). *Antroponínia na língua Oshikwanyama* [Dissertação de mestrado, FCSH-Nova de Lisboa]. RUN – Repositório da Universidade Nova de Lisboa. http://hdl.handle.net/10362/22256

Halme, R. (2004). *A tonal grammar of Kwanyama*. Rüdiger Köppe Verlag.

Halme, R. (2006). *Encontro sobre a questão do alfabeto e ortografia deoshikwanyama*. Ondjiva.

Hartmann, R. R., & James, G. (1998). *Dictionary of lexicography*. Routledge.

Holnselaar, W. (2003). Examples of design and production criterio for major dictionaries. In O. P. Sterkenburg, *A practical guide to lexicography* (pp. 323–332). John Benjamins Company.

Instituto Nacional de Estatística. (Ed.). (2016). Resultados definitivos do recenseamento geral da população e da habitação de Angola 2014. INE.

Kirkness, A. (2004). Lexicography. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 54–81). Blackwell. https://doi.org/10.1002/9780470757000.ch2

Krienger, M. (2011). Programas de educação no Brasil e dicionários para a escola fundamental. In O. S. Cardoso, S. Mejri, & J. A. Mota (Eds.), *Os dicionários: Fontes, métodos e novas tecnologias* (pp. 145–187). Instituto de Letras da Universidade Federal da Bahia.

Lara, L. F. (2004). O dicionário e suas disciplinas . In O. A. Isquerdo & M. D. Krienger (Eds.), *As ciências do léxico: Laxicologia, lexicografia e terminologia* (pp. 133–184). UFMS.

Leconte, P. E. (1949). *Pequeno método de aprender português para o uso dos povos do Cuanhama (Sul de Angola)*. Missão da Huíla.

Lidinikeni, V. (2017). *Onyika Yetu yOshikwanyama nOshiputu "Pequeno dicionário Oshikwanyama Português"*. Papelmunde.

Ministério da Educação de Angola. (s.d.). *Manuais escolares*. Xilonga. https://xilonga.med.gov.ao/manuais-escolares

Ministério da Educação de Angola. (2021a). *Educação manual e plástica, 1.ª classe* [Manual escolar]. Popular.

Ministério da Educação de Angola. (2021b). *Eduação manual e plástica, 2.ª classe* [Manual escolar]. Popular.

Ministério da Educação de Angola. (2021c). *Educação musical, 1.ª & 2.ª classes* [Manual escolar]. Moderna.

Ministério da Educação de Angola. (2021d). *Estudo do meio, 2.ª classe* [Manual escolar]. Editora das Letras.

Ministério da Educação de Angola. (2021e). *Estudo do meio, 3.ª classe* [Manual escolar]. Progresso.

Ministério da Educação de Angola. (2021f). *Língua portuguesa, 1.ª classe* [Manual escolar]. Popular.

Ministério da Educação de Angola. (2021g). *Língua portuguesa, 2.ª classe* [Manual escolar]. Popular.

Ministério da Educação de Angola. (2021h). *Matemática, 1.ª classe* [Manual escolar]. Moderna.

Ministério da Educação de Angola. (2021i). *Matemática, 2.ª classe* [Manual escolar]. Mayama Editora.

Ndombele, E. D. (2017). Reflexão sobre as línguas nacionais no sistema de educação em Angola. *RILP - Revista Internacional em Língua Portuguesa*, *31*, 71–89. https://doi.org/10.31492/2184-2043.RILP2017.31/pp.71-89

Ndombele, E. D., & Timbane, A. A. (2020). O ensino de língua portuguesa em Angola: Reflexões metodológicas em contexto multilíngue. *Fólio - Revista de Letras*, *12*(1), 289–314. https://doi.org/10.22481/folio.v12i1.6604

Ngunga, A., & Simbine, M. C. (2012). *Gramática descritiva da língua Changana*. As Nossa Línguas e Centro de Estudos Africanos (CEA) - UEM.

Petter, M. (Org.). (2015). *Introdução à linguística africana*. Contexto, pp. 49–85.

Prinsloo, D. (2015). Analyzing words as a social enterprise: Lexicography in Africa with specific reference to South Africa. In J. Miller (Ed.), *Analysing words as a social enterprise: Celebrating 40 years of the 1975 Helsinki Declaration on lexicography*. *Collected Papers from AustraLex 2015* (pp. 42–59). AustraLex.

Prinsloo, D. J. (2017). Africa's response to the corpus revolution. In H. Xu (Ed.), *Proceedings of the 11th International Conference of the Asian Association for Lexicography, Asialex 2017, 10–12 June, Guangzhou, China* (pp. 20–31). AsiaLex

Prinsloo, D., & Zondi, N. (2020). Lexicographic treatment of zero equivalence in isiZulu dictionaries. *Lexikos 30*, 346–362. https://doi.org/10.5788/30-1-1605

Scantamburlo, L. (2002). *Dicionário do guineense-português: Disionariu guinensi-purtuguis*. V. II, FASPEBI.

Severo, C. G., Sassuco, D. P., & Bernardo, E. P. (2019). Português e línguas Bantu na educação angolana: Da diversidade como "problema". *Línguas Instrumentos Linguísticos*, *43*, 290–307. https://doi.org/10.20396/lil.v0i43.8658374

Shikesho, E. (2020). Towards a common Oshikwanyama official orthography for Namibia and Angola: The way forward on the conjunctive and disjunctive writing systems of Oshikwanyama. *JULACE*: *Journal of University of Namibia Language Centre*, *5*, 1–11.

Tobias, G. W., & Turvey, B. H. (1965). *English - Kwanyama dictionary*. Witwatersrand University Press.

Tonjes, H. (1910). *Worterbuch Der Ovambo - Sprache*. Druck UND Verlag Von Georg Reimer.

Verdelho, T., & Silvestre, J. P. (2011). *Lexicografia bilingue: A tradição dicionarística português-línguas* modernas. Centro de Línguas da Universidade de Lisboa/Universidade de Aveiro.

Vilela, M. (1979). *Problemas da lexicologia e lexicografia*. Livraria Civilização Editora.

Wheeler, D., & Pelissier, R. (2013). *História de Angola*. Tinta-da-China.

Zimmermann, O. W., Taapopi, G. B., & Hasheela, P. (2019). *Oshikwanyama-English, English-Oshikwanyama dictionary*. Macmillan Education Namibia.

Montes de sentidos: Uma abordagem multidimensional às expressões metafóricas

Heaps of meanings: A multidimensional approach to metaphorical expressions

Inês Mateus[1], *Helena Mendes Oliveira*[1], *José Teixeira*[2], *Ana Paula Soares*[1]

[1] Laboratório de Cognição Humana, CIPsi, Escola de Psicologia, Universidade do Minho
[2] Departamento de Estudos Portugueses e Lusófonos, Escola de Letras, Artes e Ciências Humanas, Universidade do Minho

Abstract

According to the Conceptual Metaphor Theory (CMT), conceptual metaphors (CMs) are defined as mechanisms that structure and map abstract concepts through concrete notions, thus represented in language through metaphorical expressions (e.g., the CM *happiness is up* is present in everyday discourse in expressions such as "feeling on top of the world"). Despite this theory's relevance, recent research has highlighted the necessity to include other components in order to deepen the understanding of CMs, namely by incorporating discursive, sociocultural, and empirical approaches into CM investigation. Building on these components, this study seeks to contribute to outstanding key questions on CM research, particularly with regard to CM classification. Such classification may allow for a clearer understanding of how CMs are structured, expressed, and processed, thereby ensuring greater stimulus control in future studies. Thus, this research is based on the collection of over two hundred metaphorical expressions representative of multiple CMs in both English and Portuguese. Participants evaluated these expressions on a five-point Likert scale in terms of familiarity, concreteness, valence, activation, embodiment, and transparency, variables previously identified as influencing metaphor processing. The preliminary results presented in this study, focusing exclusively on the Portuguese sample, reveal interactions that support the view of CMs as encompassing different components that transcend language, and provide a synthesised database of metaphorical expressions which will enable the development of controlled stimuli for future research in this area.

Keywords: Conceptual metaphors, metaphoric processing, subjective dimensions.

Resumo

De acordo com a Teoria das Metáforas Concetuais (TMC), metáforas concetuais (MC) definem-se como mecanismos que estruturam e mapeiam conceitos abstratos através de noções concretas, estando presentes na linguagem através de expressões metafóricas (e.g., a MC *felicidade é em cima* está presente no discurso diário através de expressões metafóricas como "estar em alta"). Apesar da pertinência desta teoria, a investigação recente salienta a necessidade de outras componentes no aprofundamento do entendimento das MC, nomeadamente incluindo abordagens discursivas, socioculturais e empíricas na investigação. Assim, alicerçando-se nas componentes destacadas, este estudo procura contribuir para questões-chave por responder, nomeadamente no que diz respeito à classificação de MC. Esta classificação pode possibilitar um melhor entendimento sobre o modo como diferentes MC são estruturadas, expressas e processadas, garantindo assim um maior controlo estimular em estudos futuros. Assim sendo, este estudo baseia-se no levantamento de mais de duas centenas de expressões metafóricas representativas de múltiplas MC tanto em inglês como em português. Os participantes avaliaram estas expressões através de uma escala de Likert de cinco pontos em termos de familiaridade, concretude, valência, ativação, corporalidade e transparência, variáveis previamente identificadas como afetando o processamento metafórico. Os resultados preliminares apresentados neste estudo,

prendendo-se exclusivamente com a amostra portuguesa, revelam interações que contribuem para a visão das MC como abrangentes de diferentes componentes que transcendem a linguagem, sintetizando também uma base de dados de expressões metafóricas que permitirá a criação de estímulos controlados para futura investigação nesta área.

Palavras-Chave: Metáforas conceptuais, processamento metafórico, dimensões subjetivas.

1. Introdução

A teoria das metáforas concetuais (TMC; *Conceptual Metaphor Theory*) apresentada por George Lakoff e Mark Johnson em 1980 define metáforas concetuais (MC) como mecanismos utilizados para caracterizarmos noções abstratas através de conceitos concretos (Lakoff & Johnson, 1980). Neste sentido, as MC - grafadas em formato de A é B, e onde A corresponde ao conceito mais abstrato (domínio alvo), que é caracterizado através de B (i.e., domínio fonte e conceito mais concreto) - revelam o sistema concetual através do qual noções abstratas são entendidas pelos falantes de uma língua (e.g., na MC *culpa é peso*, o domínio alvo de culpa é concetualizado através de características concretas pertencentes ao domínio fonte, o peso). Enquanto representações esquemáticas de um padrão de pensamento, as MC dão origem a expressões metafóricas presentes na linguagem diária (e.g., a MC *culpa é peso* refletida em expressões metafóricas como "consciência pesada" e "um peso nos ombros") e estruturam o modo como a linguagem abstrata é processada, podendo também influenciar o nosso comportamento. Deste modo, com a introdução desta teoria, as metáforas saem da esfera exclusivamente poética (onde existiam enquanto metáforas literárias) e passam a ser entendidas como MC que constituem uma parte integral do discurso diário através de expressões metafóricas, assegurando a expressão e a compreensão de conceitos abstratos.

Desde o surgimento da TMC, vários autores têm contribuído para o desenvolvimento e expansão desta teoria, realçando a natureza interdependente da linguagem metafórica enquanto fenómeno que se baseia numa pluralidade discursiva, social e cultural (Gibbs, 2017; Soares da Silva, 2021; Steen, 2011). Neste sentido, Fauconnier e Turner (2003) apresentam a teoria da integração concetual (*conceptual blending*) segundo a qual a utilização de MC proporciona a criação e integração de novos espaços mentais que podem existir exclusivamente na nossa mente ou ter uma representação física. Como exemplo ilustrativo desta teoria, os autores realçam a interface de um ambiente de trabalho num computador. Este "ambiente" integra elementos informáticos e noções básicas de um local de trabalho, dando assim origem a um novo espaço onde num ecrã virtual os "ficheiros" são organizados em "pastas" e colocados na "reciclagem" quando descartados, tal como aconteceria no espaço físico de um escritório. Neste sentido, de acordo com os autores desta teoria, a utilização de linguagem metafórica não só possibilita a conceptualização de noções abstratas através de conceitos concretos, como também permite a criação de novos espaços integrados de pensamento (*blended spaces*).

Adicionalmente, Kövecses (2020) destaca a relevância de uma dimensão sociocultural e contextual na compreensão e utilização da MC, argumentando que, apesar de várias MC existirem em diversas línguas, é possível que o papel sociocultural e contextual de uma determinada comunidade afete a compreensão e utilização da mesma MC. Além disso, o autor destaca também o contexto discursivo que acompanha uma MC como tendo um papel fundamental na interpretação da mesma, uma vez que a mesma MC pode obter sentidos distintos dependendo da situação comunicativa, do género textual e mesmo do momento histórico no qual é aplicada. Por exemplo, a MC *o governo é uma máquina* pode ser compreendida como sinónima de um governo "célere" ou "eficiente" ou como representativa de um governo "inflexível" e "impessoal", dependendo do contexto cultural e histórico da sua utilização. Assim sendo, destacando a importância de diferentes dimensões na utilização de linguagem metafórica, Kövecses realça a natureza heterogénea das MC, mesmo quando presentes em diferentes contextos.

Contudo, apesar do desenvolvimento e utilização de MC em diversas áreas de investigação, subsistem ainda questões-chave por responder, nomeadamente no que diz respeito às características das MC e suas

possíveis classificações. A falta de consenso nesta classificação constitui uma lacuna fundamental no entendimento das diferentes vertentes da linguagem metafórica, podendo contribuir para a utilização de estímulos não controlados em estudos que utilizam duas ou mais MC ou que utilizam a mesma MC em diferentes línguas. Neste sentido, ao longo dos anos foram apresentadas várias propostas de classificação de MC derivadas de uma observação teórica da linguagem (e.g., a divisão tripartida de MC ontológicas, de orientação e estruturais proposta por Lakoff e Johnson em 1980; a teoria de MC primárias e não-primárias de proposta por Grady em 1997), sendo que estas têm sido consideradas desatualizadas uma vez que vários autores defendem a necessidade de uma abordagem empírica para caracterizar a variedade de MC.

Assim, Katz e colaboradores (1988) propuseram a criação de um sistema pioneiro no qual diferentes MC seriam avaliadas em várias dimensões previamente definidas como influentes no uso e no processamento de linguagem metafórica. Assim, ao contrário das propostas que procuravam classificar MC em categorias discretas (e.g., Grady, 1997; Lakoff & Johnson, 1980), esta abordagem dimensional entende que as MC podem ser classificadas a partir de dimensões contínuas baseadas em avaliações subjetivas. Deste modo, os autores compilaram uma lista de 464 metáforas literárias (existentes em trabalhos literários, nomeadamente *sunlight is golden dust* [*a luz do sol é poeira dourada*]) e não literárias (e.g., *children are sponges,* [*as crianças são esponjas*]) em formato A é B, avaliadas por 634 participantes em dez dimensões: *compreensibilidade* (facilidade de entendimento da MC), *facilidade de interpretação* (clareza de significado da MC), *metaforicidade* (grau de verdade figurativa transmitida pela MC), *qualidade da metáfora* (nível de adequação, pertinência e apelo da MC), *imagética da metáfora* (grau de evocação de imagens mentais e/ou sons na MC), *imagética do sujeito* (avaliação da imagética do sujeito da MC), *imagética do predicado* (avaliação da imagética do predicado da MC), *familiaridade percebida* (nível de familiaridade das ideias expressas na MC), *relação semântica* (avaliação do grau de relação ou semelhança entre os dois elementos da MC) e *número de interpretações alternativas* (variedade de significados possíveis transmitidos pela MC). De acordo com os autores, estas dimensões foram adotadas dada a sua pertinência, destacada em estudos prévios de linguagem metafórica (Honeck & Hoffman, 1980; Katz et al., 1985; Marschark et al., 1983). A avaliação das MC foi realizada através de uma escala de Likert de 7 pontos. Como resultado, os autores compilaram uma base de dados extensa na qual MC são avaliadas em várias dimensões (apresentando os valores de médias e desvios-padrão para cada MC em cada dimensão), que tem sido amplamente utilizada em investigações subsequentes para garantir um maior controlo estimular (Brugman et al., 2022; Ichien et al., 2024; Jones & Estes, 2006; Schmidt & Seger, 2009).

Adicionalmente, Campbell e Raney (2016) replicaram o estudo de Katz e colaboradores (1988) utilizando uma amostra diferente de participantes. De acordo com os autores, a necessidade desta réplica prendia-se primeiramente com a evolução natural da linguagem, particularmente no que toca à interpretação de linguagem figurativa. Assim, os autores procuraram testar se as mesmas MC seriam avaliadas de forma semelhante vinte e cinco anos após o primeiro estudo, argumentando que dimensões como a familiaridade podiam sofrer alterações justificadas pela passagem do tempo. Em segundo lugar, destacaram a importância de recolher também o historial linguístico dos participantes, focando-se particularmente na influência de uma segunda língua na interpretação de linguagem metafórica. De facto, embora o estudo desenvolvido por Katz et al. (1988) tenha sido realizado no Canadá, onde o inglês e o francês constituem línguas oficiais, os autores não recolheram informações linguísticas dos participantes. Procurando colmatar esta lacuna, Campbell e Raney (2016) incluíram apenas participantes que tinham frequentado escolas de língua inglesa durante um período mínimo de 10 anos, requerendo também um historial linguístico dos participantes. Os resultados foram analisados como uma comparação geral entre as classificações obtidas e as classificações de Katz et al. (1988) e considerando as diferenças individuais dos indivíduos (com os participantes separados de acordo com o seu historial linguístico). No entanto, ambas as análises apresentaram fortes semelhanças com as classificações obtidas anteriormente, tendo os autores concluído pela robustez e fiabilidade da base de dados criada por Katz et al. (1988).

Paralelamente, em contexto lusófono, também tem sido explorada a influência de dimensões psicolinguísticas em expressões metafóricas, quando comparadas com expressões literais. Por exemplo,

Siqueira et al. (2010) compilaram uma lista de frases que incluíam expressões metafóricas e expressões literais em português do Brasil. Recorrendo a duas tarefas de avaliação, os autores pediram aos participantes que avaliassem cada item nas dimensões subjetivas de *familiaridade* (frequência do item num determinado contexto), *alerta* (grau de ativação provocado pelo item no participante), *valência* (qualidade da resposta emocional provocada pelo item no participante) e *convencionalidade* (o quão estabelecida é uma determinada MC no contexto linguístico do participante). A avaliação foi realizada através de uma escala de Likert de cinco pontos e os resultados apresentam uma lista de materiais rigorosamente avaliados em dimensões pertinentes que podem ser utilizados em investigação futura.

Apesar da pertinência dos estudos referidos, é possível argumentar que a sua abordagem monolíngue (e focada maioritariamente na língua inglesa) constitui uma lacuna na avaliação empírica de MC, uma vez que, como destacado anteriormente, a influência de fatores culturais contribui e afeta o processamento da linguagem metafórica. Neste sentido, e dado que na investigação experimental de MC frequentemente se adotam abordagens multilíngues (Kövecses, 2012; Mateus et al., 2025), é importante que os estudos normativos de avaliação de MC reflitam esta diversidade linguística, para garantir um maior controlo estimular comum a várias línguas. Adicionalmente, os estudos normativos de avaliação de MC, mais predominantes na língua inglesa, focam-se na recolha de avaliações subjetivas de MC no seu formato formal de A é B, o que não reflete o modo como as MC são representadas na linguagem quotidiana. Por exemplo, ainda que a MC *poder é em cima* seja bastante comum até em diferentes línguas (e.g., português, inglês, espanhol, francês, etc.) esta tende a não estar refletida na linguagem desta forma, mas sim através de várias expressões metafóricas que evidenciam a sua existência (e.g., "sua alteza real" ou "os seus subordinados"). Assim sendo, é possível afirmar que a apresentação e avaliação de MC apenas no seu formato formal pode constituir uma limitação, uma vez que esta não abrange a existência de expressões metafóricas comuns no dia a dia.

A fim de colmatar as lacunas anteriormente referidas, no presente estudo recolheram-se avaliações normativas de 213 expressões metafóricas (refletindo 20 MC em português europeu e em inglês britânico) em várias dimensões consideradas pertinentes no processamento da linguagem, não só metafórica como também emocional. Assim, através da dimensão subjetiva da familiaridade procurou-se avaliar o grau de familiarização dos participantes com a expressão metafórica, destacando se esta expressão era utilizada pelos participantes e se estava presente na sua linguagem diária, conforme avaliada em estudos anteriores prendendo-se com linguagem abstrata (Al-Azary & Katz, 2020; Jamrozik et al., 2016). Em seguida, a concretude, destacada enquanto dimensão essencial para o processamento de linguagem metafórica (Aono et al., 2024; Reijnierse et al., 2019), foi avaliada enquanto dimensão que procurava perceber se os participantes experienciavam a expressão metafórica através de um dos cinco sentidos (olfato, tato, audição, paladar e visão). Adicionalmente, a valência procurava testar a natureza (positiva ou negativa) do impacto da expressão metafórica, sendo que, de acordo com vários autores, esta dimensão, bem como a ativação, constituem fatores pertinentes para a avaliação da linguagem abstrata e emocional (Soares et al., 2011, 2017). Neste sentido, a dimensão da ativação procurava perceber o impacto da expressão metafórica no nível de ativação dos participantes, enquanto a corporalidade era avaliada como dimensão que procurava captar o grau de envolvimento do corpo na expressão metafórica (Khatin-Zadeh et al., 2023; Santana & Vega, 2011). Por fim, a dimensão subjetiva da transparência testava a facilidade de compreensão da expressão metafórica apenas através dos seus constituintes, como aliás testado anteriormente (Littlemore, 2004).

Deste modo, o presente estudo tem por objetivo a criação de uma base de dados que contribua para a avaliação normativa de MC de formas inovadoras através: i) da criação daquela que é, tanto quanto sabemos, a primeira base de dados de linguagem metafórica desenvolvida em português europeu; ii) da abordagem interlinguística que prevê também a recolha de avaliações em inglês britânico; e iii) da proposta de inclusão de expressões metafóricas na avaliação de MC, substituindo a habitual abordagem em formato A é B. Por fim, apesar dos objetivos deste estudo contemplarem a criação de uma base de dados bilíngue (tendo para isso selecionado expressões metafóricas existentes em português europeu e inglês britânico), destaca-se que os resultados apresentados se restringem aos dados preliminares recolhidos em português europeu.

2. Método

2.1. Participantes

Foram registadas 239 participações através da plataforma de atribuição de créditos da Escola de Psicologia da Universidade do Minho e da rede de contactos dos investigadores. Este número corresponde a um total de 205 participantes (176 mulheres; M_{idade} = 22.5, *DP* = 6.99), uma vez que 34 deles participaram duas vezes no estudo, avaliando, em cada uma delas, expressões metafóricas distintas, como verificado através do seu número de identificação. A participação foi voluntária, e, no caso dos estudantes da Escola de Psicologia, foram atribuídos créditos pela participação. Todos os participantes tinham visão e audição normal ou corrigida. Esta amostra considerou apenas falantes de nacionalidade e língua nativa portuguesas, uma vez que se pretendia evitar que fatores culturais e linguísticos pudessem afetar as avaliações. Todos os participantes forneceram o seu consentimento digital para participar e o estudo foi aprovado pela Comissão de Ética da Universidade do Minho (CEICSH 087/2024).

2.2. Materiais

Foram selecionadas 213 expressões metafóricas usadas comumente no português europeu. Estas expressões foram selecionadas utilizando motores de busca online, como o Google e o ChatGPT (OpenAI, GPT-4). De assinalar que essa seleção considerou também que as mesmas pudessem ocorrer em língua inglesa, uma vez que este estudo faz parte de um projeto de investigação mais alargado onde se pretende analisar diferenças interlinguísticas na avaliação de expressões metafóricas equivalentes.

Especificamente, a seleção das 213 expressões metafóricas obedeceu aos seguintes critérios de inclusão: (i) expressões ilustrativas de uma ou várias MC comuns no universo português e inglês - expressões metafóricas ilustrativas da MC *saudade é uma entidade*, como "alimentar a saudade", foram excluídas, dado serem específicas do contexto português; (ii) expressões que refletissem a MC em ambas as línguas (e.g, a expressão "this contribution outweighs the argument", ilustrativa da MC *considering is weighing* não foi incluída uma vez que a sua tradução para português - i.e., "esta contribuição supera o argumento" - não reflete a MC de *considerar é pesar;* (iii) expressões traduzíveis em ambas as línguas (e.g., "shoulder the guilt" reflete a MC de *culpa é peso* em inglês; no entanto o verbo "shoulder" não apresenta tradução comum em português, pelo que esta expressão metafórica não foi incluída).

Assim sendo, as 213 expressões metafóricas selecionadas são ilustrativas de vinte MC existentes nas duas línguas, sendo que este número resulta da classificação a posteriori das expressões identificadas como pertinentes para o estudo. De facto, e sendo o principal objetivo a recolha de expressões metafóricas que corroboravam com os critérios acima referidos, a atribuição de cada expressão metafórica a uma MC específica foi elaborada posteriormente e apenas para efeitos de organização da base de dados. Assim, estas expressões metafóricas pertencem às MC de *afeto é calor* (affection is warmth), *alcançar um objetivo é chegar a um destino* (reaching a goal is arriving at a destination), *compreender é ver* (understanding is seeing), *considerar é pesar* (considering is weighing), *culpa é peso* (guilt is weight), *dificuldade é peso (*difficulty is weight), *discussões são guerras* (arguments are wars), *existência é visibilidade* (existence is visibility), *felicidade é em cima* (happiness is up), *ideias são alimento* (ideas are food), *importância é peso* (importance is weight), *mais é em cima* (more is up), *o bem/moralidade é em cima* (good/morality is up), *o tempo é um recurso* (time is a resource), *poder é em cima* (power is up), *quantidade é tamanho* (quantity is size), *segredos são pesos* (secrets are weights), *uma crença é uma posição/orientação física* (a belief is a position/physical orientation), *uma situação é um local* (a situation is a place), a *vida é uma viagem* (life is a journey).

Com o objetivo de assegurar alguma uniformidade nas expressões metafóricas a incluir na base de dados, optou-se por manter as expressões metafóricas no infinitivo sempre que possível (e.g., "subir na posição social") e, nos casos em que a atribuição de género gramatical era necessária, optou-se pela forma masculina (e.g., "esmagado pela culpa"). Além disso, para garantir um maior controlo estimular e evitar a exaustão dos

participantes, as 213 expressões metafóricas foram aleatoriamente divididas por sete listas em cada idioma, cada uma contendo entre 29 e 31 expressões metafóricas, que os participantes avaliavam nas seis dimensões subjetivas. A distribuição de participantes por lista foi feita aleatoriamente pela plataforma de recolha de dados. Optou-se ainda por recolher os dados das seis dimensões separadamente para evitar possíveis vieses que poderiam surgir da avaliação simultânea de diferentes dimensões, sendo que a ordem de aparecimento das expressões em cada dimensão foi aleatorizada. Por fim, a ordem das dimensões foi fixada (i.e., familiaridade, concretude, valência, ativação, corporalidade e transparência) para garantir que dimensões que podiam ser tidas como mais semelhantes não surgissem uma a seguir à outra (nomeadamente, a familiaridade e a transparência, sendo que ambas dizem respeito ao significado e ao uso das expressões metafóricas; e a concretude e a corporalidade, que envolvem a utilização do corpo para a compreensão da expressão metafórica). A lista total de expressões metafóricas em português europeu e respetivas MC disponibiliza-se no Anexo 1.

2.3. **Procedimento**

As avaliações foram recolhidas através de um procedimento online implementado no software Qualtrics (Qualtrics, 2024). O recurso a procedimentos online na recolha de dados tem-se tornado cada vez mais comum na investigação, pois permite a recolha rápida e segura de uma grande quantidade de dados com igual fiabilidade (e.g., Balota et al., 2001; Brysbaert et al., 2013; Kuperman, 2013; Soares et al., 2011, 2017). A tarefa foi divulgada na plataforma de obtenção de créditos da Escola de Psicologia da Universidade do Minho e através da rede de contactos pessoais dos investigadores.

A participação iniciava-se assim que os participantes clicavam na hiperligação disponibilizada na plataforma da Escola de Psicologia ou através de contacto por e-mail ou WhatsApp (no caso dos contactos pessoais dos investigadores), sendo primeiramente apresentadas as instruções gerais, que informavam os participantes acerca do propósito do estudo. Os participantes eram, ainda, esclarecidos sobre a confidencialidade dos dados recolhidos e encorajados a responder de forma rápida e instintiva. Após darem o seu consentimento para participar no estudo, procedia-se à recolha de alguma informação de caracterização sociodemográfica e do historial linguístico dos participantes (e.g., idade, sexo biológico, género, línguas nativas e domínio de outras línguas). Em seguida, apresentavam-se as instruções da primeira dimensão a ser avaliada (i.e., familiaridade), cuja descrição permanecia no ecrã até que os participantes pressionassem o botão de avanço. As instruções apresentadas para cada dimensão foram baseadas em estudos anteriores sobre avaliação lexical e MC (Littlemore, 2001; Soares et al., 2011). Como na maioria destes estudos, os participantes avaliaram cada expressão metafórica utilizando uma escala de Likert, tendo-se neste caso optado por uma escala de cinco pontos por se considerar suficientemente discriminativa para a presente tarefa. Assim, a dimensão de familiaridade foi avaliada com base na frequência com que os participantes encontravam e utilizavam a expressão no seu quotidiano, sendo que a escala disponível variava entre 1 (muito desconhecida) e 5 (muito familiar), com os valores intermédios de 2 (um pouco desconhecida), 3 (nem familiar nem desconhecida) e 4 (um pouco familiar). A concretude foi avaliada considerando se a expressão poderia ser experienciada através dos cinco sentidos (olfato, visão, tato, audição e paladar). A escala variava entre 1 (muito abstrata) e 5 (muito concreta) com os valores intermédios distribuídos de forma semelhante à escala de familiaridade (i.e., 2 [um pouco abstrata], 3 [nem abstrata nem concreta] e 4 [um pouco concreta]). A valência foi definida como o impacto emocional causado pela expressão, sendo que o participante deveria avaliar se cada expressão o fazia sentir-se "muito triste" (1) ou "muito feliz" (5), com os valores intermédios representando "um pouco triste" (2), "nem triste nem feliz" (3) e "um pouco feliz (4). A ativação dizia respeito à sensação de calma ou ativação transmitida por cada expressão, utilizando uma escala de 1 (muito calmo/a) a 5 (muito ativado/a) com os valores intermédios de "um pouco calmo/a" (2), "nem calmo/a nem ativado/a" (3) e "um pouco ativado/a" (4) também disponíveis. Na corporalidade avaliava-se o grau de envolvimento do corpo humano para compreender uma expressão, sendo que os participantes deveriam avaliar se a expressão era "muito incorpórea" (1), "um pouco incorpórea" (2), "nem incorpórea nem corpórea" (3), "um pouco corpórea (4) ou "muito corpórea" (5). Por fim, a transparência

baseava-se na clareza do significado da expressão apenas com base nas palavras utilizadas e os participantes deveriam indicar se a expressão era 1 (muito opaca) ou 5 (muito transparente) com valores intermédios de "um pouco opaca" (2), "nem opaca nem transparente" (3) e "um pouco transparente" (4) também acessíveis. Para cada dimensão foram também apresentadas duas palavras ou expressões idiomáticas como referência para os extremos da escala, tendo por base dados previamente recolhidos em estudos anteriores (Brysbaert et al., 2013; Soares et al., 2011; Titone & Connine, 1994). Assim, o conjunto completo de instruções utilizadas para cada dimensão subjetiva na versão portuguesa da tarefa pode ser consultado no Anexo 2.

As expressões metafóricas surgiam isoladamente, dispostas no centro do ecrã e acompanhadas por uma escala de Likert de cinco pontos, disponibilizando-se também a opção "não conheço a expressão", caso os participantes desconhecessem a expressão metafórica. A apresentação das expressões foi realizada de forma aleatória em cada dimensão a ser avaliada. Cada expressão permanecia no ecrã até que os participantes emitissem uma resposta selecionando com o cursor o número da escala que refletia a sua avaliação (de 1 a 5) e pressionando o botão de avanço para o aparecimento de uma nova expressão. Neste sentido, os participantes eram obrigados a avaliar todas as expressões antes de avançarem para a dimensão seguinte, não sendo permitida a retificação de uma avaliação previamente submetida. Após a avaliação de todas as expressões metafóricas numa dada dimensão, os participantes recebiam instruções sobre a próxima dimensão a ser avaliada. Em seguida, as expressões eram novamente apresentadas para que fossem avaliadas nessa dimensão. Este processo repetia-se até que todas as expressões (29 a 31, dependendo da lista) fossem avaliadas. O procedimento total demorava em média 15 a 25 minutos por participante.

3. Resultados

Os resultados preliminares (número de avaliações, médias e desvios-padrão) do estudo normativo de validação das 213 expressões metafóricas do português europeu nas dimensões de familiaridade, concretude, valência, ativação, corporalidade e transparência foram calculados depois de realizarmos várias inspeções aos dados recolhidos, de forma a garantir a integridade dos mesmos. Assim, excluíram-se participações que atribuíam a mesma classificação às diferentes expressões metafóricas de uma dimensão (à exceção da dimensão de familiaridade, uma vez que, conforme disposto em seguida, as expressões foram frequentemente avaliadas como muito familiares), revelando desatenção ou falta de interesse na realização da tarefa (24 participações). Excluíram-se também participações com comportamentos identificados como desviantes, nomeadamente a repetição de um padrão de resposta ao longo de várias expressões metafóricas (e.g., avaliações repetitivas em padrões de 4, 5, 4, 5) e avaliações notoriamente desatentas (e.g., classificar uma expressão como "ser espezinhado" como "muito alegre"), perfazendo um total de quatro participações.

Excluíram-se também, para cada dimensão, as avaliações de participantes que indicavam que a expressão metafórica lhes era desconhecida. Estes casos foram muito raros, levando à exclusão de 4% de respostas para familiaridade e 3% em cada uma das restantes dimensões. Em seguida, foram calculadas as médias e desvios-padrão com base nos dados disponíveis para cada expressão em cada dimensão (perfazendo um total de N = 6835 avaliações para familiaridade, N = 6697 para concretude, N = 6470 na valência, N = 6219 para a ativação, N = 6090 para a corporalidade e N = 6006 para a transparência). De seguida, valores 2,5 *DP* abaixo ou acima da média obtida para cada expressão metafórica em cada uma das dimensões avaliadas foram eliminados, no sentido de evitar que valores mais extremados (*outliers*) pudessem distorcer os resultados e comprometer a validade das análises. Esta etapa permitiu garantir uma maior robustez estatística, assegurando que os dados refletiam, de forma mais fidedigna, a tendência geral das avaliações, ainda que o número de *outliers* identificados em cada dimensão tenha sido bastante reduzido - 2.59 % na familiaridade, 0.21% na concretude, 1.61% na valência, 1.08% para a ativação, 0.36% para a corporalidade e 1.23% na transparência.

Assim, os valores normativos apresentados basearam-se em 6658 avaliações válidas para a dimensão de familiaridade, sendo o número médio de respostas válidas por expressão metafórica de 31.3 (amplitude: 21 – 36). A média global de familiaridade foi de 4.22 (*DP* = 0.61), tendo sido várias as expressões avaliadas como

muito familiares, como aliás se demonstra na Tabela 1, que apresenta os valores descritivos (incluindo o tamanho da amostra, as médias, desvios-padrão e amplitudes) para as expressões avaliadas como mais desconhecidas e mais familiares. Como podemos observar, as expressões mais desconhecidas (i.e., “os falcões voam mais do que os pardelhos", "trajetória moral ascendente" e “ética em queda livre”) tratam ambas de noções de verticalidade (cima-baixo), ainda que para concetualizar conceitos abstratos diferentes (refletindo a MC de *mais é em* cima na expressão metafórica “os falcões voam mais do que os pardelhos" e *moralidade é em cima* nas restantes).

Tabela 1

Número de Avaliações (N), Médias (M), Desvios-Padrão (DP) e Valores de Amplitude (Mínimo–Máximo) para as Expressões Avaliadas na Dimensão Subjetiva de Familiaridade como “Mais Desconhecidas” e “Mais Familiares”.

Classificação	Expressão metafórica	*N*	*M*	*DP*	Mín.–Máx.
Muito desconhecida	“os falcões voam mais do que os pardelhos"	21	1.57	0.75	1-3
	"trajetória moral ascendente"	30	2.33	1.27	1-5
	“ética em queda livre”	21	2.62	1.36	1-5
	[…]				
	“perder tempo”	29	5.00	0.00	5
	“expectativas elevadas”	28	5.00	0.00	5
	"sentir-se nas nuvens”	31	5.00	0.00	5
	"trabalho pesado”	32	5.00	0.00	5
	“ponto de vista”	33	5.00	0.00	5
	“subir ao poder"	30	5.00	0.00	5
	“uma conversa profunda”	27	5.00	0.00	5
	"voar mais alto"	30	5.00	0.00	5
Muito familiar	"alcançar os meus objetivos"	32	5.00	0.00	5

Em seguida, os valores obtidos para o índice subjetivo de concretude para as expressões avaliadas como mais abstratas e mais concretas são apresentados na Tabela 2. Esta dimensão registou uma média global de 3.36 (*DP* = 0.54), baseando-se em 6683 avaliações válidas, com uma média de 31.4 avaliações válidas por expressão metafórica (amplitude: 23 – 35). Como podemos observar, expressões metafóricas como “ética em queda livre” e “descer ao vazio” foram avaliadas como as mais abstratas, sendo que também registaram baixos valores de

familiaridade (M = 2.62; DP = 1.36 e M = 2.93; DP = 1.46, respetivamente), enquanto as expressões mais concretas (i.e., "esta música apareceu pela primeira vez" e "uma grande multidão") registaram elevados valores de familiaridade (M = 4.18; DP = 0.94; e M = 4.87; DP = 0.35, respetivamente).

Tabela 2

Número de Avaliações (N), Médias (M), Desvios-Padrão (DP) e Valores de Amplitude (Mínimo–Máximo) para as Expressões Avaliadas na Dimensão Subjetiva de Concretude como "Mais Abstratas" e "Mais Concretas".

Classificação	Expressão metafórica	*N*	*M*	*DP*	Mín.–Máx.
Muito abstrata	"ética em queda livre"	25	1.72	0.89	1-4
	"descer ao vazio"	29	1.79	1.01	1-4
	"o peso da importância"	31	2.29	1.44	1-5
	[…]				
	"preços altos"	31	4.58	0.81	2-5
	"esta música apareceu pela primeira vez"	29	4.59	0.73	3-5
Muito concreta	"uma grande multidão"	30	4.77	0.50	3-5

Já no que se refere à dimensão da valência, obtivemos 6366 avaliações válidas, com uma média de 29.9 expressões válidas por expressão (amplitude: 18 – 33) e uma média global de 2.99 (DP = 1.14). A Tabela 3 apresenta os valores descritivos para as expressões avaliadas como mais tristes e mais alegres. Como é de notar, as expressões metafóricas destacadas como mais alegres ("alcançar o sucesso", "alcançar os meus objetivos" e "atingir a minha meta") ilustram a mesma MC (*atingir um objetivo é chegar a um destino*), tendo todas registado também elevados níveis de familiaridade M = 4.88; DP = 0.35, M = 5.00; DP = 0.00 e M = 4.71; DP = 0.46, respetivamente) e concretude (M = 3.83; DP = 1.34, M = 3.37; DP = 1.59 e M = 3.94; DP = 1.29, respetivamente).

Tabela 3

Número de Avaliações (N), Médias (M), Desvios-Padrão (DP) e Valores de Amplitude (Mínimo–Máximo) para as Expressões Avaliadas na Dimensão Subjetiva de Valência como "Mais Tristes" e "Mais Alegres".

Classificação	Expressão metafórica	*N*	*M*	*DP*	Mín.–Máx.
Muito triste					
▲	"chegar ao fundo do poço"	28	1.00	0.00	1
	"cair em desgraça"	28	1.14	0.36	1-2
	"Estar carregado de culpa"	28	1.18	0.48	1-3
	[…]				
	"Alcançar o sucesso"	29	4,97	0.19	4-5
	"alcançar os meus objetivos"	32	4.97	0.18	4-5
▼ **Muito alegre**	"atingir a minha meta"	29	5.00	0.00	5

A dimensão subjetiva de ativação registou uma média total de 3.58 (*DP* = 0.61) baseando-se em 6152 avaliações válidas e uma média de 28.9 avaliações válidas por expressão (amplitude: 17 – 33). A Tabela 4 apresenta os valores descritivos para as expressões avaliadas como mais calmas e mais ativadoras. Pertinentemente, as expressões avaliadas como mais calmas ("mais leve depois de fazer a coisa certa", "tempo poupado" e "no caminho certo na vida") foram também avaliadas como muito alegres na dimensão subjetiva de valência (*M* = 4.65; *DP* = 0.49, *M* = 4.03; *DP* = 0.85 e *M* = 4.72; *DP* = 0.46, respetivamente) enquanto as expressões mais ativadoras ("emocionalmente difícil de suportar", "lançar um ataque verbal" e "pressão esmagadora") foram avaliadas como muito tristes na mesma dimensão (*M* = 1.23; *DP* = 0.43, *M* = 1.31; *DP* = 0.47; *M* = 1.30; *DP* = 0.54, respetivamente).

Tabela 4

Número de Avaliações (N), Médias (M), Desvios-Padrão (DP) e Valores de Amplitude (Mínimo–Máximo) para as Expressões Avaliadas na Dimensão Subjetiva de Ativação como "Mais Calmas" e "Mais Ativadoras".

Classificação	Expressão metafórica	*N*	*M*	*DP*	Mín.–Máx.
Muito calmo/a	"mais leve depois de fazer a coisa certa"	30	1.87	1.33	1-5
	"tempo poupado"	33	2.15	1.33	1-4
	"no caminho certo na vida"	30	2.3	1.44	1-5
	[…]				
	"emocionalmente difícil de suportar"	29	4.72	0.45	4-5
	"lançar um ataque verbal"	26	4.73	0.45	4-5
Muito ativado/a	"pressão esmagadora"	26	4.73	0.45	4-5

Relativamente aos valores obtidos para a dimensão subjetiva de corporalidade, registamos uma média total de 2.96 (*DP* = 0.56) baseando-se num total de 6068 avaliações válidas e numa média de 28.5 avaliações válidas por expressão (amplitude: 18 – 33). A Tabela 5 contém os valores descritivos, incluindo o tamanho da amostra, as médias, desvios-padrão e amplitudes (máximas e mínimas) para o índice subjetivo de corporalidade das expressões avaliadas como mais incorpóreas e mais corpóreas. Como é possível verificar, as expressões avaliadas como mais incorpóreas prendem-se com o universo monetário e moral (i.e., "as *ações* estão a cair" , "o *valor* de uma hora" e "declínio *moral*"), sendo as expressões mais corpóreas relacionadas com a experiência física de sensações, nomeadamente da culpa e do trabalho através de peso (i.e., "um olhar pesado de culpa" e "trabalho pesado") e do afeto enquanto calor (i.e., "abraço caloroso").

Tabela 5

Número de Avaliações (N), Médias (M), Desvios-Padrão (DP) e Valores de Amplitude (Mínimo–Máximo) para as Expressões Avaliadas na Dimensão Subjetiva de Corporalidade como "Mais Incorpóreas" e "Mais Corpóreas".

Classificação	Expressão metafórica	*N*	*M*	*DP*	Mín.–Máx.
Muito incorpórea					
	"as ações estão a cair"	25	1.56	0.87	1-4
	"o valor de uma hora"	21	1.62	0.86	1-3
	"declínio moral"	31	1.80	0.98	1-4
	[…]				
	"trabalho pesado"	29	4.34	0.61	3-5
	"um olhar pesado de culpa"	25	4.36	0.70	3-5
Muito corpórea	"abraço caloroso"	25	4.48	0.96	2-5

Por último, os valores obtidos para a dimensão subjetiva de transparência para as expressões avaliadas como mais opacas e mais transparentes são apresentados na Tabela 6. Esta dimensão registou uma média total de 3.74 (*DP* = 0.51) baseando-se num total de 5932 avaliações válidas e numa média de 27.8 avaliações válidas por expressão (amplitude: 15 – 32). Como se pode constatar, mesmo as expressões avaliadas como mais opacas ("os falcões voam mais do que os pardelhos", "de mente elevada" e "entrevistas de peso") apresentam uma avaliação média superior a 2, refletindo que mesmo nestes casos há alguma facilidade em compreender o significado da linguagem metafórica. Ainda assim, a expressão avaliada como mais opaca (i.e., "os falcões voam mais do que os pardelhos", ilustrativa da MC *mais é em cima*) foi também avaliada como a menos familiar desta base de dados (conforme se pode constatar na Tabela 1).

Tabela 6

Número de Avaliações (N), Médias (M), Desvios-Padrão (DP) e Valores de Amplitude (Mínimo–Máximo) para as Expressões Avaliadas na Dimensão Subjetiva de Transparência como "Mais Opacas" e "Mais Transparentes".

Classificação	Expressão metafórica	*N*	*M*	*DP*	Mín.–Máx.
Muito opaca	"os falcões voam mais do que os pardelhos"	15	2.33	0.72	1-3
	"de mente elevada"	22	2.45	1.18	1-5
	"entrevistas de peso"	27	2.48	1.89	1-5
	[…]				
	"um enorme sucesso"	30	4.7	0.47	4-5
	"o calor de um sorriso"	29	4.7	0.45	4-5
Muito transparente	"abraço caloroso"	24	4.75	0.44	4-5

4. Discussão

Os resultados deste estudo, ainda que amplamente preliminares, permitem-nos observar vários aspetos dignos de relevo. Primeiramente, é interessante verificar que, em geral, as expressões são avaliadas como muito familiares, sendo que nove expressões metafóricas foram avaliadas como extremamente familiares por todos os participantes, atestando a existência destas expressões no universo linguístico e cultural dos participantes. Além disso, quanto às expressões avaliadas como menos familiares, nota-se que estas contêm termos que podem ser considerados menos comuns (nomeadamente, "trajetória", "pardelhos" e até a colocação "ética em queda"), o que poderá ter influenciado estas avaliações. Neste sentido, para estudos futuros, seria pertinente recolher também avaliações de familiaridade ao nível dos componentes lexicais que integram as expressões, de modo a poder comparar estas classificações.

Em seguida, a sobreposição entre expressões mais abstratas enquanto mais desconhecidas e expressões mais concretas como mais familiaridade poderá ser um indicador de que expressões mais tangíveis e percetíveis fisicamente são mais facilmente internalizadas e, portanto, reconhecidas. Neste sentido, também se revela uma relação entre as expressões destacadas como mais felizes como mais familiares. Este facto pode dever-se à maior frequência de uso de expressões emocionalmente mais positivas num contexto social, sendo por isso mais frequentes (ou familiares) no dia a dia. Já a dimensão da ativação demonstra também uma coerência com a valência, que parece sugerir, ainda que com dados preliminares e meramente descritivos, uma relação entre a intensidade emocional (mais calma ou mais ativadora) e a polaridade afetiva (mais triste ou mais feliz).

Finalmente, na dimensão da corporalidade é notório que as expressões classificadas como mais corpóreas se prendem com conceitos físicos e sensoriais (nomeadamente de peso e calor), enquanto as expressões destacadas como mais incorpóreas tratam conceitos intangíveis, como por exemplo o tempo e a moralidade. Esta observação aponta para uma relação entre a linguagem metafórica e a utilização da experiência física do mundo para exprimir conceitos abstratos que corrobora o impacto do contexto corpóreo e social nas MC destacado por vários autores (Kövecses, 2020; Littlemore, 2019). Por fim, na dimensão da transparência

observa-se que mesmo as expressões classificadas como mais opacas apresentam valores de média que se situam acima do segundo ponto da escala e que a expressão classificada como mais opaca é também a mais desconhecida na dimensão da familiaridade, evidenciando que a clareza de interpretação pode associar-se ao contacto com as expressões no dia a dia.

5. Conclusão

No presente estudo apresentam-se os dados preliminares de uma base de dados cujo objetivo é fornecer valores normativos de familiaridade, concretude, valência, ativação, corporalidade e transparência para 213 expressões metafóricas que ilustram 20 MC em português europeu e em inglês britânico. Neste sentido, este estudo apresenta uma abordagem pioneira da avaliação subjetiva de expressões metafóricas refletida de três formas. Primeiramente, este será, tanto quanto sabemos, o primeiro estudo onde se apresentam valores normativos para a linguagem metafórica avaliada em várias dimensões no português europeu, contribuindo assim para a investigação futura nesta área em português. Em segundo lugar, e ainda que os resultados preliminares apresentados se prendam apenas com o português, esta base de dados contempla uma abordagem interlinguística na qual os componentes avaliados são comuns à língua inglesa. Finalmente, procurámos incluir expressões metafóricas ilustrativas de MC, ao invés da apresentação das MC no seu formato formal de A é B utilizada até agora, a fim de contribuir para uma representação menos artificial das MC.

Assim sendo, os resultados apresentados neste artigo, ainda que tratando apenas os dados preliminares do estudo, demonstram que mesmo ao nível de expressões metafóricas parece haver uma relação essencial entre linguagem metafórica e a sua utilização do contexto social e físico, sendo possível observar o modo como diferentes dimensões subjetivas impactam a nossa interpretação de linguagem metafórica. Nomeadamente, é possível destacar o modo como dimensões como familiaridade e transparência se relacionam (sendo que expressões muito desconhecidas também se classificam como muito opacas), bem como o papel fundamental da fisicalidade na compreensão e utilização de noções mais abstratas, conforme demonstrado pela dimensão da corporalidade.

Em conclusão, e enquanto primeira base de dados desenvolvida em português europeu, com este trabalho espera-se poder contribuir para uma abordagem empírica da linguagem metafórica que inclua expressões utilizadas no dia a dia, ao invés de estímulos meramente artificiais. Por fim, espera-se também que elementos desta base possam ser utilizados para estudos futuros, nomeadamente servindo para garantir a utilização de linguagem natural controlada em diversos parâmetros considerados fulcrais no processamento metafórico.

Referências

Aono, K., Sasano, R., & Takeda, K. (2024). Verifying claims about metaphors with large-scale automatic metaphor identification. *In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 2, pp. 711–719). Association for Computational Linguistics.

Al-Azary, H., & Katz, A. N. (2021). Do metaphorical sharks bite? Simulation and abstraction in metaphor processing. *Memory & Cognition*, *49*, 557–570.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*, 639–647. https://doi.org/10.3758/BF03200465

Brugman, B. C., Droog, E., Reijnierse, W. G., Leymann, S., Frezza, G., & Renardel de Lavalette, K. Y. (2022). Audience perceptions of COVID-19 metaphors: The role of source domain and country context. *Metaphor and Symbol*, *37(2)*, 101–113.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*. https://doi.org/10.3758/s13428-013-0403-5

Campbell, S. J., & Raney, G. E. (2016). A 25-year replication of Katz et al.'s (1988) metaphor norms. *Behavior Research Methods*, *48*, 330–340.
Fauconnier, G., & Turner, M. (2003). Conceptual blending, form and meaning. *Recherches en Communication*, *19*, 131–149. https://doi.org/10.14428/rec.v19i19.48413
Gibbs, R. W. (2017). Metaphor Wars: Conceptual Metaphors in human life. Cambridge University Press.
Grady, J. (1997). *Foundations of meaning: Primary metaphors and primary scenes.* https://escholarship.org/uc/item/3g9427m2
Honeck, R. P., & Hoffman, R. R. (Eds.). (1980). *Cognition and figurative language* (1.ª ed.). Routledge. https://doi.org/10.4324/9780429432866
Ichien, N., Stamenković, D., & Holyoak, K. J. (2024). Large language model displays emergent ability to interpret novel literary metaphors. *Metaphor and Symbol*, *39(4)*, 296–309.
Jamrozik, A., McQuire, M., Cardillo, E. R., & Chatterjee, A. (2016). Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin & Review*, *23*, 1080–1089.
Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language*, *55(1)*, 18–32.
Katz, A. N., Paivio, A., & Marschark, M. (1985). Poetic comparisons: Psychological dimensions of metaphoric processing. *Journal of Psycholinguistic Research*, *14*, 365–383.
Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbolic Activity*, *3(4)*, 191–214.
Khatin-Zadeh, O., Farsani, D., Hu, J., Eskandari, Z., Zhu, Y., & Banaruee, H. (2023). A review of studies supporting metaphorical embodiment. *Behavioral Sciences*, *13*(7), 585.
Kövecses, Z. (2012). *Emotion concepts.* Springer Science & Business Media.
Kövecses, Z. (2020). Extended Conceptual Metaphor Theory. Cambridge University Press.
Kuperman, V. (2013). Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology*, *4*, 203. https://doi.org/10.3389/fpsyg.2013.00203
Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* University of Chicago Press.
Littlemore, J. (2001). The use of metaphor in university lectures and the problems that it causes for overseas students. *Teaching in Higher Education*, *6*(3), 333–349.
Littlemore, J. (2004). What kind of training is required to help language students use metaphor-based strategies to work out the meaning of new vocabulary? *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, *20*, 265–279.
Littlemore, J. (2019). *Metaphors in the mind: Sources of variation in embodied metaphor*. Cambridge University Press. https://doi.org/10.1017/9781108241441
Marschark, M., Katz, A. N., & Paivio, A. (1983). Dimensions of metaphor. *Journal of Psycholinguistic Research, 12*(1), 17–40.
Mateus, I, Oliveira, H. M, Teixeira, J., Soares, A. P. (2025). Is She Still Beneath Him? Spatial Power and the Gendered Mind, [Manuscript submitted for publication]
Reijnierse, W. G., Burgers, C., Bolognesi, M., & Krennmayr, T. (2019). How polysemy affects concreteness ratings: The case of metaphor. *Cognitive Science, 43*(8), e12779.
Santana, E., & Vega, M. (2011). Metaphors are embodied, and so are their literal counterparts. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00090
Casasanto, D., & Gijssels, T. (2015). What makes a metaphor an embodied metaphor? *Linguistics Vanguard, 1*(1), 327–337.
Schmidt, G. L., & Seger, C. A. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition, 71*(3), 375–386.
Siqueira, M., Gil, M. & Melo, T. (2010). Contribuições para a elaboração de testes psicolinguísticos: Construção de uma lista de sentenças. Gragoata, *29*(2), 127–145.
Soares da Silva, A. (ed.) (2021). Figurative Language: Intersubjectivity and Usage. John Benjamins.

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, S. (2011). The adaptation of the affective norms for English words (ANEW) for European Portuguese. *Behavior Research Methods*, *44*, 256–269. https://doi.org/10.3758/s13428-011-0131-7

Soares, A. P., Machado, J., Costa, A., Comesaña, M., & Oliveira, H. M. (2017). The Minho Word Pool: Norms for imageability, concreteness and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods*, *49*. https://doi.org/10.3758/s13428-016-0767-4

Steen, G. J. (2011). The Contemporary Theory of Metaphor – Now new and improved! Review of Cognitive Linguistics, *9*(1), 26–64.

Titone, D. A., & Connine, C. M. (1994). Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbol*, *9*(4), 247–270.

Anexo 1

Metáfora concetual	Expressão metafórica
Afeto é calor	Abraço caloroso
	O calor de um sorriso
	Resposta fria
	Uma abordagem fria
Alcançar um objetivo é chegar a um destino	Alcançar os meus objetivos
	Chegar ao sucesso
	Navegar o meu caminho
	Chegar à meta na vida
	Chegar a uma solução
	Atingir a minha meta
	Alcançar o sucesso
	Percorrer um longo caminho na vida
	Um passo em direção a um objetivo
Compreender é ver	Claro como o dia
	Ver o objetivo
	Ponto de vista
	Cego para a verdade
	A verdade é cristalina
	Ver o panorama geral
Considerar é pesar	Dar mais peso a esta argumento
	Este fator tem mais peso
	Encontrar o equilíbrio entre duas opções
	Pesar os prós e contras
	Pesar as consequências antes de agir
	Medir as consequências
Culpa é peso	Um peso sobre os meus ombros
	Uma consciência pesada
	O fardo da culpa
	Carregar culpa
	Sobrecarregado pela culpa
	Um peso na minha consciência
	O peso da culpa

	Derrubado pela culpa
	Mais leve depois de fazer a coisa certa
	Esmagado pela culpa
	Arrastar a culpa
	Um olhar pesado de culpa
	Suportar o peso da culpa
	Culpa suportada
	Estar carregado de culpa
Dificuldade é peso	Trabalho pesado
	Emocionalmente difícil de suportar
	Carregar o fardo da responsabilidade
	Carregar o peso das expectativas
	Pressão esmagadora
	Sobrecarregado por todos os desafios
Discussões são guerras	Atacar um argumento
	Defender uma posição
	Deitar abaixo uma proposta
	Contrapor as alegações
	Lançar um ataque verbal
	Manobras estratégicas nos debates
	Perder uma discussão
	Ataque de críticas numa discussão
	Ganhar numa discussão
	Retirar-se da discussão
	Preparado para a batalha em qualquer discussão
Existência é visibilidade	Esta música apareceu pela primeira vez
	Desapareceu do vocabulário de hoje em dia
	O assunto vem à tona
	Isto é só a ponta do iceberg
Felicidade é em cima	Sentir-se inferior
	Sentir-se em alta
	Cabeça erguida
	Sentir-se no topo do mundo
	Sentir-se em baixo
	Felicidade crescente

	Sentir-se nas alturas
	Sentir-se nas nuvens
	Afogar-se em tarefas
	Um aumento da felicidade
	Sentir-se abatido
	Sentir-se cabisbaixo
	Chegar ao fundo do poço
	Por água abaixo
	O auge da felicidade
	Cair aos bocados
Ideias são alimento	Devorar um livro novo
	Sedento por novas ideias
	Consumir informação
	Alimentar o meu cérebro com este livro
	Digerir os conceitos da aula
	Alimentar-se com novas descobertas
	Mastigar palavras
Importância é peso	Entrevistas de peso
	Filmes ligeiros
	Leitura leve
	Leitura pesada
	O peso da importância
	Ter peso na decisão
	Peso da responsabilidade
Mais é em cima	As vendas estão em baixo
	As ações estão a cair
	Os falcões voam mais do que os pardelhos
	Preços a subir
	Preços a descer
	Atingir níveis mais altos
	Preços altos
	Preços baixos
	Preços reduzidos
	Preços elevados
	Voar mais alto

	Subir a novos patamares
	Um pico nas vendas
	A empresa está a afundar-se
	Descer ao vazio
O bem/moralidade é em cima	Elevar-se acima do mal
	Elevação moral
	Superioridade moral
	Trajetória moral ascendente
	Baixar os padrões morais
	Declínio moral
	Ética em queda livre
	Padrões morais baixos
	Estar à altura da ocasião
	Alcançar as estrelas
	Ideais elevados
	Ascender em direção à virtude
	Uma carreira em declínio
	Alta qualidade
	Pôr alguém para baixo
	De mente elevada
	Um indivíduo de baixo nível
	Expectativas elevadas
	Expectativas baixas
	Pessoa rasteira
O tempo é um recurso	Tempo gasto
	Tempo poupado
	Tempo desperdiçado
	Investir tempo
	O valor de uma hora
	Usar o meu tempo
	O tempo que ainda me resta
	Perder tempo
	Valorizar o meu tempo
Poder é em cima	Sua alteza real
	Subordinados na equipa

	Um membro superior da equipa
	Um membro inferior da equipa
	Subir ao poder
	Trepar a hierarquia
	Chegar ao topo da empresa
	Escalar a hierarquia
	Ascender à liderança
	Elevado a um cargo de liderança
	Crescer em influência
	Escalar ao poder
	Mobilidade ascendente na minha carreira
	Oficiais de alta patente
	Uma posição no fundo da organização
	Ser espezinhado
	Cair em desgraça
	Controlo sobre mim
	Sob minha supervisão
	O topo da empresa
	Subir na vida
	Escada social
	Estatuto elevado
	Subir na posição social
	Ascender na sociedade
	Descer no estatuto social
	Indivíduo de estatuto inferior
	Indivíduo de estatuto superior
Quantidade é tamanho	Uma grande multidão
	Um enorme sucesso
	Um pedido grande
	Uma ampla gama de opções
	Uma conversa profunda
	Uma carga de trabalho pesada
	Uma reunião curta
	Um enorme esforço
	Uma pequena fração de tempo

Segredos são pesos	Carregar um segredo pesado
	Suportar o peso de um segredo
	Um segredo é um fardo pesado
	Sufocar sob o peso do sigilo
	O fardo do silêncio
	O peso esmagador de um segredo
Uma crença é uma posição/orientação física	Tomar uma posição
	Enraizado nas convicções
	Agarrar-me à minha fé
	Mudar para uma nova perspetiva
	Estar do lado da justiça
	Enfrentar oposição
	Manter-se firme contra as críticas
	Afastar-se de crenças antigas
Uma situação é um local	Tomar uma posição firme
	Uma boa posição no trabalho
	Meter-se numa situação
	Uma encruzilhada na vida
	No caminho certo na vida
	Em apuros
	No mesmo barco
	De volta ao jogo no trabalho
	A caminho de um desastre
	Num limbo
	No meio de uma crise
	Num dilema
A vida é uma viagem	Embarcar numa nova aventura na vida
	Navegar pela vida
	As reviravoltas da vida
	Encruzilhadas na vida
	Chegar a um beco sem saída na vida
	Seguir em frente na vida
	Explorar território desconhecido na vida
	Apreciar a viagem, não apenas o destino na vida
	Encontrar o meu caminho na vida

	Enfrentar batalhas difíceis na vida
	Traçar o meu caminho na vida
	O caminho menos percorrido na vida
	Aceitar os altos e baixos da viagem da vida
	O caminho da vida
	Chegar a uma encruzilhada na vida

Anexo 2

Instruções verbatim para a dimensão subjetiva de familiaridade

Algumas expressões do dia a dia podem parecer mais familiares do que outras, dependendo da frequência com que as encontramos e utilizamos nas nossas vidas diárias. Por exemplo, a palavra "limpo" é muito utilizada e pode parecer muito familiar, enquanto a palavra "afável" é menos utilizada e pode parecer mais desconhecida. De seguida, ser-lhe-á pedido para classificar uma série de expressões quanto ao quão desconhecidas ou familiares podem ser. Para isso, deverá selecionar, na seguinte escala de 5 pontos, o número que melhor representa a sua avaliação. Se não conhecer alguma expressão, por favor selecione "não conheço". Responda rapidamente. Não existem respostas certas ou erradas. Não pense demasiado. Confie no seu instinto. Concentre-se no quão desconhecida ou familiar cada expressão é para si. Pronto/a? Vamos começar!

Instruções verbatim para a dimensão subjetiva de concretude

Algumas expressões do dia a dia são experienciadas através de um dos cinco sentidos (visão, olfato, audição, paladar e tato). Outras expressões não podem ser experienciadas diretamente através de um dos sentidos. Por exemplo, "doce" geralmente suscita a experiência concreta de doçura enquanto sabor, sendo uma palavra muito concreta. No entanto, "justiça" não pode ser experienciada através dos sentidos, sendo uma palavra abstrata. De seguida, ser-lhe-á pedido que classifique uma série de expressões quanto à sua abstração ou concretude. Para isso, deverá selecionar, na seguinte escala de 5 pontos, o número que melhor representa a sua avaliação. Se não conhecer alguma expressão, por favor selecione "não conheço". Responda rapidamente. Não existem respostas certas ou erradas. Não pense demasiado. Confie no seu instinto. Concentre-se no quão abstrata ou concreta cada expressão é para si. Pronto/a? Vamos começar!

Instruções verbatim para a dimensão subjetiva de valência

Algumas expressões do dia a dia podem impactar as nossas emoções. Certas expressões podem suscitar sentimentos de alegria, satisfação ou contentamento. Outras expressões podem invocar sentimentos de tristeza, insatisfação ou descontentamento. Por exemplo, a palavra "honesto" pode provocar sentimentos de alegria e satisfação, enquanto a palavra "cruel" pode provocar sentimentos de tristeza e insatisfação. De seguida, ser-lhe-á pedido para classificar uma série de expressões quanto aos sentimentos de tristeza ou alegria que invocam em si. Para isso, deverá selecionar, na seguinte escala de 5 pontos, o número que melhor representa a sua avaliação. Se não conhecer alguma expressão, por favor selecione "não conheço". Responda rapidamente. Não existem respostas certas ou erradas. Não pense demasiado. Confie no seu instinto. Concentre-se no quão triste ou alegre cada expressão o/a faz sentir. Pronto/a? Vamos começar!

Instruções verbatim para a dimensão subjetiva de ativação

Algumas expressões do dia a dia podem provocar sentimentos de entusiasmo e ativação. Outras expressões podem invocar sentimentos de calma e relaxamento. Por exemplo, a palavra "aventura" pode suscitar

entusiasmo e ativação, enquanto a palavra "cadeira" pode provocar sentimentos de relaxamento e calma. De seguida, ser-lhe-á pedido para classificar uma série de expressões quanto aos sentimentos de calma ou ativação que invocam. Para isso, deverá selecionar, na seguinte escala de 5 pontos, o número que melhor representa a sua avaliação. Se não conhecer alguma expressão, por favor selecione "não conheço". Responda rapidamente. Não existem respostas certas ou erradas. Não pense demasiado. Confie no seu instinto. Concentre-se no quão calmo ou ativado cada expressão o/a faz sentir. Pronto/a? Vamos começar!

Instruções verbatim para a dimensão subjetiva de corporalidade

Algumas expressões do dia a dia envolvem facilmente o corpo humano, enquanto outras expressões não envolvem o corpo humano com a mesma facilidade. Por exemplo, a palavra "correr" envolve imediatamente o corpo humano, sendo uma palavra muito corpórea. No entanto, a palavra "dissolver" não envolve facilmente o corpo humano, sendo uma palavra incorpórea. De seguida, ser-lhe-á pedido para classificar uma série de expressões quanto à sua incorporalidade ou corporalidade. Para isso, deverá selecionar, na seguinte escala de 5 pontos, o número que melhor representa a sua avaliação. Se não conhecer alguma expressão, por favor selecione "não conheço". Responda rapidamente. Não existem respostas certas ou erradas. Não pense demasiado. Confie no seu instinto. Concentre-se no quão incorpórea ou corpórea cada expressão é para si. Pronto/a? Vamos começar!

Instruções verbatim para a dimensão subjetiva de transparência

Algumas expressões do dia a dia podem ser transparentes ou opacas no seu significado, dependendo do quão fácil é adivinhar o seu sentido a partir das palavras que as compõem. Por exemplo, a expressão "estar nas nuvens" é facilmente compreendida e muito transparente, enquanto a expressão "amigo da onça" não é tão facilmente compreendida e pode ser mais opaca. De seguida, ser-lhe-á pedido para classificar uma série de expressões quanto à sua opacidade ou transparência. Para isso, deverá selecionar, na seguinte escala de 5 pontos, o número que melhor representa a sua avaliação. Se não conhecer alguma expressão, por favor selecione "não conheço". Responda rapidamente. Não existem respostas certas ou erradas. Não pense demasiado. Confie no seu instinto. Concentre-se no quão opaca ou transparente cada expressão é para si. Pronto/a? Vamos começar!

Níveis e descritores de complexidade textual para adultos de baixa literacia: um referencial do projeto iRead4Skills

Levels and descriptors of textual complexity for adults with low literacy: a benchmark from the iRead4Skills project

Ricardo Monteiro[1], *Susana Correia*[1], *Raquel Amaro*[1], *Michell Moutinho*[1], *Sílvia Barbosa*[1], *Maria Leonor Reis*[1, 2]

[1] Centro de Linguística da Universidade Nova de Lisboa
[2] Academia das Ciências de Lisboa

Abstract

This paper presents the iRead4Skills benchmark of textual complexity, a novel approach to text complexity difficulty tailored to low-literacy adult native speakers. Despite the existence of various instruments and guidelines related to reading proficiency (CEFR, PISA, etc.), little attention is given to adult reading difficulties and how these manifest linguistically in texts. Recognizing the pivotal role of literacy in personal and professional development, this tool addresses the challenges of identifying and adapting texts for this demographic. Based on existent benchmarks (ANQEP, CEFR, PIAAC, ALTE), this article builds upon them and proposes the delineation of levels of textual complexity for European Portuguese, pertaining to low-literacy adults. Integrating lexical, syntactic, and cohesion features, the framework defines three levels of textual complexity—Very Easy, Easy, and Plain— roughly aligned with the CEFR levels A1, A2, and B1. Descriptors of complexity are also advanced to detail each level. Alongside this proposal, the levels and descriptors were tested against real texts, resulting in the compilation of a *corpus*, graded by levels of complexity. With the aim of testing and validating this proposal, both the levels and descriptors were evaluated and discussed by experts in Adult Learning and Vocational Education and Training. The iRead4Skills benchmark of textual complexity aims at bridging theoretical research and practical application by combining quantitative and qualitative measures, ensuring a robust and contextually relevant analysis of textual complexity. Key contributions include the framework's alignment with authentic texts and its validation by experts in Adult Learning and Vocational Education and Training. Practical applications of this framework encompass innovative tools such as a text complexity evaluation system and a writing assistant for text adaptation, addressing gaps in suitable reading materials for low-literacy adults. This research offers valuable insights for cross-linguistic studies on text complexity, paving the way for future exploration in literacy and text complexity frameworks.

Keywords: literacy, textual complexity, descriptors, benchmark, low-literacy adults.

Resumo

Este trabalho pretende descrever o processo de desenvolvimento de um referencial de complexidade textual inserido no projeto iRead4Skills. O nosso objetivo é fazer uma descrição e análise dos traços linguísticos que conferem complexidade aos textos, em particular àqueles que se destinam a adultos de baixa literacia. Apesar da existência de orientações sobre a proficiência em leitura (QECR, PISA, etc.), pouca atenção é dada às dificuldades de leitura sentidas por adultos em programas de ensino e formação de adultos (EFA), e a como estas dificuldades se manifestam linguisticamente nos textos. Baseado em referenciais usados atualmente (ALTE, ANQEP, QECR,

PIAAC), este trabalho propõe a definição de níveis de complexidade do português europeu para adultos de baixa literacia, com base na análise crítica desses referenciais, na observação e análise de textos, e na validação por profissionais da área de EFA. Composta por três níveis – Muito Fácil, Fácil e Claro –, esta proposta considera para cada nível aspetos linguísticos como a dimensão léxico-conceptual, a ocorrência de certos tipos de estruturas sintáticas e a coesão textual, como indutores e/ou marcadores de complexidade. Estas dimensões são detalhadas através de descritores associados aos diferentes níveis de complexidade textual estabelecidos. Paralelamente, e de modo interativo, os níveis e descritores de complexidade foram usados na compilação de um *corpus*, para a seleção de textos para cada nível a partir dos descritores propostos e da identificação dos traços linguísticos específicos de cada texto. Com o intuito de testar e validar a proposta, os níveis e descritores foram levados a avaliação e discussão por professores de EFA. Fazendo a ponte entre a investigação e a aplicação prática em sala de aula, o referencial aqui avançado contribui para a criação de ferramentas automáticas de análise de complexidade textual e para o desenvolvimento de assistentes de escrita de textos com diferentes níveis de complexidade, o que poderá ajudar a desenvolver materiais de leitura adequados a adultos de baixa literacia.

Palavras-chave: literacia, complexidade textual, descritores, referencial, adultos de baixa literacia.

1. Introdução

A literacia é essencial para os indivíduos prosperarem no mundo atual, afigurando-se, cada vez mais, como uma competência vital para o desenvolvimento pessoal e profissional. Ao mesmo tempo, estabelece uma correlação positiva com outros aspetos da vida dos indivíduos, desde melhor remuneração a melhores hábitos de saúde (OECD, 2013a). Com efeito, as questões à volta da literacia na sociedade são cada vez mais discutidas e reconhecidas como importantes para uma vida digna. Atualmente, a noção da literacia é entendida como operando num *continuum* de proficiência, e é sempre uma competência indispensável a jovens e adultos, independentemente da sua posição nesse *continuum* (OECD, 2021; Wagner, 2011). Neste contexto, surge o conceito de "aprendizagem ao longo da vida", incorporado nos objetivos de programas de requalificação pessoal e profissional de adultos, e que tem sido posto em prática, tanto a nível internacional como nacional, com vista à melhoria das competências literácitas de adultos. No entanto, encontrar materiais de leitura adequados para este tipo de população – especialmente aqueles que são autênticos e relevantes – pode revelar-se uma tarefa difícil. Muitos programas de ensino e formação de adultos (EFA) fazem uso de materiais de leitura provenientes de programas escolares destinados a uma população infantil. O ensino de adultos através de materiais deste género revela-se, assim, uma tarefa difícil, havendo uma desadequação dos materiais de leitura ao contexto de ensino, podendo, isso, tornar-se num obstáculo à promoção do desenvolvimento de práticas de leitura.

Este problema está na origem do projeto iRead4Skills.[1] Reconhecendo a importância das competências de leitura, este projeto, iniciado em 2023, tem como meta a promoção do desenvolvimento de práticas de leitura através do desenvolvimento de um sistema que avalia a complexidade textual e sugere textos adequados a leitores com baixa literacia. O intuito é que o acesso a textos adequados proporcione uma reaproximação dos indivíduos à leitura, o que, por sua vez, estimulará o desenvolvimento das práticas e hábitos de leitura e contribuirá para o desenvolvimento pessoal e a qualificação profissional.

[1] https://iread4skills.com/

2. Enquadramento teórico

Nas últimas décadas, o conceito de literacia tem estado cada vez mais presente no contexto das políticas educativas europeias. A literacia é um aspeto fundamental para o acesso à informação e ao conhecimento, na possibilidade de aprender ao longo da vida, bem como para construir uma análise crítica do mundo e exercer uma cidadania ativa (Ávila, 2008). Por essa razão, os sistemas educativos, nomeadamente decisores políticos, escolas e centros de formação, têm reforçado a necessidade de promover a literacia, nas suas diversas formas (por exemplo, a literacia digital, a literacia financeira ou a literacia para a saúde), como uma competência essencial para os indivíduos na sociedade atual. Apesar disso, a relação entre o nível de escolaridade de uma população e o seu perfil de literacia não é óbvia nem direta (Rothes & Queirós, 2021; Scribner & Cole, 1981). Uma sociedade altamente escolarizada não é, necessária e obrigatoriamente, uma sociedade com mais competências literácitas. Para compreender esta competência nos indivíduos, é necessário compreender igualmente como estes usam as suas aptidões de leitura e escrita enquanto participantes ativos da sociedade.

2.1. Literacia

Quando se fala de literacia, uma abordagem descritiva das suas definições, ao invés de normativa, é recomendada (Rothes & Queirós, 2021). Isto traduz-se no entendimento da literacia não como uma competência dicotómicaque se tem ou não –, mas como operando num *continuum*, que pode ser desenvolvido e variar de indivíduo para indivíduo (Hanemann, 2016; UNESCO Institute of Statistics, 2020). Assim, a literacia configura-se como um processo dinâmico e não um estado fixo (Alves, 2005), podendo assumir diferentes formas e finalidades consoante os contextos sociais e culturais em que se insere.

É igualmente importante entender a literacia como um fenómeno que ultrapassa a mera esfera individual e é, também, influenciado por fatores externos: os textos em práticas sociais, contextual e temporalmente situadas (Alves et al., 2012). Estes contextos concretos são designados por *literacy events* (Fransman, 2005). Nestes, a interação do indivíduo com uma prática social é feita através de um texto, e, de forma reflexa, a interação com o texto é feita no âmbito da prática social. Por conseguinte, os usos e práticas dos textos não apresentam sempre as mesmas características, variando conforme as esferas de vida (Dionísio et al., 2014). Desta maneira, torna-se claro que a dimensão social influencia a interação com o texto, sendo a competência em literacia de cada indivíduo dependente do contexto social no qual é posta em prática.

É, pois, na relação recíproca entre o indivíduo e a sociedade que a literacia assume um papel cimeiro. Atualmente, o impacto da literacia é visível ao nível socioeconómico e na competitividade entre países (Cunha & Dionísio, 2019). A literacia é vista como uma prática social, na qual os falantes de uma língua utilizam as suas aptidões de leitura e escrita para o seu desenvolvimento pessoal e profissional (OECD, 2021; Stromquist, 2006). Complementarmente, a Organização para a Cooperação e Desenvolvimento Económico (OECD) salienta a relação entre baixos níveis de literacia e uma população adulta desempregada (OECD, 2013a). Nesta senda, múltiplas campanhas de literacia por parte da UNESCO têm estabelecido uma ligação entre indicadores mensuráveis da literacia e outros indicadores de desenvolvimento, por exemplo, de economia ou de modernização (Wagner, 2011). Da mesma forma, uma análise baseada nos resultados do *Survey of Adult Skills-PIAAC* (OECD, 2013a), constata que, em média, indivíduos com níveis de literacia considerados altos têm salários 60% superiores a indivíduos com níveis de literacia considerados baixos (Schleicher, 2012). Comparando estas duas populações no que diz respeito à situação profissional, foi ainda demonstrado que indivíduos de baixa em literacia têm o dobro da probabilidade de se encontrarem desempregados (Schleicher, 2012). De facto, as consequências da literacia noutros aspetos da vida, à primeira vista sem relação, parecem não ficar por aqui: uma população com altos níveis de literacia goza de um maior poder de intervenção política, de uma maior participação em atividades de voluntariado, bem como de

uma saúde boa ou excelente (Schleicher, 2012). Complementarmente, baixos índices de literacia são frequentemente associados a disfuncionalidade profissional e a défices sociais.

Importa salientar que níveis de escolaridade formais não garantem, por si sós, níveis elevados de literacia. Os dados do mais recente estudo PIAAC (OCDE, 2024a) indicam que Portugal continua abaixo da média da OCDE em termos de proficiência em literacia, situando-se no penúltimo lugar entre os países participantes. Este défice abrange não apenas adultos com baixa escolaridade, mas também indivíduos com níveis de escolaridade mais avançados. Isto confirma que uma sociedade escolarizada não é automaticamente uma sociedade mais letrada (Scribner & Cole, 1981), e reforça a importância de estudar a literacia como prática situada, mais do que como mera acumulação de anos de escolaridade.

2.2. **Ensino e formação de adultos**

De modo a contrariar os efeitos negativos da baixa literacia nos indivíduos e na sociedade, têm sido postos em prática vários programas de incentivo à requalificação pessoal e profissional, tanto a nível internacional como nacional. Estes programas baseiam-se em valores humanísticos, democráticos e emancipatórios e têm como objetivo requalificar e integrar uma população menos letrada (Hanemann, 2016). Através do ensino formal e informal, destinados a várias faixas etárias e com uma vasta abrangência social, providenciam-se oportunidades de aprendizagem, seguindo o princípio que aqui tem vindo a ser mencionado: que a competência de literacia de cada indivíduo pode ser desenvolvida, independentemente dos fatores externos. De acordo com dados da EDUSTAT – Fundação Belmiro de Azevedo (2024), no ano letivo 2022-2023 estavam inscritos mais de 37.000 adultos em cursos de formação EFA, a grande maioria (78%) com idades superiores a 23 anos. Segundo Correia et al. (2024), este perfil caracteriza-se por ter idade entre 18 e 40 anos, residir na região de Lisboa, ser do sexo feminino, empregado, casado, com filhos e com ensino básico concluído.

O sucesso destes programas está, assim, fortemente condicionado à capacidade de responder às necessidades reais dos indivíduos, as quais passam, muitas vezes pela contextualização dos processos de ensino. (Alves et al., 2012; Cunha & Dionísio, 2019). Um dos principais obstáculos identificados pelos profissionais da área prende-se com a escassez de materiais de leitura adequados, tanto em conteúdo como em grau de complexidade. Este problema foi identificado durante o desenvolvimento do projeto iRead4Skills, através de contributos diretos da Direção-Geral da Educação - Plano Nacional de Leitura 2027 (PNL), entidade parceira do consórcio. O PNL reportou que, em muitos contextos de ensino de adultos, é prática comum a reutilização de materiais pedagógicos concebidos para crianças ou jovens, nomeadamente fichas de trabalho, textos escolares e exercícios de manuais do 1.º e 2.º ciclos. Esta reutilização, ainda que prática, levanta sérios desafios pedagógicos, uma vez que os textos utilizados nem sempre estão ajustados às vivências, interesses ou competências linguísticas dos adultos, podendo resultar em frustração ou desmotivação. Estes materiais são, assim, desajustados: ou por serem excessivamente simplistas ou infantis, ou por utilizarem uma linguagem ou estrutura desfasada da experiência de vida dos adultos. Por exemplo, os padrões de leitura (i.e., o hábito, frequência e preferências de leitura) tornam-se tipicamente mais ricos com a idade e, por isso, são mais avançados nos adultos do que nas crianças (Flor et al., 2013). Isto sugere que há uma possibilidade de "amadurecimento" a partir de certas estratégias de leitura, que são desenvolvidas de forma implícita em cada indivíduo. Assim, a consequência de materiais inadequados é dupla: perda de motivação e ausência de identificação com os textos propostos.

Esta inadequação é particularmente preocupante num contexto em que os cursos EFA acolhem públicos heterogéneos em termos de idade, contexto profissional e escolarização. Tal diversidade exige um olhar mais fino

sobre os níveis de proficiência reais em leitura, que nem sempre se correlacionam com a idade ou o grau de escolaridade formal. A identificação de um perfil mais preciso do público-alvo e das suas práticas de leitura é, portanto, essencial para a produção de materiais adequados.

2.3. **Complexidade textual**

Cunningham e Mesmer (2014, p. 256) definem complexidade textual como um conjunto de características inerentes e exclusivas ao texto, que podem ser analisadas e estudadas de um modo distinto das dificuldades do leitor. Por outro lado, o *Common Core State Standards/English Language Arts* (CCSS/ELA) (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) define a noção de complexidade como algo subjetivo. Desta forma, o entendimento do que torna um texto complexo é visto como um fenómeno relativo, que varia de pessoa para pessoa (Kusters, 2003). O CCSS/ELA propõe ainda um sistema tripartido para avaliar a complexidade de um texto: a sua dimensão quantitativa, qualitativa, e a tarefa do leitor para o compreender.

Numa abordagem quantitativa, a dificuldade de um texto advém, sobretudo, da quantidade de ocorrências de algo que induz complexidade. Esta ideia é, essencialmente, executada através de fórmulas de *leiturabilidade*,[2] cuja aferição passa por contar ocorrências de determinadas propriedades linguísticas (e.g., palavras compostas, número de sílabas numa palavra, número de palavras numa frase). Esta contagem resulta, depois, num valor numérico associado a um nível de complexidade específico (Goldman & Lee, 2014). Por outro lado, a abordagem qualitativa tenta ir para além da quantidade de características linguísticas, que resultam em índices de leiturabilidade, fazendo com que estas estejam dependentes de uma interpretação posterior (Hiebert, 2011). Fundamentalmente, uma análise qualitativa da complexidade procura simular a inteligência humana e a forma como esta pode percecionar a complexidade. Nesta abordagem, estão incluídas as teorias de esquemas textuais,[3] modelos mentais representativos do processo de leitura,[4] assim como gramáticas de histórias.[5] São abordagens deste género que procuram aferir a complexidade textual de uma forma mais holística e qualitativa (Harley, 2013).

Neste contexto, a proposta de referencial de complexidade apresentada neste trabalho é desenvolvida para adultos, falantes nativos de português europeu de baixa literacia, através de uma combinação de medidas de aferição quantitativa, qualitativa e subjetiva. Por meio de uma abordagem *in tandem*, direcionada para um público-alvo concreto, pretende-se colmatar as lacunas de cada abordagem específica (Goldman & Lee, 2014; Hiebert & Pearson, 2014), enquanto se emprega uma forma inovadora, potencialmente capaz de identificar manifestações de complexidade textual não analisadas até ao momento.

3. **Metodologia**

Nesta seção, é detalhado o processo de desenvolvimento de uma infraestrutura de complexidade textual destinada a adultos de baixa literacia, para avaliar a dificuldade de textos e sugerir leituras apropriadas para estes indivíduos. É, também, descrito o processo de elaboração do referencial com descritores de complexidade, de forma

[2] Tradução de *readability*.
[3] O conceito de teorias de esquemas textuais baseia-se na organização (esquematização) dos elementos do texto com base no conhecimento prévio do leitor destes mesmos elementos e a forma como se podem relacionar com outros elementos (Anderson, 2000).
[4] Os modelos mentais de leitura pretendem representar a informação transmitida por um texto à medida que esta informação é dada. O objetivo é reproduzir as imagens mentais criadas pelos leitores na sua memória de curto prazo, e como estas percecionam elementos como o tempo, espaço, causalidade, etc. (Harley, 2013).
[5] Gramáticas de histórias identificam e analisam a estrutura de uma história, assim como os seus elementos.

a contribuir para a classificação dos níveis de complexidade textual. Para alcançar esta meta, o nosso trabalho passou por três fases distintas:

i. Compilação e análise de referenciais com descritores e orientações relevantes para a avaliação da complexidade textual.

ii. Definição de uma proposta de níveis de complexidade, a partir das características linguísticas de textos adequados a adultos nativos de baixa literacia.

iii. Validação da proposta por parte de um *focus group* constituído por profissionais de EFA, com experiência em ensino de adultos de baixa literacia.

3.1. **Análise de referenciais**

Nesta primeira fase, foi efetuado um levantamento dos documentos orientadores pertinentes à complexidade textual. Quatro referenciais foram considerados relevantes: o *Quadro Europeu Comum de Referência para as Línguas* (QECR) (Council of Europe, 2020), o *Inquérito às Competências dos Adultos* – PIAAC (OECD, 2013b), o *Can Do Project*, da Associação de Examinadores de Línguas na Europa (ALTE, 2002) e o *Referencial de Competências-Chave de Educação e Formação de Adultos* – Nível Básico da Agência Nacional para a Qualificação e o Ensino Profissional (ANQEP) (Valente, 2021). À exceção do último, os documentos selecionados não tinham como público-alvo adultos de baixa literacia, abordando as competências comunicativas e linguísticas numa ótica mais geral ou no contexto de aprendizagem de língua segunda (L2)/língua estrangeira (LE).

Embora o público-alvo do presente estudo seja constituído por adultos com baixos níveis de literacia, falantes nativos do português europeu, a opção metodológica seguida foi a de abordar os textos a partir da noção de complexidade textual, em vez da noção de dificuldade textual. Esta escolha implica uma mudança de foco conceptual: enquanto a dificuldade textual parte da perspetiva do leitor e da sua interação com o texto, a complexidade textual é tratada como um conjunto de propriedades linguísticas e discursivas inerentes ao próprio texto, categorizáveis independentemente do leitor (Hiebert & Mesmer, 2013).

Optámos por esta abordagem com o intuito de avaliar e descrever os textos em termos das suas propriedades formais e estruturais, sem partir de pressupostos sobre a capacidade de leitura de um grupo específico de indivíduos. Só num momento posterior da investigação foi feita a correspondência entre os níveis de complexidade e a acessibilidade para o público-alvo. Neste sentido, a inclusão de referenciais concebidos para o ensino do português como língua segunda ou estrangeira (L2/LE) – nomeadamente descritores de proficiência do QECR – não teve como objetivo equiparar o nosso público-alvo a falantes não nativos. Tais referenciais foram usados como ferramentas operacionais, amplamente testadas, que permitiram descrever níveis de complexidade linguística com base em critérios já estabilizados na investigação aplicada, tais como o domínio de estruturas sintáticas, uso lexical ou coesão. Assim, a abordagem adotada é intencionalmente heurística. Ao descrever a complexidade dos textos com base em critérios independentes do/a leitor/a, é possível, numa fase posterior, observar com maior precisão quais níveis de complexidade se associam a que tipos de dificuldades no público-alvo. Esta metodologia também permite maior transferibilidade do referencial para outros contextos.

Na análise destes quatro documentos, foi dada especial atenção aos três níveis iniciais propostos em cada um. A razão para esta abordagem está relacionada, novamente, com o facto de estarmos perante referenciais que, na sua maioria, não incidem exclusivamente sobre adultos de baixa literacia. O facto de o conceito da literacia ser algo abrangente, desdobrável e passível de ser abordado de várias maneiras, faz com que seja difícil definir o que é um

indivíduo com baixa literacia. Ademais, ao contrário dos níveis de escolaridade formais, não é fácil fazer corresponder níveis de literacia pré-estabelecidos a leitores/as. Desta forma, a nossa abordagem passa por enquadrar a literacia no contexto pessoal e social (e.g., argumentar um ponto de vista, ler e compreender notícias, captar informação sobre transportes públicos, etc.) (Fransman, 2005; Street, 1993).

Assim, é preciso considerar que adultos de baixa literacia não se confinam apenas a um nível, mas que existe uma escala capaz de encapsular a heterogeneidade desta população. Com efeito, nos referenciais analisados, os primeiros níveis de proficiência em leitura alinham-se com as competências literácitas e géneros/tipos de texto associados com este tipo de população. Da mesma forma, estes níveis dão pistas em relação aos limiares do enquadramento da baixa literacia de adultos. Por exemplo, no QECR, o nível B2 – nível imediatamente acima dos que foram considerados pelo projeto –, descreve um indivíduo capaz de "ler artigos e reportagens sobre assuntos contemporâneos em relação aos quais os autores adotam determinadas atitudes ou pontos de vista particulares" (Council of Europe, 2020, p. 177). Uma proficiência de leitura deste género é aqui vista como não associada a um adulto de baixa literacia. No ponto oposto, o nível A0 corresponde ao estágio de alfabetização, no qual não existe qualquer atividade de leitura autónoma. Assim, a proficiência de leitura nestes dois extremos não é relevante para o público-alvo em questão, cingindo a nossa análise às características dos níveis intermédios – A1, A2 e B1.

3.2. **Proposta de níveis de complexidade**

Depois de analisados os referenciais e designados os níveis de complexidade apropriados a um contexto de baixa literacia, foram compiladas e analisadas as características associadas a cada nível. O objetivo passou por discutir descrições subjetivas e pouco precisas (e.g., *Knowledge and skill in recognizing basic vocabulary determining the meaning of sentences, and reading paragraphs of text is expected* (OECD, 2013b, p. 71)) e, sempre que possível transpô-las para – ou associá-las a – descrições que contivessem propriedades linguísticas objetivas (e.g., ausência de verbos conjugados no Modo Conjuntivo). Estas propriedades servem, assim, como descritores de cada nível.

Depois do levantamento destes descritores linguísticos (ver Anexo 1), partiu-se para uma análise da sua presença em textos autênticos de modo a:

i. Determinar quão prototípicos de um nível seriam (i.e., se um certo descritor ocorreria frequentemente nos níveis ou seria exclusivo de um nível ou género textual)

ii. Identificar outros descritores nos textos e especificar os já listados.

Este processo de validação e recolha de descritores em textos autênticos contribuiu também para o desenvolvimento de um *corpus* por nível de complexidade (Pintard et al., 2024). Este trabalho consistiu numa abordagem *in tandem*, na qual uma parte da equipa se encarregava de compilar textos para o *corpus* e outra os utilizava para a identificação e validação dos descritores (cf. Reis et al., 2024). Esta última tarefa pautava-se pela identificação de descritores de cada nível de complexidade num certo texto, de modo a atribuir um nível a esse texto. Uma tabela de ocorrências de descritores acompanhava cada texto e era nesta que as ocorrências de cada descritor eram anotadas. Depois, o número total de ocorrências era utilizado para decidir a que nível correspondia aquele texto ou se o mesmo seria descartado. O resultado deste trabalho foi um *corpus* de textos autênticos organizado por níveis de complexidade.

Por último, para além dos níveis de baixa literacia propostos, foi também postulado um nível de complexidade superior – o nível Mais Complexo. A consideração deste nível está ligada à necessidade de estabelecer um limiar de complexidade superior. Como referido anteriormente, existe também uma fronteira inferior natural de

complexidade, que distingue adultos de baixa literacia de adultos em processo de alfabetização. O objetivo destes limiares inferior e superior de complexidade é funcional, enquadrando a população de baixa literacia. O nível Mais Complexo, apesar de estar fora do escopo do projeto, foi usado para compilar outros textos, de modo a servir de referência para os níveis de complexidade considerados neste projeto.

O resultado desta etapa de trabalho foi uma proposta de três níveis de complexidade (juntamente de um quarto nível – Mais Complexo), assim como a sua descrição detalhada e os descritores linguísticos associados a cada um.

3.3. **Validação dos níveis e descritores de complexidade por formadores de adultos**

De modo a validar a proposta desenvolvida, foi pedido que pessoal formador de EFA (nomeadamente, em Centros Qualifica)[6] fizessem uma avaliação dos descritores para cada nível. Este foi um processo iterativo cujo objetivo principal não foi o de realizar uma validação estatística ou normativa, mas sim o de recolher contributos especializados para o aperfeiçoamento linguístico, funcional e pedagógico dos descritores propostos. Esta etapa visou validar a proposta inicial através da experiência de profissionais, assegurando assim uma validação situada e contextualizada.

Este processo incluiu a apresentação da proposta e o *input* dos/as formadores/as foi recolhido em duas fases:

i. Tarefa Preparatória – Esta tarefa, sob a forma de um questionário, consistiu em recolher as suas opiniões sobre os descritores de complexidade (cf. Monteiro et al., 2023). Neste questionário, eram apresentados, primeiramente, os níveis de complexidade e, depois, os descritores, estes últimos de modo aleatório. O que se pediu aos/às participantes foi que fizessem uma correspondência entre os descritores e os níveis de complexidade previamente apresentados. De modo a garantir a compreensão por parte de todas as pessoas envolvidas, independentemente da disciplina que lecionavam (e.g., matemática, língua portuguesa, etc.), os descritores foram apresentados sem recurso a terminologia linguística e acompanhados de exemplos ilustrativos.

ii. Sessão de *focus group* 1 – Esta primeira reunião com as/os formadoras/es teve como objetivo discutir os resultados da Tarefa Preparatória. Nesta reunião, foram abordados os conceitos de complexidade textual, assim como os níveis e descritores de complexidade da nossa proposta, com base nos resultados da tarefa preparatória. Foram discutidos, em particular, os descritores relativamente aos quais não houve consenso por parte das/os formadoras/es.

iii. Sessão de *focus group* 2 – A segunda reunião teve como objetivo a validação das alterações aos descritores, provenientes das conclusões da primeira reunião. Foram apresentados os descritores finais para cada nível, considerando o *feedback* dos participantes, e foi pedida uma validação final às/aos formadoras/es.

No total, 8 formadoras/es EFA participaram nas sessões de *focus group*. As sessões decorreram em ambiente colaborativo, com base em guiões estruturados.

O processo de validação e refinamento do nosso trabalho será continuado em fases posteriores do projeto, onde está prevista a aplicação de questionários e inquéritos a formadoras/es e formandas/os dos cursos EFA, com o objetivo de recolher informação empírica sobre a adequação da futura ferramenta digital, cuja base será o referencial

[6] Os Centros Qualifica oferecem educação e formação a adultos, no âmbito do programa Qualifica. Para mais informação, consulte https://www.qualifica.gov.pt.

aqui exposto. Nesta fase, a validação será centrada nos utilizadores finais e permitirá validar a proposta em termos de eficácia, clareza e relevância pedagógica.

Assim, a validação da ferramenta deve ser entendida como parte de uma estratégia de desenvolvimento cooperativo e progressivo, que articula contributos especializados na fase de construção com testes de aplicabilidade no terreno, a decorrer ao longo do projeto. Esta abordagem está alinhada com metodologias de investigação aplicada, em que a validação se realiza em ciclos sucessivos de desenvolvimento, aplicação, avaliação e reformulação.

4. Apresentação e discussão dos resultados

Nesta secção, são apresentados os resultados do processo de estabelecimento dos níveis, descritores de complexidade textual para o português europeu, desenvolvidos no projeto iRead4Skills, assim como a validação dos mesmos. Primeiramente, a análise dos referenciais supramencionados conduziu à definição de três níveis de complexidade textual: **Muito Fácil, Fácil** e **Claro**. A estes, junta-se o nível superior Mais Complexo, que tem apenas um papel delimitativo e de referência.

Para cada nível, é apresentada uma lista de marcadores linguísticos indutores de complexidade textual. Desta forma, a nossa proposta identifica um conjunto de características textuais que contribuem para a dificuldade de compreensão de um texto, incluindo características lexicais, sintáticas e de coesão textual.

4.1. Resultado da análise dos referenciais

Os resultados apresentados nesta secção correspondem à análise realizada sobre um conjunto de referenciais existentes no domínio da leitura e da proficiência linguística. Os referenciais analisados incluíram documentos destinados ao ensino do português como língua materna, língua segunda e língua estrangeira, bem como instrumentos internacionais de avaliação de literacia em adultos. A partir desta análise, foi possível isolar descritores representativos de diferentes níveis de complexidade textual, organizando-os em quatro dimensões principais, que servem de base à proposta apresentada nas seguintes secções.

4.1.1. Níveis de complexidade

Os três níveis de complexidade propostos são: **Muito Fácil**, **Fácil** e **Claro**. Importa referir, a este propósito, que estas designações para os níveis de complexidade foram estabelecidas de forma a classificar os textos e não público-alvo. A nomenclatura escolhida resultou de um processo iterativo e colaborativo no seio do projeto, em que foram testadas diferentes alternativas, acabando por ser validada com base na sua funcionalidade pedagógica e na sua inteligibilidade prática, mesmo que não perfeitamente homogénea do ponto de vista conceptual.[7] A estrutura progressiva está garantida não apenas pela ordem dos níveis, mas sobretudo pelos descritores linguísticos associados a cada nível, que refletem, de forma rigorosa, a crescente densidade lexical, complexidade sintática, riqueza coesiva e sofisticação discursiva dos textos.

A Tabela 1 apresenta uma descrição muito breve e acessível dos três níveis. Esta descrição, usada na tarefa de validação por formadoras/es de EFA, teve como objetivo apresentar uma descrição muito geral dos níveis de complexidade, para efeitos de contextualização da tarefa sem referência direta às propriedades dos textos. Não pretende apresentar de forma cabal as características de cada nível.

[7] Entende-se que os conceitos de "facilidade", "clareza" e "complexidade" podem ser vistos como distintos, operando em campos semânticos diferentes.

Tabela 1

Descrição Acessível dos Níveis de Complexidade Alvo

Nível	Descrição
Muito Fácil	Textos que são totalmente ou quase totalmente compreendidos por todas as pessoas, incluindo pessoas com escolaridade muito baixa (i.e., que não chegaram ao 6º ano do ensino básico) e que não têm quase nenhuma prática de leitura.
Fácil	Textos que são totalmente ou quase totalmente compreendidos por pessoas com escolaridade baixa (i.e., que completaram o 6º ano do ensino básico, mas que não têm mais que o 9º ano) e que têm pouca prática de leitura.
Claro	Textos que são compreendidos à primeira leitura por pessoas que completaram o 9º ano e que têm uma prática de leitura média.

Nestas descrições, cada nível de complexidade textual é associado a um nível de escolaridade para fornecer um ponto de referência familiar e pragmático. A definição dos níveis de complexidade fundamentou-se na noção de que a complexidade dos textos pode revelar-se sob a forma de três problemas distintos: problemas superficiais, de conhecimento e de conteúdo (Beck et al., 1984). O primeiro – problemas superficiais – relaciona-se com questões estritamente linguísticas, como é o caso de estruturas sintáticas complexas ou um uso inconsistente de conjunções. O segundo – problemas de conhecimento – assenta em algo potencialmente fora da dimensão linguística, ligado à falta de conhecimento prévio sobre o tópico presente (ou envolvente) no texto. O último – problemas de conteúdo – prende-se com a forma como as ideias são apresentadas e interligadas no texto, ou seja, com a forma como o texto está organizado e é comunicado. Nesta última dimensão, ideias e conceitos pouco familiares ou que se contrariam ou que levam a conclusões contraintuitivas podem levantar problemas.

A partir desta tipologia de problemas textuais, a complexidade foi analisada sob o prisma da observação de quatro dimensões textuais distintas: a dimensão léxico-conceptual, a dimensão das estruturas sintáticas, a dimensão da coesão e, por fim, a dimensão das formas verbais. Estas quatro dimensões distinguem-se daquelas existentes em referenciais de proficiência em leitura por incidirem numa população leitora diferente das desses referenciais - tipicamente, falantes não nativos ou crianças. Por exemplo, no caso de falantes não-nativos, estes tendem a encontrar complexidade, principalmente, em questões gramaticais (Heilman et al., 2007), e crianças tendem a ter menos proficiência em leitura em comparação com adultos, devido simplesmente à sua menor experiência de leitura (Mesmer et al., 2012).

A **dimensão léxico-conceptual** é a mais relevante das quatro. Questões de compreensão lexical são as primeiras a surgir quando um/a leitor/a interage com um texto (Harley, 2013; Hu & Nation, 2000; Laufer, 2020). Com efeito, para atingir 70% da compreensão de um texto, a pessoa precisa de conhecer cerca de 95% a 98% das palavras nesse mesmo texto (Schmitt et al., 2011). Tal implica que mesmo uma pequena percentagem de vocabulário desconhecido pode comprometer seriamente a compreensão global. Adultos de baixa literacia, ainda que experientes em língua oral, não têm práticas de leitura habituais e diversas e, por conseguinte, têm uma escassez de exposição a "textos formais", isto é, a textos nos quais são requeridas estratégias de compreensão diversas e mais avançadas. Desta forma, a **dimensão sintática** considera tanto medidas quantitativas como qualitativas da complexidade. Entre os marcadores incluem-se o **comprimento das frases**, **o número de orações subordinadas** e **a presença de construções sintáticas não canónicas**, como as orações passivas ou inversões. Embora adultos nativos com baixa

literacia possam dominar essas estruturas na oralidade, a sua presença em textos escritos pode representar obstáculos à fluidez da leitura e à construção de sentido.

No que toca à **coesão textual**, analisam-se sobretudo os mecanismos de **coesão referencial**, como as relações anafóricas, e **os marcadores temporais**, como a linearidade da narrativa textual (Fang & Pace, 2013; Glenberg et al., 1987). A ausência de pistas coesivas claras ou a sua ambiguidade dificulta a articulação entre frases e parágrafos, forçando o/a leitor/a a fazer inferências mais exigentes para manter a coerência do texto.

Por último, as formas e construções verbais, ainda que fazendo parte da dimensão lexical e morfossintática, são, aqui, consideradas independentemente. A razão para esta separação está relacionada com o facto de estas formas integrarem **informações aspetuais, temporais e modais**, manifestando, assim, complexidade textual de uma forma independente da estrutura sintática.

Como resultado, cada nível de complexidade possui uma descrição que engloba estas quatro dimensões, como se pode ver na Tabela 2. As informações detalhadas encontram-se sistematizadas sob a forma de descritores no Anexo 1.

Tabela 2

Níveis de Complexidade iRead4Skills

Nível Muito Fácil
Textos curtos e simples até cerca de 50 palavras, que se refiram ao desempenho de tarefas do quotidiano ou apresentem nova informação (ex.: textos didáticos). **Conceitos simples do dia a dia**, com o tópico da comunicação a ser apresentado inicialmente. Contextos básicos de comunicação: quotidianos (horários de transportes/listas/menus/informação de preço), comunicações pessoais (cartas, mensagens). **Palavras simples e frequentes**, comummente usadas em contextos do dia a dia (transporte, comida, família, trabalho). Uso predominante de palavras concretas que fazem referência a objetos concretos no mundo real (*maçã, cadeira, fogão*). Ausência de abreviaturas, siglas e estrangeirismos, à exceção daqueles muito frequentes (internet, hamburger). Palavras compostas comuns (*couve-flor, guarda-chuva*). Palavras com afixos frequentes (*infeliz, calmamente, trabalhador*). **Frases curtas e simples**, utilizando conjunções simples e na **ordem direta (Sujeito-Verbo-Complemento).** Estruturas de coordenação frequentes (copulativas, disjuntivas e adversativas) estão presentes (ex.: *O Rui e/ou a sua irmã foram à festa, mas o João não foi*). Estruturas de subordinação frequentes estão presentes (ex.: oração subordinada adverbial temporal – *Desde que voltou, o Rui não dormiu*). Uso predominante de frases ativas. **Tempos verbais no modo Infinitivo, Indicativo e Imperativo**, nomeadamente o Presente e Pretéritos Perfeito e Imperfeito. Ocorrência de construções perifrásticas comuns como *estar/começar/andar + a + infinitivo* ou *deixar/acabar + de + infinitivo*. Não figuram tempos compostos, à exceção do Futuro que é transmitido através de *ir + infinitivo*. **Cadeia referencial completa** sem omissão de elementos, com repetição em vez da utilização de pronomes (ex.: *O Rui gosta da Maria, porque a Maria é simpática*).

Localização temporal linear, dada por advérbios e conjunções simples (ex.: *Ontem, o Rui foi à praia. Os bombeiros chegaram depois da polícia).*

Ausência de figuras de estilo.

Nível Fácil

Textos curtos até 150 palavras, que são interessantes para o leitor **se informar/aprender algo novo ou para momentos de lazer.**
Conceitos mais específicos, mas do interesse do leitor (profissional, desporto, hobbies).
Contextos de comunicação profissionais e dos media (ex.: instrução profissional, notícias, anúncios, desporto).

Palavras e expressões frequentes que fazem referência a assuntos do dia a dia, mas também a outros âmbitos mais específicos (ex.: curso, ano escolar). O texto contém algumas palavras que referem ideias abstratas mais frequentes (*felicidade, imaginação, carinho*).
O texto utiliza abreviaturas comuns, depois de as ter explicado (ex.: *litro – l/metro – m*).
O texto apenas utiliza siglas e acrónimos conhecidos (CP, IVA, RTP, SIC, TVI).
Uso de estrangeirismos menos frequentes (*site, souvenir*).
O texto utiliza palavras compostas comuns com o seu sentido a poder ser compreendido pela combinação das duas partes (*óculos de sol, mal-educado, meia-idade, maus-tratos*).
Palavras com afixos menos frequentes (*pré-cozinhado, embalagem, finíssimo*).

Frases simples e complexas através de conjunções coordenadas e subordinadas (ex.: *Os músicos estavam tocando, mas poucos clientes ouviam. Quando pararam de tocar, apenas um senhor se manifestou*). Uso de orações relativas de sujeito, mas não de objeto (ex.: *O senhor que vive na casa ao lado ouviu o barulho*). O texto pode conter frases apresentadas numa ordem não linear (Complemento-Verbo-Sujeito) (ex.: *À escola, chegaram os autocarros da visita de estudo*).
Possível uso de frases passivas.

Uso de modos e tempos verbais variados.
Uso de tempos verbais compostos e de expressões verbais com *ter*/*poder*/*dever* (*posso/devo/tenho de gostar de bolo*).
Modos Conjuntivo, Condicional e Gerúndio.
Expressões com os verbos *permanecer, querer, poder, conseguir* em tempos simples (ex.: *permaneça calmo, queria almoçar, consigo ir*).
Utilização do Futuro simples (ex.: *ele gostará de te ver*).

O texto tem alguns elementos omitidos e utiliza pronomes (ex.: *Eu gostava da Maria vs. Gostava da Maria/Mas ela não gosta dele).*

Sequência ou localizações temporais lineares, dadas através de advérbios de tempo ou conjunções, **mas também por tempos verbais** (ex.: *Ele está a caminho e a festa já começou*).

Figuras de estilo frequentes são utilizadas.
O texto faz uso da **metonímia** (ex.: *beber um copo*), assim como **metáforas familiares** (ex.: *O meu pai tem olhos de águia*).

Nível Claro

Textos com cerca de 250 palavras sobre vários tópicos, sejam estes de informação ou lazer, com os leitores a conseguirem sair fora da sua zona de conforto.
Conceitos variados que necessitam de ser interligados para a sua compreensão total (ciência, política).
Contextos comunicativos variados: lazer (literatura, reportagens); **artes** (poesia); **contextos especializados** (*fóruns* online, comunicação profissional formal).

Palavras e expressões menos frequentes que fazem referência a âmbitos específicos (ex.: *doença cardíaca, artéria, tratamento*). As palavras remetem para **conceitos mais abstratos** (*envelhecimento*, *razão*, *fidelidade*).
Utilização de siglas e acrónimos não conhecidos pelo leitor juntamente com o nome a que se referem (ex.: *Organização das Nações Unidas (ONU), Instituto Público (IP*)).
O texto utiliza abreviaturas comuns, sem explicação prévia (ex.: *m, km, g., l, Exmo*.).
Uso frequente de estrangeirismos comuns (ex.: *link, hobby, marketing, mail, gay*).
O texto contém palavras que são ambíguas e/ou pouco frequentes. (ex.: *interno, representar, forma, paciente*).
Palavras com sufixos e prefixos variados (ex.: *nervosismo, lavável, curiosidade, altitude, superpoder, preconceituoso*).

Frases frequentemente combinadas através de conjunções coordenativas e subordinativas e, por vezes, com conjunções menos comuns (ex.: *apesar da confusão provocada pelos animais à solta, a situação foi controlada sem os animais serem atropelados*.)
Presença frequente de ordem não linear de constituintes(sujeito-verbo-complementos) (ex.: *de ar puro, todos precisamos*.).
Utilização habitual de frases passivas (ex.: *o almoço foi pago pela empresa*.), incluindo construções impessoais comuns e passivas de *-se* (ex.: *O trabalho faz-se bem*.).

Utilização de todos os modos e tempos verbais compostos e simples, à exceção do Pretérito Mais-que-perfeito.

Elementos omitidos (ex.: *[] mostrámos em estudos anteriores que a gordura solidifica quando [] arrefece. vs. Nós mostrámos em estudos anteriores que a gordura solidifica quando ela arrefece*). Utilização de **pronomes e anáforas lexicais** para indicar o que é referido (ex.: *no livro, nós damos uma definição complete de uma tempestade de neve e como esta se forma. Também indicamos onde o fenómeno pode ser visto*.)
Sequências ou localizações temporais não lineares (ex.: *após analisarem as respostas dos doentes que tinham sofrido um ataque cardíaco, os médicos confirmaram que a gordura é o alimento mais perigoso.*)

Uso de **figuras de estilo** menos frequentes.
Metáforas incomuns (ex.*: A sociedade está doente; ele é um lobo em pele de cordeiro*).

4.1.2. Descritores de complexidade

A partir da sistematização das dimensões textuais identificadas e dos níveis avançados, foram elaborados descritores de complexidade. Estes descritores visam traduzir, de forma operacional, as características linguísticas, conceptuais e discursivas associadas a cada dimensão, permitindo assim avaliar e classificar textos em função do seu grau de complexidade. A proposta de complexidade considera os seguintes aspetos como indutores de complexidade textual, que incluem a sua maior ou menor frequência no texto:

- Comprimento do texto (e.g., curto, longo)
- O propósito do texto (e.g., realização de tarefas quotidianas; lazer)
- O(s) conceito(s) mencionado(s) no texto (e.g., conceitos do dia a dia; noções especializadas),

- O(s) contexto(s) comunicativo(s) em que o texto circula (e.g., contextos do dia a dia, como bilhetes de transportes públicos; contextos mais específicos, como textos de opinião política)
- O léxico empregue no texto (i.e., relacionado com os conceitos e/ou com um domínio específico)
- O tipo de frase utilizadas no texto (e.g., frases com ordem não linear SVO)
- O tipo de formas e construções verbais no texto (e.g., tempos compostos, Modo Conjuntivo)
- A coesão textual (e.g., presença de conetores textuais e organização temporal)

Através da análise dos referenciais e de textos de cada nível, estes aspetos foram desdobrados e detalhados em descritores concretos (ver Anexo 1). Isto resultou na proposta de 81 descritores, contemplando os quatro níveis de complexidade (Monteiro et al., 2023). A noção de *descritores de complexidade textual* está relacionada com as manifestações concretas de complexidade no texto. O conjunto de descritores abrange dimensões textuais e permite uma análise da sua complexidade com maior granularidade.

No caso da dimensão léxico-conceptual, foram consideradas medidas tradicionais como a densidade lexical (i.e., frequência de palavras funcionais/gramaticais por palavras lexicais) e a relação entre complexidade lexical e a complexidade das ideias expressadas no texto (e.g., palavras atípicas – pouco frequentes – contribuem para tornar o texto mais abstrato (Fang & Pace, 2013). De igual forma, a linguagem figurativa também é considerada como indutora de complexidade, uma vez que sobrecarrega o processamento cognitivo do texto (Fang & Pace, 2013; Harley, 2013). A complexidade proveniente da linguagem figurativa está também relacionada com a concretude e grau de abstração da palavra: palavras que denotam conceitos mais abstratos são, geralmente, mais difíceis de processar do que palavras que denotam entidades concretas, especialmente para leitores de baixa literacia, uma vez que as primeiras não têm um referente concreto no mundo, mas as segundas têm (James, 1975). No entanto, em alguns casos, conceitos abstratos como *amor* ou *felicidade* contrariam este entendimento, possivelmente devido à familiaridade das/os leitoras/es com esses conceitos. Além disso, a frequência de conceitos abstratos nos textos é também um fator considerado (e.g., é o fator distintivo entre os descritores 21 e 22 ou a maior ou menor transparência e frequência de sentidos figurados, descritores 76 e 77).

Em relação à dimensão das estruturas sintáticas, esta proposta tem em conta medidas de *leiturabilidade* tradicionais, como a extensão da frase, responsáveis por induzir complexidade num texto, mas também manifestações de sintaxe complexa, de avaliação mais qualitativa, como estrutura não linear e ocorrências de subordinação e coordenação frequente e combinada. Para além disto, construções particulares, como frases passivas, são também consideradas como indutoras de complexidade. As construções na passiva tendem a manifestar-se em textos de caráter mais formal (Street & Dąbrowska, 2014). Por esta razão, pessoas mais proficientes nestes textos formais, isto é, com práticas de leitura mais desenvolvidas e referentes a uma maior diversidade de textos, têm mais facilidade em lidar com frases passivas do que indivíduos de baixa literacia.[8]

A flexão verbal rica de uma língua românica como o português conduziu ao entendimento de que esta dimensão teria de ser considerada independentemente. Assim, a complexidade verbal foi analisada à luz dos traços de tempo, modo e aspeto. Por exemplo, formas verbais no Infinitivo ou Presente do Indicativo são consideradas

[8] Para além da frequência, as razões por detrás de a passiva ser indutora de complexidade podem ser de natureza distintas: Harley (2013) afirma que, para a passiva ser processada cognitivamente, esta requer que a pessoa conceba, primeiramente, o equivalente na frase ativa, para depois converter a frase de volta para a passiva. Por outro lado, Ambridge et al. (2016) destacam as propriedades semânticas do verbo, nomeadamente os seus papéis temáticos (θ-*roles*), como facilitadores ou agravantes na compreensão da estrutura passiva, sendo verbos com os papéis temáticos de Agente-Verbo-Tema os mais fáceis em compreender em frases passivas, e verbos com os papéis temáticos envolvendo Tema-Experienciador, os mais difíceis.

como menos complexas, comparativamente com formas no modo Conjuntivo. De igual modo, formas verbais compostas menos frequentes são mais desafiantes, independentemente da frequência de uso ou densidade conceptual da unidade lexical (e.g. "Quando chegar a altura do teste, ele terá estudado bastante"). Com efeito, isto pode dever-se à frequência de uso de dada forma verbal ou da associação desta com um género textual específico (e.g., o pretérito mais-que-perfeito associado a uma linguagem literária). A juntar a isto, as formas verbais, consoante o seu tempo, modo e aspeto, podem criar espaços temporais ou mundos possíveis objetivamente mais complexos e difíceis de processar cognitivamente (Carreiras et al., 1997). Por fim, a criação de sequências temporais que aludem a mais do que uma localização temporal, como acontece várias vezes com a utilização de tempos verbais compostos, conduzem a uma dificuldade na conceptualização da situação em questão (e.g., "O João tinha comido o bolo quando o Rui entrou em casa"). Os verbos são, assim, tratados como manifestações linguísticas muito flexíveis a vários níveis e que, por esta razão, podem ser geradores de uma complexidade particular.

Nesta senda, também as perífrases verbais foram contempladas. Expressões como "*estar a* + Verbo no Infinitivo", "*começar a* + Verbo no Infinitivo", etc., são manifestações linguísticas de aspeto gramatical, que, embora possuam uma natureza composta, se revelam bastante acessíveis para todos os falantes. A forma como o tempo e aspeto são codificados nas expressões linguísticas, seja por flexão verbal, seja por meio de advérbios ou através de perífrase verbal, parece revelar uma hierarquia de complexidade que segue o entendimento da complexidade textual de Dahl (2007). Este considera a quantidade de informação linguística embutida numa dada manifestação linguística como reveladora da sua complexidade: quanto mais informação linguística existir (e.g., marcação de género gramatical, tempo, aspeto), maior será a complexidade do enunciado/da forma em questão. Assim, tendo em conta que é através da flexão verbal que mais informação linguística é concentrada numa única palavra, e que as perífrases verbais, por vezes, diluem esta informação linguística, podemos considerar que a flexão verbal revela mais complexidade do que certas perífrases verbais.

Por último, a dimensão de coesão textual tem em conta a coesão referencial, que se reflete no uso de anáforas, elipses, assim como no tipo de relação com o antecedente. Além disto, também são consideradas as manifestações espácio-temporais, no texto. As/Os leitoras/es constroem um modelo mental das entidades presentes num texto, bem como da informação espácio-temporal associada a estas. Estas informações condicionam o processamento do texto por parte do indivíduo: quando mais emaranhada a estruturação de informações, maior dificuldade existe para compreendê-la. Efetivamente, um aumento de localizações espácio-temporais levam a que a compreensão do texto seja afetada (Glenberg et al., 1987).

Com todos estes fatores em mente, a nossa proposta foca-se, principalmente, na aferição qualitativa da complexidade de um texto, através de descritores concretos associados a manifestações de complexidade específicas. Contudo, uma dimensão quantitativa também é tida em conta, nomeadamente no que respeita à extensão do texto.

4.2. **Resultados da validação pelo *focus group***

Como mencionado anteriormente, estas dimensões da complexidade textual foram desdobradas em 81 descritores concretos, agrupados em 12 categorias distintas: tamanho do texto; propósito; contexto comunicativo; diversidade conceptual; concretude; vocabulário/conjunto de léxico; morfologia; morfossintaxe; sintaxe; coesão textual: pronomes, anáforas e elipses; coesão textual: referências temporais; estilo. Estes descritores foram decorrentes da análise dos referenciais pela equipa do projeto e, posteriormente, validados por formadoras/es de EFA.

A participação e validação de profissionais de EFA contribuiu para um melhor entendimento de uma característica específica destes descritores: a sua frequência de ocorrência. Ao contrário da suposição de que uma

frequência mais alta de ocorrências de fenómenos indutores de complexidade aumenta a complexidade num texto, as reuniões com as/os formadoras/es EFA conduziram-nos à conclusão de que isto não era universalmente verdadeiro. Algumas características foram consideradas como indutoras da mesma complexidade, independentemente da sua frequência de ocorrência num texto. Ou seja, foi considerado que o nível de complexidade associado a estes traços seria o mesmo, independentemente do número de ocorrências no texto (mais/menos frequência não equivale a mais ou menos complexidade desse traço). Consequentemente, a nossa proposta inicial foi revista, de modo a privilegiar descrições centradas em características linguísticas observáveis, desligadas da frequência de ocorrência dos fenómenos, ainda que esse aspeto seja considerado posteriormente na aplicação prática do referencial em ferramentas de avaliação de complexidade textual. Porém, também foi reconhecido que existem exceções a este entendimento, podendo manifestar-se um *continuum* de frequência os seguintes descritores:

- A frequência do grau de concretude das palavras (descritores 20, 21)
- A frequência dos estrangeirismos presentes no texto (descritores 35, 36, 37).
- A frequência de orações relativas (descritores 53, 54, 55, 56)
- A frequência da presença frases passivas (descritores 57, 58, 59)

Os descritores mencionados acima enquadram-se no campo léxico-conceptual (os dois primeiros) e sintático (os dois últimos). A continuidade dos dois primeiros descritores segue em linha com a importância do conhecimento de 95% a 98% de palavras no texto, por parte do/a leitor/a (Schmitt et al., 2011). Ou seja, aquando da consideração de complexidade lexical, existe uma margem – ainda que curta – para considerar a frequência de um dado fenómeno linguístico – neste caso, certas palavras. Por sua vez, os outros dois descritores, de natureza sintática, enquadram-se numa avaliação de *leiturabilidade* (Dale & Chall, 1948; Flesch, 1948; Gunning, 1952; Kincaid et al., 1975; Lorge, 1948; McLaughlin, 1969; Spache, 1953), cuja medida de aferição é o número de ocorrências presentes no texto.

O referencial de complexidade para adultos de baixa literacia aqui proposto é integralmente apresentado no Anexo 1 e textos exemplificativos de cada nível, provenientes do *corpus* iRead4Skills,[9] estão presentes no Anexo 2.

5. Conclusão

Este trabalho apresentou uma proposta de referencial de complexidade textual para adultos, falantes nativos de baixa literacia, desenvolvida no âmbito do projeto iRead4Skills. Considera-se que os resultados e a metodologia aqui detalhados contribuem para uma melhor compreensão do fenómeno da complexidade textual, propondo uma abordagem multifacetada que combina dimensões quantitativas e qualitativas com a variabilidade de competências literácitas de um público-alvo específico. Reconhecendo o fenómeno da complexidade como algo que diverge de indivíduo para indivíduo, a nossa proposta reconhece o *continuum* de competências em literacia e as diferentes posições dos leitores, através do estabelecimento de três níveis: Muito Fácil, Fácil e Claro.

A contribuição da nossa proposta assenta no estabelecimento de características linguísticas e de coesão textual num sistema de classificação da proficiência em leitura que visa adultos falantes nativos. Para além disto, a

[9] Os textos exemplificativos são provenientes do *corpus* iRead4Skills. Todo o *corpus* está disponível, assim como outros recursos produzidos no projeto, em https://zenodo.org/communities/iread4skills/records?q=&l=list&p=1&s=10&sort=newest .

validação da proposta por profissionais de EFA – bem como a integração do seu conhecimento e experiência empírica – é um fator distintivo deste novo referencial, nomeadamente no que diz respeito à sua potencial aplicação em contextos reais de sala de aula. Em termos práticos, e no contexto do projeto iRead4Skills, a nossa proposta de complexidade contribui para o desenvolvimento de ferramentas inovadoras, tais como um sistema inteligente para a avaliação da complexidade de textos e uma ferramenta de adaptação de textos para níveis de complexidade adequados a públicos específicos. Contribui, também, para auxiliar profissionais da área a melhor compreender e avaliar competências e dificuldades de leitura das/os suas/seus formandas/os. Em termos teóricos, a natureza objetiva e a caracterização linguística detalhada dos descritores propostos contribuem para o avanço de conhecimento na área da complexidade textual e da literacia, permitindo também estudos comparativos entre diferentes populações-alvo.

No decorrer da elaboração deste trabalho, foram divulgados os resultados do Inquérito às Competências dos Adultos de 2023, desenvolvido pelo PIAAC, que colocam Portugal abaixo da média da OCDE no que diz respeito aos níveis de proficiência em literacia, no penúltimo lugar entre os países participantes (OECD, 2024a). Além disso, é importante realçar que as descrições dos níveis de proficiência presentes no guia-referência do PIAAC foram ligeiramente alteradas relativamente ao documento anterior (OECD, 2013b). Estas novas descrições são mais extensas e incluem elementos descritivos mais relacionados com o conteúdo dos textos. Por exemplo, no primeiro nível de proficiência (abaixo do Nível 1), há agora referências explícitas à consistência do uso de palavras e à forma como esta consistência contribui para o significado global da frase. Ao nível da frase, a adjacência dos conceitos e a forma como se relacionam entre si são também consideradas (OCDE, 2024b). Estas revisões, que enfatizam a coerência interna e a conetividade semântica dos textos – em contraste com as anteriores, que se centravam mais nas tarefas que se esperava que os indivíduos realizassem em cada nível e na forma como a informação era apresentada no texto como um todo –, alinham-se com a abordagem e as conclusões da nossa proposta. No entanto, os níveis e descritores propostos neste trabalho são significativamente mais pormenorizados, oferecendo uma caraterização mais rica e mais fina da complexidade do texto. Isto assegura a nossa proposta como uma ferramenta valiosa e altamente relevante para responder às necessidades específicas de adultos com baixa literacia.

Em suma, o referencial de complexidade textual iRead4Skills representa um contributo relevante para o estudo sobre os desafios de leitura em adultos de baixa literacia, estabelecendo uma ponte entre a investigação teórica e a sua aplicação. O melhor entendimento dos fenómenos da complexidade textual e o seu condicionamento em adultos de baixa literacia proveniente deste trabalho espera-se benéfico profissionais da educação e investigação, bem como em lugares de decisão política, que ambicionam promover o desenvolvimento das competências de literacia e potenciar as competências de leitura das pessoas, com retorno para a sociedade.

Agradecimentos

Este trabalho foi financiado pela Comissão Europeia (projeto iRead4Skills, número 1010094837, HORIZONCL2- 2022-TRANSFORMATIONS-01-07). Os pontos de vista e opiniões expressos neste trabalho são da exclusiva responsabilidade dos/as autores/as e não refletem necessariamente os da União Europeia ou da Agência de Execução Europeia da Investigação. Nem a União Europeia nem a autoridade que concedeu o financiamento podem ser responsabilizadas pelos mesmos. Parte deste trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e Tecnologia, I.P., no âmbito do projeto UID/03213 – Centro de Linguística da Universidade NOVA de Lisboa (CLUNL).

Referências

ALTE (2002). *The ALTE Can Do Project (1992–2002)*. https://www.alte.org/can_do/alte_cando.pdf .

Alves, J. (2005). Literacia e modernização económica. In H. Moura (Org.), *Diálogos com a literacia* (pp. 15–22). Lisboa Editora.

Alves, A. C. & Arqueiro, A. S. & Dionísio, M. (2012). Vidas letradas de adultos em processo de Reconhecimento, Validação e Certificação de Competências. In I. Pereira, I. Dias, H. Pinto, H. Menino & R. Cadima (Orgs.), I *Conferência Internacional — investigação, práticas e contextos em Educação* (pp. 123–129). Escola Superior de Educação e Ciências Sociais – Instituto Politécnico de Leiria.

Ambridge, B., Bidgood, A., Pine, J. M., Rowland, C. F., & Freudenthal, D. (2016). Is passive syntax semantically constrained? Evidence from adult grammaticality judgment and comprehension studies. *Cognitive Science*, *40*(6), 1435–1459. https://doi.org/10.1111/cogs.12277

Anderson, J.R. (2000). *Cognitive psychology and its implications*. W.H. Freeman.

Beck, I. L., McKeown, M. G., Omanson R., & Pople, M. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly*, *19*, 263–277.

Carreiras, M., Carriedo, N., Alonso, M., & Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Memory & Cognition*, 25(4), 438–446. https://doi.org/10.3758/bf03201120

Correia, S., Amaro, R., & Monteiro, R. (2024). *iRead4skills – reading skills survey* [Relatório]. Zenodo. https://doi.org/10.5281/ZENODO.10179536

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. https://www.coe.int/lang-cefr .

Cunha, J. & Dionísio, M. (2019). O lugar da literacia nas ofertas de educação e formação de jovens em Portugal. In T. Dantas, M. Dionísio & M. Loffin (Orgs.), *Educação de jovens e adultos: Políticas, direitos, formação e emancipação social* (pp. 195–212). Editora da Universidade Federal da Bahia (EDUFBA).

Cunningham, J. W., & Mesmer, H. A. (2014). Quantitative measurement of text difficulty: What's the use? *Elementary School Journal*, 115, 255–269.

Dahl, Ö. (2007). Definitions of complexity. *Proceedings of the Colloquium on complexity, accuracy and fluency in second language use, learning & teaching,* 41-46.

Dale, E. & Chall, J. (1948). A formula for predicting readability. *Educational Research Bulletin*, *27*(1). 11–20, 37–54.

EDUSTAT – Fundação Belmiro de Azevedo. (2024). *Distribuição de adultos matriculados nos diferentes tipos de curso orientados para formação de adultos*. https://www.edustat.pt/indicador?id=1

Fang, Z., & Pace, B. G. (2013). Teaching with challenging texts in the disciplines. *Journal of Adolescent & Adult Literacy, 57*(2). 104–108.

Flor, M., Klebanov, B., & Sheehan, K. (2013). Lexical tightness and text complexity. In *Proceedings of the Workshop on Natural Language Processing for Improving Textual Accessibility* (pp. 29–38). Association for Computational Linguistics. https://aclanthology.org/W13-1504/

Fransman, J. (2005). *Understanding literacy: A concept paper. Background paper for EFA global monitoring report*. UNESCO.

Glenberg, A.M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*.

Goldman, S.R., & Lee, C.D. (2014). Text complexity. State of the art and the conundrums it raises. *Elementary School Journal*, *115*, 290–300.

Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
Hanemann, U. (2016). Shaping the literacy agenda from a lifelong learning perspective. *Adult Education and Development, 83*, 14–17.
Harley, T. A. (2013). *Psychology of language: From data to theory*. Taylor and Francis.
Heilman, M., Collins-Thompson, K., Callan, J., & Eskenazi, M. (2007). Combining lexical and grammatical features to improve readability measures for first and second language texts. *Proceedings of the NAACL Human Language Technology Conference*, 460–467.
Hiebert, E.H. (2011). *Using multiple sources of information in establishing text complexity* [Reading Research Report #11.03]. TextProject, Inc.
Hiebert, E. H., & Mesmer, H. A. E. (2013). Upping the ante of text complexity in the Common Core State Standards: Examining its potential impact on young readers. *Educational Researcher*, *42*(1), 44–51. https://doi.org/10.3102/0013189x12459802
Hu, M., & Nation, I. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430.
Kincaid, J. P., Fishburne, R. P. J., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas Automated Readability Index, Fog Count and Flesch Reading Ease Formula for Navy enlisted personnel* [Research Branch Report 8-75]. Naval Technical Training.
Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, *54*(4), 1076–1085. https://doi.org/10.1002/tesq.3004
Lorge, I. (1948). The Lorge and Flesch readability formulae: A correction. *School and Society*, *67*, 141–142.
Mclaughlin, G. H. (1969). SMOG grading — A new readability formula. *Journal of Reading*, *128*, 639–646.
Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading Research Quarterly*, *47*, 235–258. https://doi:10.1002/rrq.019.
Monteiro, R., Amaro, R., Correia, S., Pintard, A., Gauchola, R., Moutinho, M., & Blanco Escoda, X. (2023). *iRead4Skills - Complexity Levels* (V1.0). Zenodo. https://doi.org/10.5281/zenodo.10459090
OECD (2013a). *Technical report of the survey of adult skills (PIAAC)*. OECD Publishing.
OECD. (2013b). *The survey of adult skills reader's companion*. OECD Publishing.
OECD (2021). *The assessment frameworks for cycle 2 of the programme for the International Assessment of Adult Competencies*. OECD Publishing. https://doi.org/10.1787/4bc2342d-en
OECD (2024a). *Do adults have the skills they need to thrive in a changing world survey of adult skills 2023.* OECD Publishing. https://doi.org/10.1787/b263dc5d-en
OECD (2024b). *Survey of adult skills – Reader's companion: 2023*. OECD Publishing. https://doi.org/10.1787/3639d1e2-en
Pintard, A., François, T., Nagant de Deuxchaisnes, J., Barbosa, S., Reis, M. L., Moutinho, M., Monteiro, R., Amaro, R., Correia, S., Rodríguez Rey, S., Garcia González, M., Mu, K., & Blanco Escoda, X. (2024). *iRead4Skills Dataset 1: corpora by complexity level for FR, PT and SP* (2.1.) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.13768477
Reis, M., Barbosa, S., Moutinho, M., Monteiro, R., Correia, S., & Amaro, R. (2024). Intelligent support for low literacy adults: The European Portuguese iRead4Skills corpus. *International Journal of Emerging Technologies in Learning (iJET)*, *19*(08), 61–81. https://doi.org/10.3991/ijet.v19i08.52023
Rothes, L & Queirós, J. (2021). Práticas de leitura e competências de literacia: Resultados e ilações do PIAAC. *Entreler*, *1*, 32–42.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Harvard University Press.

Spache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal*, *537*, 410–413.

Street, B. (1993). Introduction: The new literacy studies. In B. Street (Ed.), *Cross-cultural approaches to literacy* (pp. 1–21). Cambridge University Press.

Street, J. A., & Dąbrowska, E. (2014). Lexically specific knowledge and individual differences in adult native speakers' processing of the English passive. *Applied Psycholinguistics*, *35*(1), 97–118. https://doi.org/10.1017/S0142716412000367

Stromquist, N. P. (2006). Women's rights to adult education as a means to citizenship. *International Journal of Educational Development, 26*(2), 140–152. https://doi.org/10.1016/j.ijedudev.2005.07.02

UNESCO Institute of Statistics. (2020, 22 junho). *Literacy*. https://uis.unesco.org/node/3079547

Valente, A. (Coord.) (2021). *Referencial de Competências-Chave de Educação e Formação de Adultos – Nível Básico.* ANQEP.

Wagner, D. (2011). What happened to literacy? Historical and conceptual perspectives on literacy in UNESCO. *International Journal of Educational Development*, *42*(3), 319–323.

Anexo 1 – Descritores por nível de complexidade

		Descritores	Muito fácil	Fácil	Claro	+ complexo
tamanho	1	O texto tem até 50 palavras.				
	2	O texto tem até 150 palavras.				
	3	O texto tem até 250 palavras.				
	4	O texto tem até 500 palavras.				

		Descritores	Muito fácil	Fácil	Claro	+ complexo
propósito	5	O texto serve como apoio para desempenhar tarefas familiares/quotidianas (e.g., comprar bilhetes de autocarro, identificar tipo e preço de produtos).				
	6	O texto apresenta informação nova (e.g., uma introdução sobre a importância da reciclagem como *A reciclagem é quando transformamos o lixo em coisas novas e úteis*.).				
	7	O texto tem o propósito de entreter o leitor em momentos de lazer (e.g., desporto, crónica, notícias).				
	8	O texto serve para aprender/realizar tarefas novas (e.g., instruções sobre como utilizar o e-mail da empresa.				
	9	O propósito do texto pode ser sobre vários tópicos de interesse do leitor, quer seja para o informar ou para lazer (e.g., artigo de revista sobre natureza).				

		Descritores	Muito fácil	Fácil	Claro	+ complexo
contexto comunicativo	10	O texto circula em comunicações do dia a dia (e.g., horários de transportes, ementas, instruções, preços de produtos, comunicações pessoais, etc.).				
	11	O texto é utilizado em contextos profissionais ou de comunicação dos *media* (e.g., comunicação ou instrução profissional, notícias, anúncios, desporto).				
	12	O texto é utilizado em contextos de lazer (e.g., literatura e ficção, reportagens, artigos de opinião).				
	13	O texto circula em contextos artísticos (e.g., literatura, poesia).				

	14	O texto é empregue em contextos especializados (e.g., comunicação profissional formal, fóruns *online*).			
	15	O texto é utilizado em contextos científicos e académicos (e.g., ensaios, relatórios, artigos científicos).			

		Descritores	**Muito fácil**	**Fácil**	**Claro**	**+ complexo**
diversidade conceptual	16	O texto menciona conceitos relacionados com atividades do dia a dia (e.g., comida, lida da casa, roupa).				
	17	O texto menciona conceitos presentes em âmbitos específicos, mas que são de interesse do leitor (e.g., profissional, desporto, hobbies).				
	18	O texto refere conceitos que os leitores precisam de interligar para compreender totalmente (e.g., estrelas, galáxias, vulcões).				
	19	O texto faz referência a conceitos novos e multifacetados (e.g., matéria negra, desenvolvimento do cérebro, estatísticas).				

		Descritores	**Muito fácil**	**Fácil**	**Claro**	**+ complexo**
concretude	20	O texto apresenta mais palavras que referem objetos concretos (e.g., *cadeira, maçã, martelo, computador*).				
	21	O texto contém poucas palavras que referem ideias abstratas (e.g., *felicidade, imaginação, amor*).				
	22	O texto emprega palavras que remetem para conceitos e ideias abstratas (e.g., *envelhecimento, razão, fidelidade*).				
	23	O texto usa palavras que referem conceitos e ideias abstratas em contextos multifacetados e/ou âmbitos especializados (e.g., *poesia, democracia, fusão nuclear*).				

		Descritores	**Muito fácil**	**Fácil**	**Claro**	**+ complexo**
vocabulário/ conjunto de léxico	24	O texto não emprega abreviações.				
	25	O texto utiliza abreviações comuns após ter explicado o que significam (e.g., *l – litro, m – metro*).				
	26	O texto utiliza abreviações comuns (e.g., *m, km, g., l, Exmo.*).				

27	O texto não utiliza siglas.					
28	O texto apenas utiliza siglas e acrónimos conhecidos (e.g., *CP, IVA, RTP, SIC, TVI*).					
29	O texto utiliza siglas e acrónimos não conhecidos juntamente com o nome a que se referem (e.g., *Organização das Nações Unidas (ONU), Instituto Público (IP)*).					
30	O texto emprega todos os tipos de siglas (e.g., *ADN, UNESCO, EU, UNICEF*).					
31	O texto utiliza palavras e expressões simples relacionadas com assuntos do quotidiano, fazendo uso de substantivos frequentes (e.g., *batatas, azeite, legumes*).					
32	O texto tem palavras e expressões frequentes que fazem referência a assuntos do dia a dia, mas também a outros âmbitos específicos ou profissionais (e.g., *segunda-feira, curso, ano escolar*).					
33	O texto tem palavras e expressões um pouco menos conhecidas e frequentes que fazem referência a âmbitos específicos (e.g., *doença cardíaca, artéria, tratamento*).					
34	O texto utiliza palavras e expressões de âmbitos específicos, jargão ou termos arcaicos (e.g., *enfarte do miocárdio, artéria coronária, dissecção espontânea, genoma*).					
35	O texto não faz uso de estrangeirismos ou faz uso de estrangeirismos muito frequentes e conhecidos (e.g., *internet, donut*).					
36	O texto tem poucos estrangeirismos comuns (e.g., *site, stress*).					
37	O texto faz uso de estrangeirismos comuns frequentemente (e.g. *link, hobby, marketing, mail, gay*).					
38	O texto contém palavras polissémicas (palavras que possuem mais do que um sentido) que são ambíguas e/ou pouco frequentes. (e.g., *interno, representar, forma, paciente*).					
39	O texto contém palavras polissémicas (palavras que possuem mais do que um significado) ambíguas e/ou bastante frequentes (e.g., *base, experiência, leve, seca, aplicar, banco*).					

		Descritores	Muito fácil	Fácil	Claro	+ complexo
morfologia	40	O texto utiliza palavras compostas comuns (e.g., *feijão verde, couve-flor*) ou palavras compostas transparentes, ou seja, palavras ou expressões que podem ser facilmente compreendidas através da combinação das duas partes (e.g., *tira-caricas*).				
	41	O texto utiliza palavras compostas comuns e o seu sentido pode ser percebido pela combinação das duas partes (e.g., *óculos de sol, mal-educado, meia-idade, maus-tratos*).				
	42	O texto tem palavras compostas cujo sentido não é facilmente compreendido através da combinação das duas partes (e.g., *meio-termo, porta-voz, obra-prima*).				
	43	O texto possui várias palavras compostas com sentidos que não são facilmente compreendidos pela combinação das duas partes (e.g., *pós-graduação, mais-valia, pé-de-meia, cavalo de batalha*).				
	44	O texto tem palavras derivadas com prefixos e sufixos comuns e familiares em grande número (e.g., *infeliz, calmamente, organizar, trabalhador, casinha*).				
	45	O texto tem algumas palavras derivadas mais comuns, com prefixos e sufixos frequentes (e.g. *pré-cozinhado, desorganizar, embalagem, organização, jardineiro, finíssimo*).				
	46	O texto utiliza palavras derivadas com sufixos e prefixos variados (e.g., *nervosismo, lavável, curiosidade, altitude, superpoder, preconceituoso*).				
	47	O texto utiliza palavras derivadas com sufixos e prefixos menos comuns (e.g., *megatonelada, tendinite, geopolítico, neoliberal*).				

		Descritores	Muito fácil	Fácil	Claro	+ complexo
morfossintaxe	48	<u>Modos e Tempos verbais</u>				
		Verbos no infinitivo (e.g., *gostar, comer, fugir*).				
		Verbos no Presente do Indicativo (e.g., *gosto, comem, foge*).				
		Verbos no Pretérito Imperfeito do Indicativo (e.g., *gostavam, comia, fugias*).				
		Verbos no Pretérito Perfeito do Indicativo (e.g., *gostou, comi, fugiram*).				

	Verbos no Pretérito Mais-que-perfeito do Indicativo (e.g., *gostara, comeras, fugira*)			■
	Verbos no Futuro Simples do Indicativo (e.g., *esquecerá, gostará*).		■	
	Verbos no Gerúndio (e.g., *gostando, comendo, fugindo*).		■	
	Verbos no Imperativo (e.g., *come!, falai!; não comas!; não falem!)*.	■		
	Verbos no Condicional (e.g., *gostaria, comerias, fugiriam*).		■	
	Verbos no Presente do Conjuntivo (e.g., *gosta, coma, fujam*)			■
	Verbos no Pretérito Imperfeito do Conjuntivo (e.g., *gostasses, comessem, fugisse*).		■	
	Verbos no Futuro do Conjuntivo (e.g., *se eu gostar, se tu comeres, se eles fugirem*).			■
	Verbos no Pretérito Perfeito Composto do Indicativo (e.g., *tenho gostado, tem comido, têm fugido*).			■
	Verbos no Pretérito Mais-que-perfeito Composto do Indicativo (e.g., *tinhas gostado, tinham comido, tinha fugido*).			■
	Verbos no Futuro Composto do Indicativo (e.g., *terei gostado, terás comido, terão fugido*).			■
	Verbos no Pretérito Perfeito Composto do Conjuntivo (e.g., *tenha gostado, tenha comido, tenham fugido*).			■
	Verbos no Pretérito Mais-que-perfeito Composto do Conjuntivo (e.g., *tivesses gostado, tivessem comido, tivesse fugido*).			■
	Verbos no Futuro Composto do Conjuntivo (e.g., *tiverem gostado, tiver comido, tiveres fugido*).			■
	Verbos auxiliares *poder, ir, estar* no Indicativo (e.g., *posso gostar, vai comer, estão a fugir*).		■	
	Expressões com os verbos *permanecer, querer, poder, conseguir* em tempos simples (e.g., *permaneça calmo, queria almoçar, consigo ir*).		■	
	Expressões não muito extensas com os verbos estar, ser, permanecer, querer, poder e dever, em tempos compostos (e.g., *tenham estado ligados; devíamos ter acabado; temos de permanecer calmos; devíamos ir*).			■

		Descritores	Muito fácil	Fácil	Claro	+ complexo
sintaxe	49	No texto, há mais frases curtas e diretas ou combinadas através de conjunções familiares (e.g., *A criança fugiu ontem. O pai trabalha na fábrica, mas a mãe fica em casa.*).				
	50	O texto combina frases através de conjunções coordenadas e subordinadas (e.g., *Os músicos estavam tocando, mas poucos clientes ouviam. Quando pararam de tocar, apenas um senhor se manifestou*.).				
	51	As frases no texto são frequentemente combinadas através de conjunções coordenadas e subordinadas e, por vezes, com conjunções menos comuns (e.g., *Apesar da confusão provocada pelos animais à solta, a situação foi controlada sem os animais serem atropelados*.).				
	52	As frases no texto são longas e compostas por diversas outras frases combinadas através de conjunções coordenadas e subordinadas (e.g., *Dado que grande parte deste livro fala na gestão das partes difíceis do trabalho a solo e analisa modos de as corrigir ou de as transcender, primeiro quero deixar bem claro que trabalhar sozinho é fantástico: é por isso que tantas pessoas o consideram uma opção*.).				
	53	O texto tem poucas orações relativas referentes ao sujeito da frase (e.g., *O senhor que vive na casa ao lado ouviu o barulho*).				
	54	O texto tem orações relativas que referem o sujeito da frase (e.g., *O senhor que vive na casa ao lado ouviu o barulho*.).				
	55	As frases no texto incluem frequentemente orações relativas correspondentes ao sujeito da frase (e.g., *Francisca, a quem os amigos tratam por «Kika», começou a surfar muito pequena.*).				
	56	As frases no texto incluem frequentemente orações relativas correspondentes ao sujeito ou ao objeto da frase (e.g., *O rapaz que abraçou a mãe dirigiu-se à entrada; O rapaz que a mãe abraçou dirigiu-se à entrada*.).				
	57	No texto, há mais frases ativas (e.g., *O cão comeu o bolo* (ativa) vs. *O bolo foi comido pelo cão*. (passiva)).				
	58	O texto tem poucas frases passivas (e.g., *A visita foi paga pelo patrão.*).				
	59	O texto apresenta frases passivas (e.g., *O almoço foi pago pela empresa*), incluindo construções impessoais comuns e passivas de -se (e.g., *O trabalho faz-se bem*.).				
	60	O texto tem frases simples apresentadas numa ordem básica e linear (sujeito-verbo-complementos) (e.g., *Eu ajudei os meus filhos a comprar casa.*).				

	61	O texto contém poucas frases simples apresentadas numa ordem não linear (sujeito-verbo-complementos) (e.g., *À escola, chegaram os autocarros da visita de estudo.*).			
	62	As frases do texto não estão necessariamente numa ordem básica e linear (sujeito-verbo-complementos) (e.g., *De ar puro, todos precisamos*.).			
	63	As frases no texto são apresentadas em diversas ordens não lineares (sujeito-verbo-complementos; verbo-complementos-sujeito, etc.) (e.g., *Até final de março, porque antes seria impossível encontrar todos os ingredientes, haverá na ementa vários pratos de caça*).			

		Descritores	**Muito fácil**	**Fácil**	**Claro**	**+ complexo**
coesão textual: pronomes, anáforas e elipses	64	O texto tem nenhum ou poucos elementos omitidos (e.g., *O acidente causou dois feridos leves. Os bombeiros levaram os feridos para o hospital.*).				
	65	O texto tem elementos repetidos em vez de pronomes (e.g., *O João não gosta da Maria porque a Maria não é simpática.*).				
	66	O texto tem alguns elementos omitidos (e.g., *Gostava de fazer artesanato*. vs. *Eu gostava de fazer artesanato.*).				
	67	O texto usa pronomes (e.g., *O João não gosta da Maria porque ela não é simpática.*).				
	68	O texto apresenta elementos omitidos (e.g., *[] Mostrámos em estudos anteriores que a gordura solidifica quando [] arrefece*. vs. *Nós mostrámos em estudos anteriores que a gordura solidifica quando ela arrefece.*).				
	69	O texto apresenta pronomes e palavras relacionadas para indicar o que é referido (e.g., *No livro, nós damos uma definição completa de uma tempestade de neve e como esta se forma. Também indicamos onde o fenómeno pode ser visto.*).				

		Descritores	Muito fácil	Fácil	Claro	+ complexo
estilo	74	O texto não faz uso de sentidos figurativos (e.g., *Este verão os preços subiram* vs. *Este verão os preços dispararam.*).				
	75	O texto apresenta sentidos figurativos comuns (e.g., *Este verão os preços dispararam.*).				
	76	O texto apresenta sentidos figurativos menos comuns (e.g., *O Barcelona varreu o Real Madrid*; *O Estoril esmagou o Santa Clara*.).				
	77	O texto apresenta sentidos figurativos incomuns (e.g., *Os primeiros raios de sol aconchegaram-me as faces.*).				
	78	O texto utiliza apenas comparações, personificações e hipérboles mais comuns (e.g., *Ela é linda como uma flor; As minhas costas estão a dar cabo de mim; A tua mala pesa uma tonelada.*).				
	79	O texto faz uso da metonímia (e.g., *beber um copo*), assim como metáforas familiares (e.g., *Meu pai tem olhos de águia.*).				
	80	O texto faz uso de metáforas incomuns (e.g., *A sociedade está doente; ele é um lobo em pele de cordeiro.*).				
	81	O texto apresenta diversas figuras de estilo (e.g., *O cão da fábrica de curtumes acendia latidos fosforescentes nas noites de julho, quando o pólen da acácia me cobria as pálpebras, eu, morto de amores pela mulher de Sandokan, descobria-me unicórnio trancado na retrete da escola*.).				

Anexo 2 – Textos exemplificativos por nível de complexidade

Exemplo de texto do nível Muito Fácil

Irlanda, 16 de novembro de 2014

Cara Ana,
Muito obrigado pelas fotos de Inglaterra que me mandaste! São muito bonitas.
Na Irlanda é tudo muito bonito. As pessoas cá são simpáticas, e já sei palavras e expressões novas que desconhecia. Este ano passamos bons momentos juntas. Agora seguimos caminhos diferentes. Vou mandar-te umas lembranças de cá como uns trevos e uns duendes (em peluche). Tens de vir ver a minha nova casa. Nas férias passa por cá e ficas aqui uma se maninha ou duas.
Dá notícias.

Um beijinho,
Bruna

PS. - Um dia destes vou aí visitar-te. Manda beijinhos ao Harry.

In
MAGMA studio. (2022). https://magmastudio.pt/

Exemplo de texto do nível Fácil

Pelas Florestas de Portugal

O verão está a chegar e o calor já começa a apertar. Felizmente, vivemos em Portugal, um país repleto de paisagens que concilia água e floresta e ainda nos oferece umas boas idas à praia. Estás pronto para entrar nesta aventura e conhecer melhor este nosso retângulo?
Vira a página!

In
Pelas florestas de Portugal. (2021). *Dá a Mão à Floresta*, *42*, 3.

Exemplo de texto do nível Claro

Sentei-me no muro a comer uma bola de Berlim, uns instantes antes de caírem umas gotas grossas de chuva. As pessoas na praia começaram a arrumar as coisas apressadamente. Fechei os olhos por um instante, uma ou duas gotas caíram-me sobre a testa. O cheiro da areia molhada é diferente do petricor, aquele nome que dão ao cheiro da terra molhada. É um cheiro quente, metálico, de maresia misturada com pó. A propósito de nada lembrei-me daquela telenovela dos anos oitenta, baseada no romance do Sttau Monteiro, a "Chuva na Areia". Estendi a mão para que me caísse um pingo ou dois na palma, as pessoas corriam na areia, naquele movimento um bocado ridículo, em câmara lenta e de pernas presas na areia solta, alguns com um saco ou toalha na cabeça, a dizerem para os miúdos correrem. Continuei sentada no muro, limpei o açúcar do canto da boca, com o dedo indicador, que me esqueci de levar lenços e enquanto estendia de novo a palma da mão, muitos dos carros arrancaram, ficaram poucos mais as carrinhas dos surfistas, a praia esvaziou-se, o silêncio alargou o horizonte sobre o mar, e eu, sentada no muro, depois de uma semana cansativa, sem me chamar Guida nem agarrar mais uma vez o Verão, finalmente respirei.

In
Soares, C. N. (2023, 25 agosto). Agarra o Verão, Cristina. *Já disse que não quero nada*. https://naoqueronada.blogs.sapo.pt/agarra-o-verao-cristina-2323

Interfaces na construção do conhecimento fonológico – a lateral em Coda silábica

Interfaces in the construction of phonological knowledge – the lateral in syllabic Coda

Ana Ruth Moresco Miranda[1], *Carmen Lúcia Barreto Matzenauer*[1]

[1]Universidade Federal de Pelotas

Abstract

The article deals with the process of language acquisition by Brazilian children, focusing on the construction of phonological knowledge of segmental units of the language, specifically the lateral liquid /l/. The discussion includes an analysis of the internal structure of the segment, the syllabic constitution of the Coda, the morphophonological behavior of the segment, and the phonology/orthography relationship established in the process of writing acquisition. The Bidirectional Parallel Model of Phonology and Phonetics (BiPhon), which is configured as a Model of Grammar and L1 Processing (Boersma, 2007, 2011; Boersma & Hamann, 2009), supports the formalization of input conditioning in the construction of abstract categories of phonological grammar. The possibility of change in representations in a continuous process of phonological knowledge development is also considered on the basis of BiPhon, enriched by the incorporation of the <Written Form> as input for the formation of phonological categories, as proposed by Hamann and Colombo (2017).

Keywords: language acquisition, phonological knowledge, lateral liquid, syllabic Coda, Brazilian Portuguese

Resumo

O artigo trata do processo de aquisição da linguagem por crianças brasileiras, focalizando a construção do conhecimento fonológico de unidades segmentais da língua, especificamente da líquida lateral /l/. A discussão contempla a análise da estrutura interna do segmento, a constituição silábica da Coda, o comportamento morfofonológico do segmento e a relação fonologia/ortografia estabelecida no processo de aquisição da escrita. O Modelo Bidirecional Paralelo de Fonologia e Fonética (BiPhon), que se configura como Modelo de Gramática e de Processamento de L1 (Boersma, 2007, 2011; Boersma & Hamann, 2009), dá suporte à formalização do condicionamento do input na construção de categorias abstratas da gramática fonológica. A possibilidade de mudança nas representações em um processo continuado de desenvolvimento do conhecimento fonológico é considerada também com base no BiPhon, enriquecido pela incorporação da <Forma Escrita> (<*Written Form*>) como input para a formação de categorias fonológicas, conforme proposta de Hamann e Colombo (2017).

Palavras-chave: aquisição da linguagem, conhecimento fonológico, líquida lateral, Coda silábica, português brasileiro

1. Introdução

Este artigo focaliza a construção do conhecimento fonológico de unidades segmentais da língua, especificamente a líquida lateral /l/, no processo de aquisição da linguagem por crianças brasileiras. O exame desse processo de formação da representação fonológica se desenvolve a partir de quatro eixos: (a) das unidades que constituem a estrutura interna do segmento, (b) dos constituintes silábicos para os quais o segmento é licenciado, com atenção especial para a Coda, (c) do comportamento morfofonológico do segmento na

derivação de palavras da língua, (d) da relação fonologia/ortografia estabelecida no processo de aquisição da escrita.

Para chegar ao funcionamento da língua alvo, a aquisição gradual da gramática fonológica pela criança exige não apenas o conhecimento concernente a cada um dos quatro eixos antes referidos, mas das interações entre eles. Essas interfaces evidenciam a complexidade da construção de representação de unidades da gramática e também a possibilidade de sua reestruturação em um processo continuado de desenvolvimento do conhecimento fonológico.

Começa-se com a discussão sobre a interface constituída pelos eixos (a) e (b) e com a observação do comportamento da líquida lateral /l/ junto às outras consoantes líquidas do Português. Após, é feita referência ao comportamento morfofonológico da lateral em Coda, eixo (c), para que então sejam apresentados resultados referentes a dados de aquisição da linguagem, da fala e da escrita, os quais dão subsídios à discussão sobre a construção das representações fonológicas, à luz do modelo de gramática proposto por Boersma (2007, 2011).

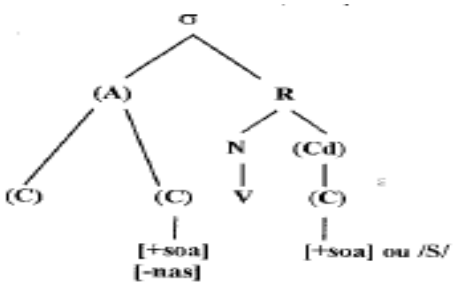
2. O funcionamento das líquidas do português e seu licenciamento prosódico

A fonologia do Português integra quatro consoantes líquidas: duas laterais (/l/ e / ʎ/) e duas róticas (/ɾ/ e /ʀ/). O licenciamento desses segmentos está diretamente vinculado, na gramática da língua, à estrutura interna da sílaba.

Na Figura 1 apresenta-se o molde silábico proposto por Bisol (1999, p.703) para o Português.

Figura 1

Molde Silábico do Português



Nota. Fonte: Bisol (1999, p.703)

Levando em conta a estrutura na Figura 1, tem-se na Tabela 1 a explicitação do licenciamento das líquidas do Português nos constituintes silábicos marginais – Ataque (Onset) e Coda – em relação à palavra.

Tabela 1

Distribuição das Líquidas no Português

Ataque absoluto	Ataque medial	Ataque complexo (C_2)	Coda medial	Coda final
/l/	/l/	/l/	/l/	/l/
	/ʎ/			
/ʀ/	/ʀ/			
	/ɾ/	/ɾ/	/ɾ/	/ɾ/

Exemplos do emprego das consoantes líquidas e de seu licenciamento em constituintes silábicos na gramática da língua são mostrados na Tabela 2.

Tabela 2

Exemplos das líquidas do Português de acordo com a posição silábica

Ataque Absoluto	Ataque Medial	Ataque Complexo (C_2)	Coda Medial	Coda Final
/l/ /l/*ata, /l/uva*	/l/ *bo/l/a, ga/l/o*	/l/ *p/l/anta, b/l/oco*	/l/ *ba/l/de, ca/l/ça*	/l/ *ane/l/, so/l/*
	/ʎ/ *fo/ʎ/a, te/ʎ/a*			
/ʀ/ /ʀ/*oda, /ʀ/ato*	/ʀ/ *ca/ʀ/o, ba/ʀ/iga*			
	/ɾ/ *pi/ɾ/ata, ba/ɾ/ata*	/ɾ/ *p/ɾ/ato, g/ɾ/ama*	/ɾ/ *ca/ɾ/ta, po/ɾ/ta*	/ɾ/ *flo/ɾ/, tambo/ɾ/*

As restrições de licenciamento das líquidas como constituintes silábicos expostas em (3) encontram motivação especialmente em três aspectos:

(a) no grau de sonoridade (*sonority*) do segmento (o molde silábico do Português atende ao Princípio de Sonoridade – Clements, 1990));

(b) na estrutura interna do segmento (a lateral palatal /ʎ/ tem o licenciamento restrito ao Ataque medial por ser uma consoante geminada – Giangola, 1994; Matzenauer, 1999; Wetzels, 1997);

(c) em classes de segmentos (as duas classes que integram a grande classe das consoantes líquidas – a classe das laterais e a classe das róticas – podem ocupar o constituinte Coda, cada uma representando um exemplar fonológico).

Os registros na Tabela 2 também expõem que a líquida lateral /l/ é a única licenciada em todos os constituintes marginais da sílaba.

Reunindo-se essas observações referentes aos dados expostos na Tabela 2, é possível encontrarem-se fundamentos para três fatos que estão na essência deste estudo, quais sejam, o foco nas líquidas laterais; o foco no constituinte Coda da estrutura silábica; o foco na interface estrutura interna do segmento/constituintes silábicos, a qual reúne os dois primeiros eixos da investigação (conforme já referido).

Com relação ao constituinte Coda da sílaba, vale registrar que, de acordo com Bisol (1999), a fonologia do português brasileiro (PB) licencia apenas quatro segmentos consonantais para ocupar esse espaço (fricativa

coronal, nasal, líquida lateral e rótica),[1] o que significa dizer, conforme Matzenauer (2024), que há o licenciamento de quatro "categorias segmentais". É por este funcionamento que são observadas muitas formas variantes na posição de Coda silábica e todas as várias formas fonéticas que, nesta posição silábica, venham representar a categoria "fricativa coronal", ou a categoria "nasal", ou a categoria "líquida lateral" ou a categoria "rótica" têm o seu valor fonológico neutralizado entre si. E é o que se verifica com diferentes formas fonéticas que, nesta posição, podem representar a categoria "líquida lateral", por exemplo. Espiga (2000), em investigação de caráter sociolinguístico em região fronteiriça do Brasil com o Uruguai (país de fala espanhola), identificou cinco formas no espaço da lateral em Coda silábica – [l] ~ [ɬ] ~ [l^{w}] ~ [w] e zero fonético – todas elas representam, portanto, a categoria fonológica "líquida lateral" em Coda.

Com foco na líquida lateral, estudos de cunho variacionista sobre o PB, ao mesmo tempo em que apontam para a estabilidade na representação do segmento no Ataque de sílaba, revelam haver variação nas formas que ocupam o espaço de Coda silábica, seja medial ou final de palavra, sendo prevalente a forma vocalizada [w] (Bisol 1996/2014; Brandão 2021; Collischonn & Costa, 2003; Espiga, 2001; Mateus & d'Andrade, 2000; Quednau, 1993; Tasca, 1999; Wetzels & Schwindt, 2016).

Considerando a realidade fonética que, no PB, representa a categoria "líquida lateral" no constituinte Coda silábica, a questão que se discute é o encaminhamento, no processo de aquisição da linguagem por crianças brasileiras, da construção do conhecimento fonológico relativo à líquida lateral nesta posição da sílaba, à medida que o input recebido pela criança é de uma semivogal dorsal (*carnava*[w]*, ane*[w] para *carnaval* e *anel*, respectivamente).

A questão do descompasso entre a forma fonológica e a forma fonética da lateral em Coda no PB é complexa e exige um exame do comportamento morfofonológico do segmento.

3. O comportamento morfofonológico da lateral em Coda no português brasileiro

No português brasileiro a lateral pós-vocálica, predominantemente produzida como semivogal nas posições medial e final, implica uma produção, juntamente com a vogal precedente, que se mostra homófona aos ditongos decrescentes constituídos por vogal mais vogal alta dorsal. Observa-se, porém, que a distribuição dos ditongos com a vogal alta dorsal, redundantemente arredondada, /u/, é bastante restrita na língua, havendo apenas o licenciamento com as vogais não arredondadas, como nas palavras ***Eu**ropa*, ***céu*** e *degr**au***, por exemplo.[2] Já a sequência vogal mais lateral apresenta uma distribuição mais ampla no sistema, sendo possível encontrá-la com todas as vogais na posição medial e final conforme os exemplos na Tabela 3:

[1] As análises fonológicas para o Português Europeu não preveem que a Coda silábica seja ocupada por um segmento nasal (Costa & Freitas, 2001; Mateus & d'Andrade, 2000).

[2] O ditongo 'iu' resulta do processo flexional em verbos de terceira conjugação (*dormiu* e *sentiu*) e o 'ou' em verbos de primeira conjugação (*andou* e *cantou*). O caso do ditongo 'ou', em posição medial de nomes, que apresenta a combinação de vogais arredondadas, é justamente aquele em que não há contrastividade e há alternância total na língua [ow] ~[o], o que, de acordo com Camara Jr., é evidência de que tal ditongo não seria fonológico, mas um recurso estilístico.

Tabela 3

Exemplos da Forma Fonética da Líquida Lateral em Coda Silábica no PB

Coda medial	Coda final	Forma Fonética
filme	funil	[iw]
felpa	delével	[ew]
selva	papel	[ɛw]
salva	canal	[aw]
volta	sol	[ɔw]
polvo	gol	[ow]
pulso	azul	[uw]

Os exemplos apresentados são relativos à presença da lateral pós-vocálica em radicais da língua e ilustram o processo de semivocalização predominante em manifestações fonéticas do português brasileiro. São, portanto, estas as formas fonéticas que estão na base da construção das representações fonológicas das crianças antes do contato com a escrita alfabético-ortográfica. Há, no entanto, alternâncias que podem ser observadas na derivação e na flexão, as quais podem revelar um comportamento morfofonológico da lateral capaz de fornecer pistas relativas à sua presença em Coda, como mostram os exemplos na Tabela 4.

Tabela 4

Exemplos de Alternâncias da Líquida Lateral em Coda Silábica – Formas Primitivas e Derivadas

funil	[w]	funil+eiro	[l]
papel	[w]	papel+ada	[l]
canal	[w]	canal+ete	[l]
sol	[w]	sol+aço	[l]
gol	[w]	gol+ear	[l]
azul	[w]	azul+ar	[l]

Na Tabela 4, a derivação por meio da sufixação é fato que expõe a forma fonética da lateral nas palavras em que o segmento ocupa a posição de borda na base. Apenas nos casos das formas terminadas em *-el*, as quais correspondem prosodicamente às palavras paroxítonas, sejam derivadas ou não do sufixo *-vel*, não há ocorrências em que a derivação revele a lateral foneticamente.[3]

Já a flexão de número das palavras terminadas em lateral apresenta, na língua, comportamento condicionado pelo acento e as formas resultantes não trazem a lateral à superfície, como mostram os exemplos na Tabela 5:

[3] Embora exemplificada a sequência [ol], em *gol*, trata-se de um exemplo único no léxico, pois a tendência é de uma vogal média baixa na sequência vogal+lateral.

Tabela 5

Exemplos de Alternâncias da Líquida Lateral em Coda Silábica na Flexão

papel	papéis	l-is
canal	canais	l-is
azul	azuis	l-is
sutil	sutis	l-s
funil	funis	l-s
fácil	fáceis	il-eis
hábil	hábeis	il-eis

Como mostram os exemplos, as palavras oxítonas selecionam *-is* e *-s* para a formação do plural e as paroxítonas, *-eis*. Tal informação decorrente do funcionamento morfológico pode servir como pista para diferençar sequências de vogal mais lateral (VL) ou vogal mais vogal alta dorsal (VU), à medida que para as primeiras a forma de flexão selecionada contém uma vogal, como se observa nos casos de palavras terminadas em outras consoantes (*flor-flores; paz-pazes*), enquanto para as últimas apenas o *-s* é selecionado (*degrau-degraus; museu-museus; chapéu-chapéus*).

No que diz respeito à adjunção do sufixo *-(z)inho*, cujo funcionamento é peculiar, em se comparando aos sufixos que afetam o acento da palavra (***ro**sa-ro**sei**ra-rosei**ral***) e/ou produzem modificação na classe gramatical (*rosa-rosado*), pode diminuir a opacidade apresentada pelo /l/ na posição pós-vocálica assim como os exemplos em (6), à medida que a adjunção de *-(z)inho* produz formas que deixam em evidência a flexão ocorrida, conforme observado em *papelzinho-papeizinhos* ou *anelzinho-aneizinhos*, por exemplo.

O funcionamento morfofonológico da lateral em posição final, diferentemente do que se observa na posição medial, é, portanto, capaz de fornecer pista referente à presença de /l/ em estruturas com Coda silábica. Mesmo que prepondere na pronúncia dos falantes a forma vocalizada nos radicais, a palavra derivada, por ressilabação, faz emergir a lateral em Ataque.

Observa-se, pois, que, quanto ao comportamento morfofonológico do segmento na derivação e na flexão de palavras da língua, há indícios da presença de uma lateral em posição de coda, especificamente em posição de final de palavra, seja na derivação (*sal-saleiro*) seja na flexão de número em que palavras com VL se diferenciam daquelas terminadas em VU (*mural-mura**is***; *degrau-degrau**s***).

É importante mencionar que, além de fatos morfofonológicos, referidos nesta seção, o conhecimento fonológico da líquida lateral em Coda integra a estrutura interna do segmento, o seu licenciamento como constituinte silábico, a sua representação na manifestação escrita da língua, bem como o input fonético que a criança recebe. É esse complexo de fatores que compõe a formação do nível da representação fonológica da líquida lateral.

4. A aquisição da líquida lateral em Coda silábica por crianças brasileiras – o estabelecimento da representação fonológica

A gradual aquisição da fonologia por crianças brasileiras evidencia o condicionamento da estrutura interna do segmento neste processo. Ao observar-se a classe das consoantes líquidas, embora todos os segmentos compartilhem propriedades fonéticas e traços fonológicos, são detectadas especificidades, as quais são determinantes dos contrastes entre os segmentos e também de uma ordem de aquisição. Na aquisição do PB, as crianças tendem a apresentar a ordem de aquisição das consoantes líquidas como na Tabela 3, seguindo Miranda, (1996), Rangel (1998), Matzenauer (2003) e Lamprecht et al. (2004):

(1) /l/ > / ʀ / > /ʎ/ > /ɾ/

Esta ordem de aquisição é condicionada pela estrutura silábica, já que os segmentos emergem foneticamente e são fonologicamente estabilizados em diferentes estágios do desenvolvimento linguístico. Quanto à líquida lateral /l/, por exemplo, as crianças brasileiras tendem a apresentar o ordenamento mostrado em (2), de acordo com Matzenauer-Hernandorena e Lamprecht (1997).

(2) /l/ - posição de Ataque simples absoluto – exemplo: [l]obo para *lobo* – idade: 2:8 (anos: meses)

/l/ - posição de Ataque simples medial – exemplo: bo[l]o para *bolo* – idade: 3:0

Subsequentemente à aquisição da lateral na posição de Ataque simples, observa-se aos 5 anos a sua estabilização fonológica na posição de Ataque complexo, em palavras como *f*[l]*or* e *b*[l]*usa*, por exemplo. No que se refere à Coda em posição medial e final de palavra, estudos como os de Matzenauer (1990) e Miranda (2009) assumem que, nesta posição da estrutura silábica, tendo em vista a preponderância do PB de uma sequência vogal mais glide dorsal, a coda com lateral não seria computada pela criança. Tal proposta é corroborada pelos dados que mostram a precocidade das produções com ditongos desde as primeiras palavras em exemplos como *ba*[w]*de* e *so*[w], para os vocábulos *balde* e *sol*.

Ao se examinarem as estratégias empregadas pelas crianças até a estabilização dos segmentos em seu inventário fonológico, verifica-se também um tratamento diferenciado ao se observar como as crianças brasileiras ocupam o espaço fonético-fonológico da líquida lateral /l/ até a sua plena aquisição. Esta ocupação do espaço de /l/ também sofre condicionamento do constituinte da sílaba. Vejam-se os dados em (3).

(3) /l/ - posição de Ataque silábico simples – exemplo: [∅]*obo*, *bo*[j]*o* e *ma*[l]*a*, para *lobo*, *bolo* e *mala*, respectivamente, com ocupação do espaço: ∅ > j > l

/l/ - posição de Coda – exemplo: *sa*[∅]*to* e *sa*[w]*to*, para *salto*, com ocupação do espaço: ∅ > w

A ocupação do espaço fonológico de uma consoante líquida por um glide, seja em Ataque ou em Coda, pode ser considerado natural por serem ambos os segmentos integrantes da classe [+aproximante] – e esse tipo de ocupação de espaço é registrado no processo de aquisição de diferentes línguas, como o Inglês (Ingram, 1985) e o Francês (Yamaguchi, 2012), por exemplo.

Chama atenção, no entanto, o emprego de diferentes glides, no espaço da lateral /l/, em razão do constituinte silábico: o glide coronal [j] no espaço da lateral na posição de Ataque[4] e o glide dorsal-labial [w] no espaço da lateral na posição de Coda. Esse fato levou à observação do que ocorre no processo de aquisição fonológica por crianças falantes nativas do Português Europeu (PE). De acordo com Matzenauer e Freitas (2024), para as crianças portuguesas as estratégias de reconstrução prevalentes para o alvo /l/ em Ataque simples são estas em (4), sendo o uso de [w] a estratégia mais produtiva, seguida da aproximante labiodental e, em índice menor, seguida do zero fonético.

(4) /l/ - posição de Ataque silábico simples – ocupação do espaço: w > ʋ > ∅

As diferenças de ocupação do espaço da líquida lateral /l/ até a sua aquisição, observadas entre os constituintes silábicos e entre variedades de uma mesma língua, levam à interpretação de que também há condicionamento do input fonético que as crianças recebem. É a partir do input fonético que as crianças constroem as representações fonológicas, com a mediação da capacidade inata de estabelecer categorias linguísticas.

[4] Nos estudos sobre a aquisição fonológica de crianças brasileiras, o glide dorsal-labial [w] ocupa o espaço da líquida lateral /l/ em contexto de vogal dorsal-labial (ex.: *bo*[w]*a* para *bola*).

Embora não haja correspondência direta entre a realidade fonética e a realidade fonológica, há inegável relação entre ambas. Nesse entendimento, concorda-se com Downing e Hamann (2021), que retomam Hyman (2001), no sentido de que propriedades fonéticas não equivalem a unidades fonológicas categóricas abstratas, ou seja, as propriedades fonéticas não implicam diretamente valores fonológicos. O que ocorre é que determinadas propriedades fonéticas são interpretadas, no contexto de uma gramática fonológica, como unidades capazes de alterar significado. A percepção do input fonético, portanto, está relacionada à construção de categorias fonológicas. Sendo assim, as crianças constroem o conhecimento fonológico condicionadas pelo input linguístico que recebem.

Entende-se que o processamento do input linguístico, na aquisição dos segmentos fonológicos, ocorre a partir de categorias maiores, com a gradual especificação de unidades menores. Ao tratar-se das líquidas, seguindo-se o argumento de Matzenauer (2020), o desenvolvimento de segmentos fonológicos ocorre pela categorização da classe de consoantes aproximantes, o que permite que glides ocupem o espaço de consoantes líquidas, sendo subsequentemente ativado o traço fonológico [lateral], com a ativação também, por relação implicacional, do traço não marcado [coronal], motivando o conhecimento fonológico da líquida /l/. É o que acontece com relação ao Ataque silábico, já que o input que as crianças recebem deste constituinte é de uma forma fonética consonantal lateral alveolar [l].

Com relação à líquida lateral na posição de Coda da sílaba, o input que as crianças brasileiras recebem é a forma fonética de um glide [w], conforme já referido (exs.: *ane*[w] para *anel*, *so*[w] para *sol*). O processamento deste input deve categorizar este segmento como pertencente à classe de aproximantes. Com essa interpretação, Matzenauer e Freitas (2024) sustentam que, na aquisição do PB, a representação para o alvo lateral em Coda é um segmento vocálico, /u/, havendo, portanto, a construção de uma categorização fonológica assimétrica da lateral em Ataque e em Coda de sílaba no desenvolvimento fonológico de crianças brasileiras, isto é, uma vogal alta dorsal em Coda e uma lateral em Ataque. Essa categorização poderá passar a ser simétrica subsequentemente, por meio do processo de escolarização das crianças, como se verá a seguir, pelo fenômeno de reestruturação da representação fonológica, como defende Miranda (2019).

A visão aqui exposta de que, na construção do conhecimento fonológico como parte do desenvolvimento da linguagem, a aquisição dos segmentos é condicionada pelos traços fonológicos que compõem a sua estrutura interna, pelo constituinte silábico que ocupam e pelas formas de input que as crianças recebem pode ser formalizada pelo BiPhon, o modelo de processamento e de gramática proposto por Boersma (2007, 20011), a ser apresentado na seguinte seção.

5. Modelo Bidirecional Paralelo de Fonologia e Fonética (BiPhon) e a aquisição da lateral em Coda de sílaba

Bases para a formalização da construção do conhecimento fonológico da líquida lateral em Coda silábica podem ser encontradas no Modelo Bidirecional Paralelo de Fonologia e Fonética (BiPhon), que se configura como Modelo de Gramática e de Processamento de L1, de Boersma (2007, 2011) e Boersma e Hamann (2009).

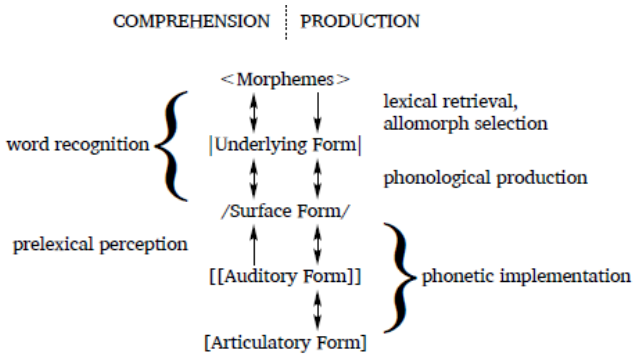
5.1. Caracterização do BiPhon

Como um modelo de processamento e de gramática, particularmente de gramática fonológica, o BiPhon reúne Fonética e Fonologia ao considerar que categorias fonológicas são representações constituídas, pelo falante/ouvinte (ou aprendiz de L1 e/ou de L2), a partir de pistas fonéticas. Assume diferentes níveis de representação, sendo que aqui interessam os níveis identificados como [Forma Fonética] (a qual inclui a [[Forma Auditiva]] e a [Forma Articulatória]), a /Forma Fonológica de Superfície/ e a |Forma Fonológica Subjacente|, presentes em dois módulos: o Módulo da Compreensão linguística e o Módulo da Produção linguística.

Considerando estar o foco do presente estudo na construção do conhecimento fonológico de uma unidade segmental (a líquida lateral), vai-se atentar ao Módulo da Compreensão, na arquitetura do BiPhon mostrado na Figura 2.

Figura 2

Representação de Modelo de Fonética e Fonologia



Nota. Fonte: Boersma et al. (2020, p.107)

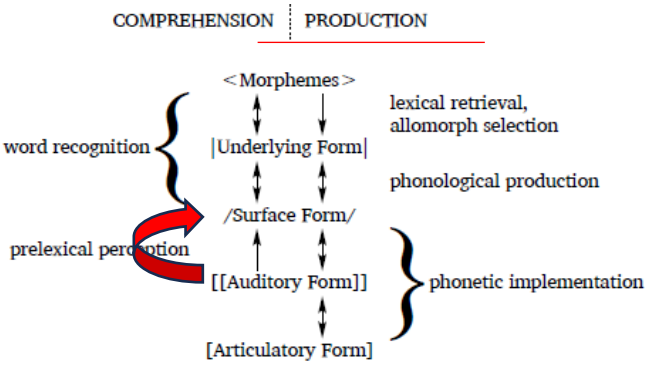
Na Figura 2, observa-se que a Percepção foi incorporada ao Modelo, sendo interpretada como o mapeamento da uma [[Forma Auditiva]] em /Forma Fonológica de Superfície/. As flechas, na Figura 2, indicam o mecanismo do processamento; a articulação entre os níveis cujas designações estão expostas nas bordas da Figura corresponde aos movimentos que caracterizam a gramática. Esse mapeamento implica o processamento do input fonético e a sua interpretação em categorias fonológicas, ou seja, ocorre a interpretação de propriedades acústicas em unidades abstratas da gramática da língua, já que a /Forma Fonológica de Superfície/ se constitui em uma estrutura fonológica que consiste em elementos fonológicos abstratos, como traços, segmentos, sílabas e pés.

É, portanto, na interação entre estes níveis, no Módulo da Compreensão Linguística, que começa a construção do conhecimento das unidades fonológicas, conforme está indicado na Figura 3, a seguir.

E esta /Forma Fonológica de Superfície/ é usada para encontrar uma |Forma Fonológica Subjacente| (*Underlying Form* – UF), a qual se constitui em uma sequência de estruturas fonológicas discretas associadas a morfemas no léxico.

Figura 3

Representação de Modelo de Fonética e Fonologia, com a Indicação da Interação entre os Níveis em que Começa a Construção do Conhecimento Fonológico



Nota. Fonte: Boersma et al. (2020, p.107)

5.2. **A categorização fonológica da lateral, segundo o BiPhon**

Retomando-se o processo de aquisição da líquida lateral, traz-se para consideração inicialmente que as formas fonéticas que a representam, no uso da língua por falantes adultos do PB, são diferentes em Ataque silábico simples e em Coda. Como essas formas fonéticas, segundo o BiPhon, constituem o input que as crianças recebem (são os elementos da sua [[Forma Auditiva]]), constituem a matéria objeto de interpretação para a categorização das unidades abstratas da gramática fonológica (/Forma Fonológica de Superfície/). Diante desse entendimento, o esperado é que, desse processamento, resultem, na aquisição da linguagem por crianças brasileiras, representações fonológicas diferentes para a lateral em Ataque silábico e em Coda: a representação /l/ para o Ataque silábico e a representação /u/, realizada foneticamente como glide [w], para a Coda. É o que defendem Matzenauer e Freitas (2024) e Miranda e Costa (2024) em estudos comparativos entre a aquisição fonológica do PB e do PE.

À luz da formalização do BiPhon exposta na Fig.3, tem-se, então, que as crianças brasileiras, ao receberem a [[Forma Auditiva]] da lateral [l] em Ataque de sílaba (como em [l]*obo* para *lobo*, por exemplo), dispõem de material acústico para processar o segmento como /l/ na /Forma Fonológica de Superfície/. Diferentemente, ao receberem a [[Forma Auditiva]] de glide dorsal em posição de Coda silábica (como em *ane*[w] para *anel*, por exemplo), as crianças dispõem de material acústico para processar o segmento como /u/ na /Forma Fonológica de Superfície/.

Portanto, pelo fato de o BiPhon incluir a Fonética na arquitetura do modelo, reconhecendo na [[Forma Auditiva]] um suporte para a construção de categorias fonológicas, oferece um meio para representar o condicionamento do input fonético que as crianças recebem na formação das unidades que vão compor a sua gramática fonológica, ou seja, das unidades que irão integrar a /Forma Fonológica de Superfície/ e a a |Forma Fonológica Subjacente|,

Indo além, este artigo considera que as representações fonológicas não são estáticas: podem sofrer alterações, no curso do desenvolvimento linguístico, em razão da aquisição de alternâncias morfológicas (conforme discussão na Seção 3) e, de modo particular, em decorrência da aquisição da escrita. É a favor de alterações das representações fonológicas que se argumenta com base na observação do comportamento da líquida lateral na posição de Coda silábica em dados de escrita de crianças brasileiras.

6. A aquisição da escrita da líquida lateral pós-vocálica por crianças brasileiras – a reestruturação da representação fonológica

A aquisição da escrita é parte integrante do desenvolvimento da linguagem (Abaurre, 1991; Miranda, 2014; Miranda & Matzenauer, 2010; Tolchinsky & Berman, 2023). Neste sentido, o *continuum* observado no desenvolvimento do conhecimento linguístico, em todas as suas dimensões (estruturais e discursivas), ganha novo caráter a partir da alfabetização e das práticas de leitura e escrita desenvolvidas em contexto escolar ou mesmo fora dele. Se nos primeiros anos da vida da criança a aquisição da linguagem ocorre de modo natural e espontâneo, sem a necessidade de instrução específica, nos anos subsequentes, em que o processo de escolarização tem seu início, a instrução será necessária para que o sistema de escrita alfabética (SEA) seja apreendido.

A compreensão dos princípios do SEA irá modificar de forma indelével o processamento da linguagem, especialmente o processamento fonológico, uma vez que a base da escrita alfabética se ancora nos elementos constitutivos da gramática sonora, quais sejam, os fonemas e suas combinações. A descoberta pelas crianças de que a língua também é forma sonora e não apenas significado é o *turning point* para a mudança na percepção e para o potencial rearranjo das representações fonológicas.

A partir desta perspectiva, as pesquisas desenvolvidas pelo Grupo de Estudos sobre a Aquisição da Linguagem Escrita, o GEALE, têm analisado aspectos do conhecimento fonológico que se manifestam nas escritas iniciais e são, portanto, reveladores da construção do conhecimento da língua.

Os erros referentes às sílabas complexas, especialmente em estruturas com a Rima ramificada, as quais englobam Rimas com Coda, se destacam pela variedade e quantidade de estratégias empregadas pelos aprendizes na reconstrução de tais unidades. Dentre as quatro classes de segmentos licenciados para a posição, as nasais e as laterais se destacam por serem as que se mostram mais problemáticas às crianças e são estas as estruturas, VL e VN, que paradoxalmente são adquiridas primeiro no desenvolvimento da fonologia. Com base em dados de escrita, Miranda (2009, 2018, 2019) e Pachalski (2020) arguem serem tais estruturas constituídas por núcleos ramificados, sendo consideradas Codas verdadeiras nas gramáticas infantis apenas aquelas com a rótica e a fricativa coronal. As mudanças nas representações estariam condicionadas à continuidade do desenvolvimento fonológico decorrente do processo de escolarização.

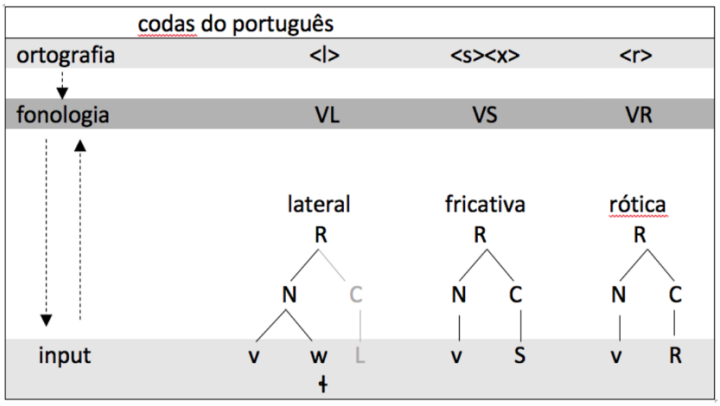
Miranda et al. (2023) analisaram, em um estudo sobre as grafias do glide [w], dados de escrita extraídos de textos produzidos por alunos de 1º ao 3º ano do Ensino Fundamental de 9 anos (faixa etária correspondente a 6 a 8 anos). As grafias para o ditongo derivado da lateral mostram uma distribuição de erros que alcança índices de, aproximadamente, 50% em relação aos acertos. O índice é alto e, dentre as estratégias, preponderam a substituição de <l> por <u>, <o>, <m> ou <n>; e a omissão do grafema, em geral nos contextos de vogal posterior não baixa. Para as autoras, o uso de grafemas nasais é considerado sugestivo, no sentido de ser indiciário sobre a maneira como a criança interpreta as estruturas VN e VL, o que vai ao encontro da ideia de serem a nasal e a lateral pós-vocálica pertencentes ao núcleo e não à Coda na gramática infantil.

Miranda e Costa (2024), em seu estudo sobre as grafias de Rima ramificada extraídas de textos de crianças brasileiras e portuguesas, mostram que a lateral, nos dados do PB, tem uma taxa em torno de 41% de erros, o que não se verifica no PE, cujo índice é de 9%. A diferença observada nos percentuais de erros reforça a ideia de ter o input um papel crucial no desenvolvimento da gramática, pois são as pistas acústicas e articulatórias informações relevantes para a construção do sistema linguístico.

Duas variedades de uma mesma língua, a brasileira e a portuguesa, revelam diferenças no processo de aquisição que apresenta momentos peculiares a cada uma, a depender do input que está disponibilizado pela comunidade de fala. Essas diferenças, no entanto, tendem a ser homogeneizadas pela escrita ortográfica que tem o poder de promover mudanças nas representações levando os usuários a reestruturarem unidades da prosódia, como no caso em estudo. O exame dos resultados de análises da escrita inicial e de estudos do desenvolvimento fonológico leva Miranda e Costa (2024) a ilustrarem a representação das Rimas ramificadas, conforme mostra a Figura 4.

Figura 4

Relação Fonética, Fonologia e Ortografia para as Rimas Ramificadas



Nota. Fonte: Miranda e Costa (2024, p.148)

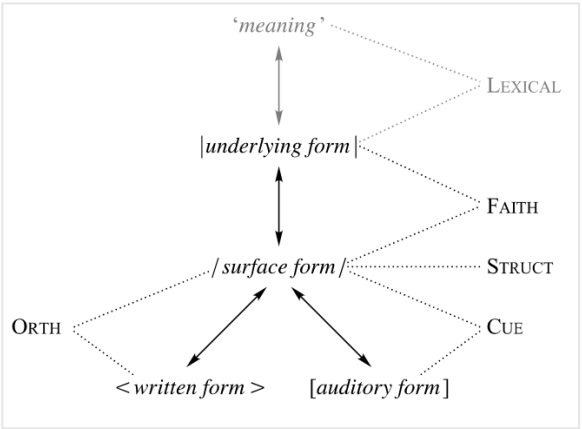
A assimetria verificada entre o PB e o PE, cujos inputs para a estrutura de Rima ramificada com lateral são distintos, quais sejam [w] para o primeiro e [ɫ] para o segundo, vem a reforçar a ideia, antes mencionada, de que a aquisição fonológica está suscetível ao input recebido. A assimetria vai ao encontro da proposta de serem as representações fonológicas passíveis de reestruturações, a partir do conhecimento ortográfico que tende a se consolidar ao longo da escolarização.

Hamann e Colombo (2017), em um estudo que trata de aprendizes de L2, abordam o efeito da ortografia sobre as representações e acrescentam restrições ortográficas (*orthographic* ou *ORTH constraints*) ao modelo BiPhon, buscando formalizar o efeito da ortografia na gramática fonológica.

O resultado da proposta das autoras está reproduzido na Figura 5.

Figura 5

Modelo BiPhon com a incorporação da <forma escrita> (<written form>) e das restrições ortográficas (ORTH).



Nota. Fonte: Hamann e Colombo (2017, p. 701)

De acordo com a figura, a [forma auditiva] assim como a <forma escrita> contribuem para a formação da gramática fonológica, sendo ambas mapeadas para a /forma de superfície/ a partir de dois tipos de restrições, as de pista e as ortográficas, as quais operam em paralelo na avaliação de candidatos. As autoras enfatizam a direção da compreensão no processamento, o que equivaleria à leitura em se tratando da ortografia e asseveram que para dar conta da direção da produção, isto é, da escrita, bastaria inverter a direção do processamento como no modelo BiPhon para a fala, sendo que a mesma lógica de operacionalização e as mesmas restrições podem ser utilizadas.

Uma implicação da proposta de Hamann e Colombo (2017) para a formação da /Forma Fonológica de Superfície/ é o fato de que uma representação fonológica construída a partir do input fonético, ou seja, da [Forma Auditiva], pode ser reestruturado com o acesso, pela criança, à <Forma Escrita>, já que ambas as formas alimentam a construção da /Forma Fonológica de Superfície/. Entende-se haver um processo de reestruturação diante da realidade de o processo de aquisição da escrita ser subsequente à aquisição da fala.

Assim, a ausência de um input fonético capaz de oferecer às crianças elementos para diferençar as estruturas silábicas ($CVV_{[alto]}$ e $CVC_{[lateral]}$) em palavras como *pauta* e *falta*, cujas pronúncias são em PB de uma sequência de vogal mais *glide* [w], mostra-se como um limitador para a construção da fonologia de estruturas $CVC_{[lateral]}$, ainda que aspectos da morfofonologia possam trazer informações de um 'l' pós vocálico. A explicação para o não impacto dessa informação morfológica na fala inicial deve-se ao fato de a morfologia tornar-se mais flexível e acessível ao acesso consciente em decorrência da escolarização que, em seu início, é pautada pela descoberta das relações fonema-grafema. Por esta perspectiva, seria o acesso às formas escritas o gatilho para a reestruturação das representações fonológicas para sílabas com a lateral pós-vocálica.

7. Considerações finais

Neste artigo, a aquisição da fonologia é compreendida como um processo desenvolvimental que extrapola os primeiros anos de construção da gramática sonora, período em que os conhecimentos basilares do sistema são adquiridos. Buscou-se discutir o efeito da aprendizagem da escrita sobre as representações fonológicas de crianças brasileiras e, para tanto, a líquida lateral foi escolhida como tema, uma vez que se trata de um segmento com potencial para o estudo, dada a assimetria verificada em seu funcionamento no Ataque e na Coda silábica em PB, [l] na primeira posição e [w] na última. A opacidade observada neste caso é considerada um elemento dificultador para a construção das representações referentes às sílabas VL, postuladas como estruturas de Rima ramificada com Coda pelos estudos fonológicos da gramática-alvo (Bisol, 1999).

Além de examinar o processo de formação da representação fonológica a partir dos eixos analisados, quais sejam: (a) das unidades que constituem a estrutura interna do segmento, (b) dos constituintes silábicos para os quais o segmento é licenciado, com atenção especial para a Coda, (c) do comportamento morfofonológico do segmento na derivação de palavras da língua e (d) da relação fonologia/ortografia estabelecida no processo de aquisição da escrita, o estudo propiciou a reflexão sobre a importância do input para o desenvolvimento das representações fonológicas, bem como a formalização das mudanças representacionais, com a modelagem proposta por Boersma (2011), Boersma et al. (2020) e Hamann e Colombo (2017).

Ademais, defende-se ocorrer um processo de reestruturação da representação fonológica da líquida lateral em Coda silábica na construção do conhecimento por crianças brasileiras. Pelo input linguístico que recebem (a forma [w] é a manifestação fonética da lateral em Coda), as crianças brasileiras constroem primeiramente a representação de um segmento vocálico, /u/, para o alvo lateral em Coda. Com o processo de aquisição da escrita e com as evidências do comportamento morfofonológico do segmento em Coda final, esta representação tende a ser reestruturada, passando a ser a lateral /l/. Com o suporte do modelo BiPhon, enriquecido pela incorporação da <Forma Escrita> (<*Written Form*>) como input para a formação de categorias fonológicas, conforme proposta de Hamann e Colombo (2017), o processo de reestruturação fonológica também pôde ser captado e formalmente representado.

Financiamento

Artigo desenvolvido no âmbito de projetos financiados pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior– CAPES - Processo n. 88887.937773/2024-00; e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Processos n. 312387/2020-2, 423038/2021-4 e 308970/2023-3.

Referências

Abaurre, M. B. (2001). Dados da escrita inicial: indícios da construção da hierarquia de constituintes silábicos? In C. L. B. Matzenauer (Org.), *Aquisição de língua materna e de língua estrangeira*. EDUCAT.

Bisol, L. (1999). A sílaba e seus constituintes. In M. H. de M. Neves (Org.), *Gramática do português falado* (Vol. 7). Humanitas/FFLCH/USP; Ed. da Unicamp.

Bisol, L. (Org.). (2014). *Introdução a estudos de fonologia do Português Brasileiro* (5.ª ed.). EDIPUCRS.

Boersma, P. (2007). Some listener-oriented accounts of h-aspiré in French, *Lingua, 117*, 1989–2054. https://doi:10.1016/j.lingua.2006.11.00

Boersma, P. (2011). A program for bidirectional phonology and phonetics and their acquisition and evolution. In: A. Benz; J. Mattausch (Eds.), *Bidirectional optimality theory* (pp. 33–72). John Benjamins. https://doi.org/10.1075/la.180.02boe

Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & W. L. Wetzels (Eds.), *Loanword phonology* (pp. 11–58). John Benjamins. https://doi.org/10.1075/cilt.307.02boe

Boersma, P., Benders, T., & Seinhorst, K. (2020). Neural network models for phonology and phonetics. *Journal of Language Modelling 8*(1), 103–177. https://doi.org/10.15398/jlm.v8i1.224

Brandão, S. F. (2021). Vocalização da lateral em Coda silábica em duas variedades do Português. *Labor Histórico*, *7*(2), 87–106.

Camara Jr., J. M. (1970). *Estrutura da língua portuguesa*. Vozes.

Camara, J. M. (1975). *História e estrutura da língua portuguesa*. Padrão.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman. (Eds.), *Papers in laboratory phonology I. between the grammar and physics of speech* (pp. 283–333). Cambridge University Press.

Collischonn, G., & Costa, C. (2003). Resyllabification of laterals in Brazilian Portuguese. *Journal of Portuguese Linguistics, 2*(2), 31–54.

Costa, J., & Freitas, M. J. (2001). Sobre a representação das vogais nasais em português europeu: Evidências dos dados da aquisição. In C. Matzenauer-Hernandorena (Org), *Aquisição de língua materna e de língua estrangeira: aspectos fonético-fonológicos* (pp. 87–110). EDUCAT/ALAB.

Espiga, J. W. da R. (2000). *O Português dos campos neutrais: Um estudo sociolinguístico da lateral posvocálica nos dialetos fronteiriços de Chuí e Santa Vitória do Palmar*. [Tese de doutorado não publicada]. Pontifícia Universidade Católica do Rio Grande do Sul.

Giangola, J. P. (1994). Complex palatal geminates in Brazilian Portuguese. In R. Aranovich, W. Byrne, S. Preuss & M. Senturia (Eds.), *The proceedings of the thirteenth West coast conference on formal linguistics* (pp. 46–61). Center for the Study of Language and Information.

Hamann, S., & Colombo, I. E. (2017). A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. *Natural Language and Linguistic Theory, 35*(3), 683–714. https://doi.org/10.1007/s11049-017-9362-3

Ingram, D. (1989). *First Language Acquisition: method, description and explanation*. Cambridge University Press.

Lamprecht, R. R. (2004). *Aquisição Fonológica do Português. Perfil de desenvolvimento e subsídio para terapia*. Artmed.

Mateus, M. H. M., & D'Andrade, E. (2000). *The phonology of Portuguese*. Oxford University Press.

Matzenauer-Hernadorena, C. L.B. (1999). Aquisição da fonologia e implicações teóricas: Um estudo sobre as soantes palatais. In R. R. Lamprecht (Org.), *Aquisição da linguagem: Questões teóricas e aplicadas*. EDIPUCRS.

Matzenauer, C. L. B. (2003). Restrições segmentais e prosódicas na aquisição das líquidas do Português Brasileiro e do Português Europeu. *Boletim da ABRALIN*, *26*(1), 223–225.

Matzenauer, C. L. B. (2020). Traços e classes de segmentos na arquitetura da gramática fonológica. *Fórum Linguístico*, *17*, 4612–4635. https://doi.org/10.5007/1984-8412.2020v17nespp4612

Matzenauer-Hernandorena, C. L. B., & Lamprecht, R. R. (1997). Aquisição das consoantes líquidas do Português. *Letras de Hoje, 32*(4), 7–22.

Matzenauer, C. L. B., & Freitas, M. J. (2024). Aquisição de Assimetria na Gramática Fonológica do Português: o exemplo da líquida lateral /l/. In D. Hora & C.L.B Matzenauer (Eds.), *Fonologia do Português e interfaces: Fenômenos da Aquisição e da Variação* (pp. 63–79). Routledge Taylor & Francis. https://doi.org/10.4324/9781003294344

Miranda, A. R. M. (1996). *A aquisição do 'r': uma contribuição à discussão sobre seu status fonológico* [Dissertação de Mestrado não publicada]. Pontifícia Universidade Católica do Rio Grande do Sul.

Miranda, A. R. M. (2014). A fonologia em dados de escrita inicial de crianças brasileiras. *Linguistica*, *30*(2), 45–80.

Miranda, A. R. M. (2019). As sílabas complexas: Fonologia e aquisição da linguagem oral e escrita. *Fórum Linguístico, 16*(2), 3825–3848. https://doi.org/10.5007/1984-8412.2019v16n2p3825

Miranda, A. R. M., & Matzenauer, C. L. B. (2010). Aquisição da fala e da escrita: relações com a fonologia. *Cadernos de Educação, 35*, 359–405. https://doi.org/10.15210/caduc.v0i35.1626

Miranda, A. R. M., Pachalski, L., & Reinehr, N. V. (2023). O glide [w] nas escritas iniciais: ortografia e fonologia [The glide [w] in early writing: orthography and phonology]. *Organon, 38*(76). https://doi.org/10.22456/2238-8915.135081

Miranda, A. R. M., & Costa, T. (2024). As consoantes em final de sílaba nos dados da aquisição da escrita: uma análise comparativa entre o português do Brasil e o português europeu. In D. Hora & C. L. B. Matzenauer (Eds.), *Fonologia do português e interfaces: Fenômenos da aquisição e da variação* (pp. 63–79). Routledge Taylor & Francis. https://doi.org/10.4324/9781003294344

Quednau, L. R. (1993). *A lateral pós-vocálica no Português gaúcho: análise variacionista e representação não-linea*r [Dissertação de Mestrado não publicada]. Universidade Federal do Rio Grande do Sul.

Rangel, G. A. (1998). *Uma análise auto-segmental da fonologia normal: Estudo longitudinal de três crianças de 1:6 a 3:0* [Dissertação de mestrado não publicada]. Pontifícia Universidade Católica do Rio Grande do Sul.

Tasca, M. (1999). *A lateral em Coda silábica no sul do Brasil* [Tese de doutorado não publicada]. Pontifícia Universidade Católica do Rio Grande do Sul.

Tolchinsky, L., & Berman, R. (2023) *Growing into language*. Oxford University Press. https://doi.org/10.1093/oso/9780192849984.001.0001

Wetzels, W. L. (1997). The lexical representation of nasality in Brazilian Portuguese. *Probus, 9*, 203–232. https://doi.org/10.1515/prbs.1997.9.2.203

Wetzels, W. L., & Schwindt, L. C. (2016). The morphology and phonology of inflection. In W. L. Wetzels, J. Costa & S. Menuzzi (Eds.), *The handbook of Portuguese linguistics*. John Wiley & Sons.

Yamaguchi, N. (2012) *Parcours d'acquisition des sons du langage chez deux enfants francofones* [Tese de doutorado não publicada]. Université Sorbonne Nouvelle – Paris 3.

Advérbios disjuntivos de estilo simples e multipalavra em português

Simple and multiword disjunctive adverbs of style in Portuguese

Izabela Müller[1], *Jorge Baptista*[1]

[1] Universidade do Algarve, Faculdade de Ciências Humanas e Sociais &
INESC-ID Lisboa – Human Language Technology Lab

Abstract

This paper examines the syntactic properties of some Portuguese simple and multiword adverbs, focusing specifically on disjunctive *adverbs of style* (noted PS). Also referred to as *enunciative*, *metalinguistic*, or *illocutionary* adverbs, PS adverbs function as sentence-external modifiers, typically conveying the speaker's attitude toward a given statement. They are systematically associated with performative operators, whose zeroing yields their particular syntactic properties. This study is based on 87 multiword disjunctive adverbs of style that have been identified among the approximately 3,700 multiword adverbs cataloged in our research (from both Brazilian and European Portuguese), along with 9 single-word adverbs ending in *-mente* from the same class. The aim of this paper is to define the formal properties and semantic nuances of this adverbial class. To that end, we propose a tentative semantic classification of these adverbs and distinguish them from other related constructions.

Keywords: multiword adverbs, compound, style disjunctive adverbs, Portuguese, syntax.

Resumo

Este artigo examina as propriedades sintáticas de alguns advérbios simples e locuções adverbiais do português, com foco específico nos advérbios disjuntivos de estilo (notados PS). Também chamados de advérbios enunciativos, metalinguísticos ou ilocutórios, os advérbios PS funcionam como modificadores externos à sentença, transmitindo tipicamente a atitude do falante em relação a uma determinada afirmação. Eles estão sistematicamente associados a operadores performativos, cujo apagamento dá origem às suas propriedades sintáticas particulares. Este estudo baseia-se em 87 advérbios disjuntivos de estilo formados por locuções, identificados entre aproximadamente 3.700 locuções adverbiais catalogadas em nossa pesquisa (abrangendo o português europeu e o português brasileiro), além de 9 advérbios simples terminados em *-mente* pertencentes à mesma classe. O objetivo do artigo é definir as propriedades formais e as nuances semânticas dessa classe adverbial. Para isso, propomos uma classificação semântica preliminar desses advérbios e os distinguimos de outras construções afins.

Palavras-chave: advérbios multipalavra, advérbios disjuntivos de estilo, português, sintaxe.

1. Introduction

This paper aims to investigate a set of adverbs in Portuguese, more specifically, multiword disjunctive adverbs of style. The objectives of this paper are threefold: (1) to provide a syntactic description of this class of adverbs, grounded in the theoretical principles of Transformational Operator Grammar (Harris, 1976, 1991), namely by analyzing them with the help of the metalinguistic performative operator *Eu digo* 'I say'; (2) to distinguish their use as disjunctive style adverbs, from other classes of adverbs, mentioned in the literature, in

order to clarify differences in meaning and distribution; and (3) to analyze their behavior in discourse through corpus analysis.

This paper is organized as follows. We began with a general introduction to the objectives of the paper. Section 2 includes the definition and theoretical description of Disjunctive Style Adverbs. In Section 3, we describe the current work on Multiword Adverbs in Brazilian Portuguese, providing an overview of the construction of the lexicon and the syntactic-semantic classification proposed for Portuguese. We also make a tentative semantic classification for *Disjunctive Style Adverbs* and present a full list of these adverbs. Section 4 highlights the differences and similarities between Conjunctive (PC) and Disjunctive Style (PS) adverbs. Section 5 describes the relations between Disjunctive Style (PS), Manner (MV), and Subject-oriented adverbs (MS), demonstrating that certain adverbs may assume multiple semantic roles depending on whether they modify the sentence as a whole or a specific constituent. Section 6 examines PS adverbs in combination with other performative metalinguistic operators. Section 7 reviews related work across different languages, and finally, Section 8 concludes this work with future perspectives in analyzing this class of adverbs in a corpus. The paper concludes with an Annex presenting the style-disjunctive adverbs here studied.

2. Disjunctive adverbs of style: general definition and theoretical description

Disjunctive adverbs of style (henceforward *PS adverbs*) reflect the speaker's attitude with respect to the content of his/her utterance. These adverbs usually occur at the beginning of a sentence, establishing the tone of the speaker's statement. They include adverbs such as *francamente*/*com* (*toda a*) *franqueza* 'frankly, in all honesty', e.g.,

(1) Francamente/com toda a franqueza*, a cidade está imunda* [1]
'Frankly/in all honesty, the city is filthy'

In this example, the adverb *francamente* points to the *manner* in which the speaker utters the sentence. This meaning can be made overt by inserting the adverb in a sentence with a performative metalinguistic operator *eu digo* (Harris, 1991; Gross, 1986, p. 93):

(2) *Eu* digo francamente/com toda a franqueza *que a cidade está imunda*
'I say frankly/in all honesty that the city is filthy'

In this sentence, the manner, subject-oriented, adverb *francamente*/*com toda a franqueza* modifies the verb *digo*. Before proceeding, let us define more precisely what manner, subject-oriented adverbs are considered here. To illustrate this, we will use a similar example but with a different subject (3):

(3) *O Pedro disse isso* francamente 'Pedro said this frankly/in a frank manner'

Based on Molinier / Levrier classification framework on French adverbs ending in *ment* '-ly', which we adopted in this study, the adverb in (3) is considered to be a sentence-internal modifier of the verb because he does not admit *simultaneously* the following two properties: Firstly, it *cannot* be detached to the beginning of the sentence when the verb is negated, since, as a verb modifier, it falls under the scope of the negation adverb, (3a) and (3b):

[1] Unless otherwise stated, all examples are original and produced by the authors. Translations and glosses are approximate and primarily literal, aiming to highlight the linguistic phenomena under analysis. Their grammatical acceptability—or lack thereof—is not relevant to the discussion.

(3a) *O Pedro não disse isso* francamente
'Pedro did not say that frankly'

≠

(3b) ºFrancamente, *o Pedro não disse isso*
'Frankly, Pedro did not say that'

Notice that sentence (3b) is an acceptable utterance but with a different meaning than that of (3a), which is signaled by the symbol 'º'. Hence, we are dealing with *two distinct constructions* here. While in (3a) the adverb indicated the manner of saying (and the corresponding attitude of the sentence's *subject*), in (3b) it conveys the *speaker's* attitude towards the content of the main clause. This second use corresponds, then, to the disjunctive-style adverbial construction we aim at in this paper. Secondly, the adverb *can* undergo clefting with *ser … que* 'be … that', like many other sentence-internal constituents, especially in a contrastive context, which enables one to continue the sentence with an adversative clause (3c):

(3c) *Foi* francamente *que o Pedro disse isso* (*e não hipocritamente*)
'It was frankly that Pedro said it (and not hypocritically)'

Finally, the adverb *francamente* is said to be *subject-oriented* due to the equivalence with an adjectival sentence taking the same subject as the verb (3d):

(3d) *O Pedro foi* franco (*ao dizer isso*) 'Pedro was frank (in saying that)'

This adjectival construction can also appear within a relative clause, in the analytical construction of the corresponding manner adverb, as the modifier an operator-noun *maneira* 'manner' (and its equivalent forms *modo*, *forma,* and *jeito*[PB] 'way/manner'):[2]

(3e) *O Pedro*$_i$ *disse isso de um modo que foi* franco *da sua*$_i$ *parte*
'Pedro said that in a way that was frank on his part'

The equivalent multiword adverb *com toda a franqueza* has practically the same properties as *francamente*, occurring in the same syntactic slots as the simple-word adverb, except for the corresponding paraphrasis for (3d), since this is a predicative noun. One can, however, suggest a support-verb construction with *ser de* 'be of' (Baptista, 2005) as the base form for this noun, (3f) and (3g):

(3f) *O Pedro foi de uma certa* franqueza (*ao dizer isso*)
'Pedro was of (=showed) a certain frankness (in saying that)'

(3g) *O Pedro*$_i$ *disse isso de um* modo *que foi de uma certa* franqueza *da sua*$_i$[3] *parte*
'Pedro said that in a way that was of a certain frankness on his part'

In the remainder of the examples, we simplify the description by using the simple form *francamente*, referring to the multiword expression only when relevant differences warrant separate discussion.

[2] The 'PT' and 'PB' indices indicate the language variety the expression belongs to, namely the European or the Brazilian varieties of Portuguese, respectively.
[3] The indice '*i*' indicates coreference

Let us now return to sentence (2) with the verb *digo* 'say'. This verb holds a special status within the theoretical framework adopted here (Harris, 1991, pp. 123–128). Transformational Operator Grammar posits that natural language contains its own metalanguage; more precisely, there is no external metalanguage to natural language. According to the theoretical principle of the *least grammar*, grammatical analysis should employ the minimal set of constructs necessary, so that constraints on word combination are kept to a minimum and are not redundantly stated in the generation of utterances (Harris, 1991, pp. 31–36). Within this framework, sentences such as (2), with the metalinguistic performative operator *Eu digo* 'I say' are considered to underlie all actual (declarative) utterances of the language, with the subject *Eu* 'I' corresponding to the speaker and the clausal complement representing the content of the utterance[4]. Furthermore, *Eu digo* 'I say' is used in the present-indicative tense, temporally anchored to the moment of utterance. When the metalinguistic operator *Eu digo* is zeroed—and the sentence is actually uttered, the adverb becomes a sentence-external modifier, that is, the performative operator is pragmatically nullified, yielding the sentence-initial adverbial element that appears detached from the sentence's core structure. Thus, the adverb comes to modify the entire sentence and exhibits the characteristic properties of this type of construction, as defined by Molinier and Levrier (2000), which will be further explained below. Notice that the Harrissian framework, zeroing operations do not change the information constraints on the base sentences, only their surface form. In this sense, the zeroing of *Eu digo* does not alter the constraints of *francamente* on its argument (the performative *digo*) but can help explain its syntactic behavior as a sentence-external modifier once the performative has been zeroed.

To illustrate these properties, we resort to a clear-cut example of the style disjunctive construction of the adverb, (4a) and (4b). These properties consist of two, *simultaneously* observed, patterns: (i) Sentence-external adverb modifiers are typically placed at the beginning of the sentence, where they remain unaffected by the scope of the negation of the main clause, as illustrated in (4a); moreover, (ii) these adverbs cannot undergo *clefting* — in other words, they cannot be extracted using structures like *ser...que* 'it is...that', as demonstrated in (4b):

(4a) Francamente*, eu* não *confio nele*
'Frankly, I do not trust him'

(4b) **É* francamente *que eu confio nele*
'It is frankly that I trust [in] him'

As mentioned above, sentence (4b) is acceptable only with a different meaning from that of (4a). In (4a), *francamente* 'frankly' expresses the speaker's attitude toward the content of the utterance and falls outside the scope of the negation *não* 'not'. In contrast, in (4b)—which is barely acceptable—the adverb would be interpreted as referring to the subject's attitude while performing the act of *confiar* 'trust'. This reading is semantically incongruous, which explains the negative acceptability judgment associated with (4b).

The analysis of the adverb *francamente* presented above considers, then *two* constructions: (a) a base construction as a sentence-internal, subject-oriented, manner adverb, in which it can modify a large set of verbal predicates; and (b) a special construction as the modifier of the performative metalinguistic operator *Eu digo*, which, after its reduction, leaves the adverb as a sentence-external modifier.

The use of a verb like *confiar* 'trust' in sentences (4a) and (4b) can be seen as a clear-cut example of the style disjunctive construction of the adverb *francamente,* as this adverb, from a semantical viewpoint, *cannot*

[4] For simplicity, in the following analysis we ignore the indirect (dative) complement of the verb *dizer* 'say' (*Eu digo*-te 'I say to_you') and later on, the second complement of the verb *falar* 'talk/speak' (*Eu falo* contigo 'I speak to_you', corresponding in both cases to the **addressee** of the **speech act**.

be a manner, subject-oriented modifier of such verb.[5] Notice that placing this adverb at the right of the verb, as in (5a) and (5b), produces unacceptable utterances, as this adverb cannot be a proper modifier of this verb:

(5a) **Eu confio nele* francamente
'I trust him frankly'

(5b) º*Eu confio* francamente *nele*
'I trust frankly [in] him'

Note that sentence (5b) is acceptable, but its meaning differs: in this case, *francamente* functions as a quantifier or intensity modifier (class MQi), typically applied to adjectives, as illustrated in (6). Sentence (5a), on the other hand, is even less acceptable, even when interpreted with the same quantifying sense.

(6) *O Pedro está francamente/muito cansado*
'Pedro is frankly/very tired'

The corresponding multiword adverb does not exhibit this behavior, as the sentence **O Pedro está com toda a franqueza cansado* is unacceptable. Naturally, this usage falls outside the scope of the present analysis.

Since the adverb *francamente* (in the sense here relevant) cannot be a modifier of *confiar*, its presence in sentence (4a) can only be explained by deriving this sentence from the one with the performative metalinguistic operator *Eu digo* 'I say', where the adverb occurs with a normal distribution (4c):

(4c) *Eu digo* francamente *que não confio nele*
'I say frankly that I don't trust him'

As mentioned before, PS adverbs typically introduce the tone of the speaker or their attitude concerning the content of the utterance. It may also express a belief, a criticism, a judgment, skepticism, or a mere impression. They present specific distinguishing characteristics (Molinier & Levrier, 2000), namely:

A. The property of functioning as manner adverbs modifying a performative verb *dizer* 'say' Harris (1976, pp. 155–158; 1991, pp. 137–144), found in a higher-level clause as demonstrated in (5):

(5) *Eu* digo francamente/com toda a franqueza *que eu não confio nele*
'I frankly say that I do not trust him'

[5]A search in the entire Portuguese *ptTenTen23* corpus (Kilgarriff et al., 2014), (https://www.sketchengine.eu/pttenten-portuguese-corpus/; 20 billion tokens), using Sketch Engine (https://app.sketchengine.eu/) and the CQL pattern [lemma="confiar"] []{0,5} [lemma="francamente"]—that is, the lemmas of the verb *confiar* 'to trust' and the adverb *francamente* 'frankly' within a window of up to five intervening words—returned only 17 occurrences. However, in most of these, the two words appear in separate sentences. Only three instances include both the verb and the adverb in the same sentence, and even then, they are syntactically unrelated. For example: (1) [...] *eu não* ***confiava*** *em nenhum deles para conversar* ***francamente*** ('I didn't trust any of them to speak frankly'), where the adverb modifies *conversar*, not *confiar*; or cases in which the adverb appears in a separate clause or parenthetical comment, functioning as a style disjunct rather than a manner adverb: e.g., (2) *Confio em si, meu caro doutor, é agora a única pessoa em que eu posso* ***confiar****, mas,* ***francamente****, não alcancei ainda compreender o seu plano!* 'I trust you, my dear doctor—you are now the only person I can rely on. But, frankly, I still haven't managed to understand your plan!'; or (3) [...] *como espera que depois de adolescente venham a* ***confiar*** *em você,* ***francamente*** *não vejo como* 'How do you expect them to trust you after adolescence? Frankly, I just don't see how.'). The intuitive unacceptability of the word combination *confiar–francamente* (functioning as manner, subject-oriented) appears to be supported by the absence of empirical evidence for such a combination in a corpus of this magnitude.

B. The property of appearing as a manner modifier of a gerundive form of *falar* 'say', *falando* 'saying' which functions in a similar way as a *verbum dicendi* (Baptista, 2010), as in example (6); in Portuguese, the adverb and the verb in the gerund can permutate: *falando francamente/francamente falando*:

(6) Falando francamente/Francamente falando/Com toda a franqueza*, eu (não) confio nele*
'Frankly speaking/Speaking frankly/With all honesty, I (do not) trust him'

Other constructions of *falar* and the adverb *francamente* can also be found, besides the gerundive, namely, *a*/*para falar francamente* 'to speak frankly'. In this case, the permutation is not acceptable: **a*/*para francamente falar* 'to frankly speak'. For simplicity, we will focus on the gerundive structure alone.[6]

The use of the gerund form *falando* ('speaking') calls for a more nuanced analysis. It may be argued that *falar* ('to speak') also assumes a metalinguistic function. Unlike the verb *dizer* ('to say'), which foregrounds the speaker's attitude towards the propositional content of an utterance, *falar* ('to speak') highlights the manner—or other pragmatic features—through which the utterance is conveyed. The occurrence of *falar* in a reduced gerundial clause, as seen in sentence (6), suggests that such constructions result from the reduction of an underlying structure involving a complex combination of two metalinguistic operators: *dizer* and *falar*, linked via the gerundive suffix *-ndo*, which is analyzed here as a subordinative-conjunctive (bound) morpheme. This type of concatenation is exemplified in (7a) and (7b), where the symbol '#' indicates the sentence boundaries, and *-ndo* is the bound conjunctive morpheme concatenating the two base sentences[7]:

(7a) Eu digo *que não confio nele* # -ndo # Eu falo *francamente/com toda a franqueza*

(7b) Falando *francamente/com toda a franqueza, eu digo que não confio nele*
'Frankly speaking, I say that I (do not) trust him'

It is worth noting that, although the adverb *francamente* can, in principle, modify both verbs (e.g., *dizer/falar francamente* 'to say something/to speak frankly'), the two metalinguistic operators serve two distinct functions.

This use of the verb *falar* 'to speak' acts as a *framing device* for the entire utterance—that is, a (subordinate, gerundive) construction that appears at the left periphery of the sentence and sets the speaker's stance or interpretive mode without affecting the core propositional content—where it introduces the adverb. The verb *dizer* 'to say', then, introduces the propositional content of the utterance. This distinction reinforces the sentential scope of the adverb. It may also explain why the two verbs are not interchangeable in such constructions:

[6] In the Portuguese *ptTenTen23* corpus (Kilgarriff et al., 2014), we encountered the following constructions with ***falar*** 'speak' and the adverb ***francamente*** 'frankly': *falando francamente* 'speaking frankly' (731 occurrences) in both the initial position of a sentence, followed by a comma, and detached in the middle of the sentence, also separated by commas, which indicates in a clearer way their use as the style–disjunctive adverbial expressions here analyzed. The lack of these punctuation patterns yields many spurious instances of this word combination.The inversion *francamente falando* 'frankly speaking', following the same punctuation pattern, is less frequent (178), however, still significant. The word combination *a falar francamente* 'to speak frankly', although present in the corpus (12 occurrences), does not carry the adverbial function that is the object of this study. On the other hand, *para falar francamente* 'to speak frankly' (65 occurrences), appearing mainly in the middle of the sentence but not at the beginning, were practically all true instances of the target expression. No cases of adverb-verb inversion were found. The queries of the combinations of the verb ***dizer*** 'say' and the adverb ***francamente*** 'frankly': showed scarce number occurrences: *dizendo francamente* 'saying frankly' as an adverbial appears 6 times out of 7 in total; while the inversion *francamente dizendo* 'frankly saying' showed up twice, both corresponding to the adverbial. Lastly, *a dizer francamente* 'to say frankly' was not found in the corpus, and *para dizer francamente* showed 8 occurrences, all of them cases of the target expression. Again, no cases of adverb-verb inversion were found.

[7] For simplicity, we do not detail the transformation here. It entails the subordination of the secondary clause to the first one, reducing of repeated coreferential material, and the permutation of the gerundive to the front of the main clause.

(8) **Dizendo francamente/com toda a franqueza, eu falo que [eu] (não) confio nele*
'Frankly saying, I speak that I (do not) trust him'

It should also be noted that, although the verb *falar* ('to speak') can appear with an object subclause in Brazilian Portuguese (BP)—as in *Eu falo que não confio nele* ('I say I do not trust him')—this construction is not usually accepted in European Portuguese (EP). While the unacceptability of example (8) in EP could be attributed to the absence of this syntactic pattern, the same explanation does not account for its unacceptability in BP.

This distinction between *dizer* and *falar* is also reflected in English, where *say* and *speak* exhibit a similar functional differentiation, as well as in other Romance languages, such as French. In English, *say* cannot be replaced by *speak* in comparable constructions, as the unacceptability of the translation of sentence (8) demonstrates:

(9a) *Frankly* speaking*, I* say *that I (do not) trust him.*

(9b) **Frankly* saying*, I* speak *that I (do not) trust him.*

A parallel pattern is observed in French, with the verbs *dire* 'say' and *parler* 'speak':[8]

(10a) À *franchement* **parler***,* [*je* ***dis*** *que*] *je* (*n'*) *ai* (*pas*) *de confiance de lui*

(10b) *À *franchement* **dire***,* [*je* ***parle*** *que*] *je* (*n'*) *ai* (*pas*) *de confiance de lui*

These examples underscore the cross-linguistic robustness of the functional asymmetry between the two verbs, further supporting the analysis of *falar* as a performative framing operator and *dizer* as a performative content introducer.

C. Another key property of sentence-modifying, style-disjunctive adverbs such as *francamente* 'frankly' (or *com toda a franqueza* 'with complete frankness') is their restriction to occur only in assertive utterances, that is, with the performative operator *Eu digo* 'I say'. These adverbs are typically excluded from sentences with interrogative (11a), imperative (11b), or exclamative (11c) modality:

(11a) **Francamente, eu* (*não*) *confio nele*?
'Frankly, do I trust him?'

(11b) **Francamente*, *confia* (*tu*) *nele* / *não confies* (*tu*) *nele*.[9]

[8] In the French *frTenTen23* corpus, with +23 billion words, we find several variants of these constructions, including the permutation of the verb and the adverb, and the variation of the initial preposition along with the verb form (*à* + infinitive, *en* + gerund). The following distribution is observed for the verb *parler* 'speak': *à parler franchement* (230 occurrences, with the targeted pattern of the detached adverbial attested), *à franchement parler* (174, idem), *en parlant franchement* (33, idem), *en franchement parlant* (0; no occurrences found), *parlant franchement* (80, idem), *franchement parlant* (39, idem). Interestingly, with *dire* 'say', the situation is different: *à dire franchement* occurs 714 times, but none of them corresponds to the adverbial in detached position; with permuted adverb, *à franchement dire*, only 4 occurrences were found, but they all correspond to the target expression; the *en* + gerund pattern *en disant franchement* occurs only 19 times, and none is the target expression, while no cases were found of the permuted adverb, *en franchement disant*. No instances of the gerund construction without preposition *disant franchement* were found that would correspond to the target expression. It seems, then, that in French, the combination with *dire* 'say', *à franchement dire*, is combinatorially more constraint than the more variable structure with *parler*.

[9] Notice that, in the case of the imperative sentence (11b), a 2nd-person *tu* 'you' must be used as required by the imperative modality. For the interrogative, the use of a 1st-person *eu* 'I', while not impossible—as a rhetorical device, for example—is pragmatically awkward, since the person asking is also the person that is best placed to provide the answer. The sentence is much more natural with a 2nd-person, as illustrated in (11a).

'Frankly, do / do not trust him.'

(11c) **Francamente, eu confio nele!*
'Frankly, I do trust him!'

This is considered a direct consequence of the fact that the adverb *francamente* is semantically (distributionally) incompatible with the metalinguistic performative operators *Eu pergunto* 'I ask/question', *Eu peço/ordeno* 'I ask/command' and *Eu exclamo* 'I exclaim', which are considered to underly such sentences (Harris, 1976, 1991):

(12a) **Eu pergunto francamente se confias nele*
'I ask/question you frankly whether you trust him'

(12b) **Eu peço/ordeno francamente que confies nele*
'I ask/command you frankly to trust him'

(12c) **Eu exclamo francamente que confio nele*
'I exclaim frankly that I trust him'

This distributional restriction is also closely linked to the incompatibility of such adverbs with fronted detached structures when the base sentence is negated. As Molinier and Levrier (2000, p. 65) explain, this constraint arises from the impossibility of simultaneously modifying how a declarative statement is uttered and superimposing an alternative modality (interrogative, imperative, or exclamative) onto it. The adverb presupposes a statement-like speech act, a meaning that is overtly made explicit in the base sentence with the performative metalinguistic operator *Eu digo* 'I say', which is disrupted when another modality is introduced. As negation does not change the modality of the main clause, this is further indirect evidence for the independence of PS adverbs regarding the utterance under their scope.

This said, there are instances of this and other similar adverbs appearing in (apparently) non-assertive utterances.[10] These sentences required further analysis. For example, in (13a):

(13) *Francamente, fizeste isso?*
'Frankly, did you do this?'

the adverb's role in the sentence cannot be put in relation with *perguntar* 'ask/question', since there is no sentence **Eu pergunto-te francamente se fizeste isso* 'I ask you frankly whether you did this'. On the contrary, *francamente* it is interpreted as a (negative) side comment, eventually expressing incredulity or criticism. This is also used in isolation, with an exclamative value:

(14) *Fizeste isso? Francamente!*
'Did you do this? Frankly!'

The orthographic contrast of the two sentences above clarifies that the adverb is being used in (13) in much the same way as in (14), as a single-word, verbless utterance that is now clearly distinguished from the previous sentence using standard punctuation. Therefore, in (13), the adverb is not part of the interrogative, even if superficially it looks like it is, due to the punctuation here used.

Adopting the classification principles of the Lexicon-Grammar framework (Gross, 1986, pp. 97–99), these verbless, single-word utterances could have been classified as *frozen sentences* (Antunes et al., 2025, class

[10] The authors wish to thank the anonymous reviewers who called our attention to such phenomena.

C0E), within the same category that includes other interjective/exclamative expressions. However, as the author notes, this represents a borderline case of the adverb category. Therefore, we have followed M. Gross (1986) perspective and treated these forms as adverbs (class PADV), representing them in the Lexicon-Grammar with an exclamation mark (*francamente!*).

3. The broader picture: building the lexicon-grammar of multiword adverbs in Portuguese

This study forms part of a broader research initiative aimed at constructing a comprehensive lexicon of multiword adverbs in Portuguese. The analysis is conducted within the Lexicon-Grammar theoretical and methodological framework developed by Maurice Gross (1975, 1981, 1986), which is itself rooted in Zellig S. Harris's (1991) Transformational Operator Grammar. In this framework, the elementary sentence is considered the primary unit of linguistic analysis, as both the syntactic properties and the semantic interpretation of linguistic expressions can only be fully characterized within this minimal context. Maurice Gross (1986, 1996) proposed a formal classification of adverbial expressions based on the notion of *adverbe généralisé* (ing. *generalized adverb*), which encompasses both compositional, syntactically transparent expressions and frozen, idiomatic, non-compositional forms. His approach relies on a taxonomic method that organizes expressions according to their internal grammatical structure, following the base sequence *Prep Det N Modif*, where *Prep* denotes a preposition, *Det* a determiner, *N* a nominal element, and *Modif* a modifier. In Gross (1986, pp. 93–100), the author revisits the Harrissian analysis of adverbs that have scope over different types of sentences (declarative, interrogative, or imperative), interpreting them as reductions of the corresponding performative verbs. Consequently, the distinction between the scope over a sentence ("portée sur une phrase") and the scope over verbs ("portée sur des verbes") is viewed as merely terminological.

Building on this theoretical foundation, the present study applies Gross's classification model to a large set of data, resulting in the identification of approximately 3,700 multiword adverbs in Portuguese, encompassing expressions from both European Portuguese (EP)—some of which were previously documented by Palma (2009)—and Brazilian Portuguese (BP), as described by Müller et al. (2022, 2023). The lexicon incorporates 12 out of the 16 formal subclasses proposed by Maurice Gross (1986). Table 1 provides an updated overview of the distribution across these formal classes, detailing the internal structure of each adverbial construction, accompanied by representative examples, and presenting the number of occurrences in EP, in BP, those shared by both varieties (EP-BP), and the overall total.

Table 1

Current Distribution of the Lexicon by Formal Classes in Portuguese

Formal Class	Internal Structure	Examples	EP	BP	EP-BP	Total
PC	*Prep C*	*em vão* 'in vain'	28	256	673	957
PDETC	*Prep Det C*	*pelo menos* 'at least'	56	227	526	809
PAC	*Prep Adj C*	*de má vontade* 'unwillingly'	11	46	232	289
PCA	*Prep C Adj*	*por maioria absoluta* 'by absolute majority'	22	69	269	360
PCDC	*Prep C1 de C2*	*por conta da casa* 'on the house'	22	81	211	314
PCPC	*Prep C1 Prep C2*	*da cabeça aos pés* 'head to toes'	46	103	269	418
PCONJ	*Prep C1 Conj C2*	*em verso e prosa* 'in verse and prose'	10	75	169	254
PF	*Frase fixa*	*dito isso* 'this said'	2	43	89	134
PV	*Prep V W*	*até dizer chega* lit.: 'until say enough', 'a lot'	4	4	27	35
PJC	*Conj C*	*e por aí vai* 'and so on'	3	48	33	84
PACO	*<ADJ> como C*	*<surdo> como uma porta* 'deaf as a door'	0	7	3	10
PVCO	*<V> como C*	*<trabalhar> como uma mula* 'work like a mule'	0	26	25	51
		Total	**204**	**985**	**2,526**	**3,715**

In addition to this formal classification, we adopt the syntactic-semantic framework proposed by Molinier and Levrier (2000), which distinguishes two major types of adverbs: (i) internal modifiers (M), which target specific components within the sentence and are divided into seven subclasses, and (ii) external modifiers (P), which operate at the propositional level and are organized into six subclasses. Table 2 presents the current syntactic-semantic distribution of multiword adverbs in our lexicon, organized into two major groups and their respective subclasses. It also provides figures for expressions exclusive to each variety (EP and BP), as well as those common to both (EP-BP). The totals in Table 2 slightly exceed those in Table 1, as some expressions occur in more than one syntactic-semantic class.

Table 2

Current Semantic Distribution of Multiword Adverbs in Portuguese

Class	Subclass	Examples	EP	BP	EP-BP	Total
PC conjunctive		*afinal de contas* 'after all'	3	44	203	250
PS disjunctive of style		*com toda a franqueza* 'in all honesty'	4	9	74	87
PA disjunctive of attitude						
PAa	evaluative	*por pura sorte* 'by sheer luck'			21	21
PAm	modal	*com certeza* 'certainly'	1	10	25	36
PAs	subject-oriented	*pelo meu lado* 'for my part'		5	1	6
PAh	habit	*de costume* 'usually'		3	15	18
MV manner verb-oriented		*a todo o gás* 'at full throttle' *de cabo a rabo* 'from beginning to end'	221	754	1,469	2,444
MS manner subject-oriented		*de boa fé* 'in good faith' *sem pestanejar* 'without hesitation'	3	21	103	127
MT time						
MTd	date	*a horas mortas* 'at dead of night'	24	77	309	410
MTf	frequency	*dia sim dia não* 'every other day'	5	24	55	84
MTu	duration	*anos a fio* 'for years on end'		12	35	47
MP point of view		*na prática* in practice'			10	10
MQ quantitative		*aos montes* 'in abundance'	5	63	122	190
MF focus		*em especial* 'especially'		10	19	29
ML locative		*nos confins do mundo* 'at the ends of the world'	3	80	121	204
Total			**269**	**1,111**	**2,582**	**3,962**

Our investigation has identified 87 multiword disjunctive style adverbs in Portuguese. Their distributions are as follows: 4 exclusives to European Portuguese (signaled with the index '^{PE}'), 9 expressions exclusive to Brazilian Portuguese (signaled with the index '^{PB}'), and 74 shared by both varieties. To this we add 9 simple-word adverbs ending in *-mente*, common to both varieties. The list of these adverbs is provided in the Annex. Below, we present a tentative semantic classification of these adverbs.

Request: *por favor* 'please', *por especial favor*[11] 'please', *por gentileza*^{PB} 'kindly', *por obséquio* 'please', *por Deus* 'lit.: by God, please', *por tudo o que há de mais sagrado* 'for all that is most sacred', *por/pelo amor de Deus* 'for God's sake', *se faz/fizer favor* 'please/if you please';

Manner of speaking: *concretamente* 'specifically', *francamente* 'frankly', *honestamente* 'honestly' *literalmente* 'literally', *modestamente* 'modestly', *seriamente* 'seriously', *sinceramente* 'sincerely'; *com o devido respeito* 'with all due respect', *com franqueza* 'frankly', *com toda a franqueza* 'in all honesty',

[11] Although the expression *por favor especial* is grammatically correct, it is observed that the form with the adjective preceding the noun—*por especial favor*—is much more common, both in European Portuguese and Brazilian Portuguese, with 20 and 11 occurrences respectively in the ptTenTen23 corpus, compared to 0 occurrences in PT and 1 in BR of the form with the adjective following the noun. For this reason, we classify this expression as a compound adverb, class PAC.

com toda a sinceridade 'in all sincerity', *falando francamente*/*francamente falando* speaking 'frankly/frankly speaking', *modéstia à parte* 'modesty aside', *para ser sincero*[PB] 'to be honest';

Clarification: *a bem da verdade* 'to tell the truth'*, a bem dizer* 'strictly speaking', *a propósito* 'by the way', *a rigor* 'strictly speaking', *em abono da verdade* 'in all honesty', *em linhas gerais* 'generally speaking', *em bom português* 'in plain Portuguese', *em português claro* 'in clear Portuguese', *em rigor* 'strictly speaking', *em*/*no sentido* (*mais*) *estrito*/*restrito* (*do termo*/*da palavra*) 'strictly speaking', *em termos gerais* 'in general terms', *em termos simples* 'in simple terms', *em toda a extensão da palavra* 'in the full sense of the word', *na acepção*[PB]*/aceção*[PE] *da palavra* 'in the meaning of the word', *na extensão da palavra*[PB] 'in the full sense of the word', *para seu governo* 'for your information', *para todos os efeitos* 'for all intents and purposes', *por assim dizer* 'so to speak', *por bem dizer* 'so to speak', *por outras palavras* 'in other words', strictu sensu 'strictly speaking';

Validation: *fora de brincadeira* 'no kidding', *na boa*[PB] 'honestly', *na moral*[PB] 'honestly', *para abreviar razões*[PE]*,* 'to cut a long story short' *para começo de conversa,* 'to begin with' *para dizer a verdade* 'to tell you the truth', *para dizer/falar depressa e bem*[PE] 'to put it bluntly' *sem brincadeira*[PB]*,* 'no kidding', *sem exagero* 'without exaggeration', *verdade seja dita*[PB] 'truth be told';

Concessive: *salvo erro ou omissão* 'unless otherwise stated'*, salvo indicação contrária,* 'unless otherwise indicated', *salvo indicação em contrário,* 'unless otherwise indicated', *salvo melhor juízo,* 'tentatively, pending better judgment', *salvo melhor entendimento* 'unless otherwise interpreted', *salvo melhor opinião* 'unless there is a better opinion';

Self-reference concessive: *até onde sei/sabemos* 'as far as we know', *até onde é do meu/nosso conhecimento* 'to the best of my/our knowledge', *pelo que me consta*[PB]*,* 'as far as I know', *pelo que sei* 'as far as I know'; *salvo engano,* 'unless I am mistaken', *salvo erro* (*da minha parte*) ' unless I am mistaken', *salvo erro de memória* 'unless my memory fails me', *se a memória não me falha* 'if my memory serves me well/correctly', *se não me falha a memória* 'if my memory does not fail me', *se bem me lembro/recordo* or *se me lembro/recordo bem* 'if I remember correctly/well', *tanto quanto me lembro/recordo,* 'as far as I remember';

Irony: *entre aspas*, 'in quotation marks', *salvo seja* 'except be';

Conclusion: *resumidamente,* 'in short', *sumariamente* 'briefly'; *em breves palavras* 'in a nutshell'*, em duas palavras* 'in two words'*, em poucas palavras* 'in a few words'*, em suma* 'in short', *em resumo* 'in summary', *em síntese* 'in synthesis', *numa palavra* 'in a word' (see Section 4., below);

Secrecy (gossip): *aqui entre nós,* 'between us' *cá entre nós* 'between us';

The list of adverbial expressions presented above illustrates the richness and complexity of this class in Portuguese. The proposed semantic organization should be viewed as a preliminary attempt to systematize these expressions. It is noteworthy, however, to specify certain expressions regarding their distribution across both varieties. For instance, *por gentileza* 'please', while understood in European Portuguese, is more common and significantly more frequent in Brazilian Portuguese. Similarly, *por obséquio* 'please, kindly', tends to occur in more formal or even outdated registers in European Portuguese, while in Brazilian Portuguese it remains active and without any formality in register. The expression *na/numa boa* 'easily' presents an interesting case of functional variation: while in both EP and BP it maintains a more literal meaning as a manner adverb (modifying the verb) in both varieties,

(15) *O Pedro fez isso* na boa
'Pedro did it easily'

In Brazilian Portuguese, it also functions as a PS, such as in the following example (16), in which we can clearly place the performative metalinguistic operator *eu digo*:

(16) (Eu digo) Na boa*, quem entendeu errado foi você!*[12]
'Honestly, you're the one who got it wrong!'

As for the expression *a bem da verdade* 'to tell the truth', it is attested in both European and Brazilian Portuguese, whereas the variant *a bem verdade* appears to be exclusive to Brazilian Portuguese. In Silva (2013), the form *a bem verdade* is listed within the context of Brazilian Portuguese, and we consider it equivalent in meaning and function to *a bem da verdade*. The *ptTenTen23* corpus confirms this observation, showing multiple occurrences of *a bem verdade*, exclusively in Brazilian usage.

A particularly interesting subgroup within these classes involves concessive expressions, which can be compared to those that are self-referential in nature. The main difference lies in who or what the concessive element refers to. In expressions like *salvo indicação em contrário* 'unless otherwise stated', the concessive clause doesn't necessarily point back to the speaker; it can refer to some other entity (*salvo indicação minha/de outra pessoa em contrário* 'unless otherwise indicated by me/another person'). On the other hand, expressions like *salvo engano* (*meu*) 'unless I am mistaken', are clearly self-referential, tied to the speaker as the subject of the utterance. This contrast is similar to what we see in examples such as

(17) *Eu digo que isso foi assim* salvo erro (*meu*/**teu*)
'I tell you that that was like that, unless (I'm/*you are) mistaken'

(18) *Eu digo que isso foi assim* até onde (*eu sei*/**tu sabes*)
'I tell you that that was like that as far as (I/*you) know

The expression *salvo seja* 'save be' functions as a marker that refers back to the speaker's own utterance, often signaling that a term or expression has been used in a possibly inappropriate, or non-literal, way. In other words, it invites the listener to interpret the marked expression with caution. It is very similar to the expression *entre aspas* 'in quotation marks', in which its oral usage is frequently reinforced by accompanying gestures with two stretched fingers of each hand representing the quotation marks.

4. Distinction between conjunctive and disjunctive style adverbs

The distinction between conjunctive adverbs (Müller et al., in press), noted (PC), and disjunctive style adverbs (noted PS) is not absolute. Molinier and Levrier (2000) argue that PS adverbs share similarities with some *Reformulative* and *Conclusive* classes of conjunctive adverbs.

> Certains adverbes, du type reformulatif (cf. *en bref*, *en résumé*, *en conclusion*, *en d'autres termes* ...), répondent à la fois à la définition des adverbes conjonctifs et des adverbes disjonctifs de style. En raison de l'ordre institué dans le principe de classement, ces adverbes ont été versés dans la classe des conjonctifs. Sémantiquement, ces adverbes indiquent un certain type de rapport entre le locuteur et son destinataire ou l'énoncé lui-même, ou encore, ils formulent un commentaire du locuteur sur la forme de l'énoncé (Molinier & Levrier, 2000, p. 66).[13]

[12] Sketch Engine Token number: 11726755654. https://www.forumcarros.com.br/index.php?/profile/3624-everton-florentino-bueno/content/&type=forums_topic_post&change_section=1

[13] Translation: Some adverbs of the reformulative type (cf. *en bref*, *en résumé*, *en conclusion*, *en d'autres termes*...) meet the definition of both conjunctive adverbs and disjunctive adverbs of style. Because of the order instituted in the principles of classification, these adverbs have been placed in the conjunctive class. Semantically, these adverbs indicate a certain type of relationship between the speaker and the

The emphasis of the disjunctive style is not so much on the content of an utterance but on how something is said. In Portuguese, we observed that a limited number of multiword adverbs may be accommodated in both classes, PC and PS. Some semantic sub-classes of conjunctive adverbs (PC), particularly *Contrastive and Conclusive* and *Order and Rank*, share certain syntactic features with disjunctive style adverbs (PS). Semantically, these adverbs indicate a type of relation between the speaker and the recipient or the statement itself, or they provide the speaker's commentary on the form of the statement. Some examples of these cases include:

i. Contrastive: *só para contrariar* 'just to contradict', *só para variar* 'just for a change';

ii. Conclusive: *em definitivo* 'definitely', *em resumo*, *em síntese*, *em suma*, 'in short', *em breves termos* 'in brief';

iii. Order and Rank: *antes de mais nada* 'first of all', *em primeiro lugar* 'firstly'.

A key property of conjunctive adverbs is their inability to appear at the absolute beginning of a discourse, though they might appear at an initial position of a clause. However, their function is to establish a nexus between different parts of the discourse, which might be a previous sentence, paragraph, or even other parts of the text, as, for instance, *em poucas palavras* 'in a few words', in the next example, which we interpret with a **reformulative** value:

(19a) [Sugerir uma linha de ação claramente definida que facilite a sistematização de importantes dados e informações que vêm sendo construídas sobre os esportes na natureza, mas, em geral, de maneira mais ou menos dispersa.] ***Em poucas palavras****, é essa a intenção deste trabalho*[14]
'[To suggest a clearly defined course of action that facilitates the systematization of important data and information that has been developed on nature sports, but which, in general, remains more or less scattered] In a few words, that is the intention of this work'

Still, *em poucas palavras* 'in a few words' can also be classified as a PS since it may be associated to a performative 'say' verb signaling how the content of the utterance is said. These types of sentences can be understood as a reduced version of the performative verb 'say', where this verb has been omitted, such as in the following example:

(19b) (*Eu digo*) *Em poucas palavras* (*que*) *é essa a intenção deste trabalho*
'(I say) In a few words (that) that is the intention of this work'

Disjunctive of style adverbs (PS) typically appear at the beginning of a sentence, as in (19a). However, there is nothing to prevent them from also appearing at the initial position of a discourse. Based on the work of Molinier and Levrier (2000), our classification system for multi-word adverbs also assigns precedence to PC adverbs over PS adverbs. In our lexicon, we have identified 18 PC multi-word adverbs that have similar characteristics to PS adverbs. To prevent multiple and redundant entries in the lexicon-grammar, PC adverbs holding a PS behaviour were given a feature indicating that, for those entries, this additional semantic value could be expressed.

addressee or the utterance itself, or else they formulate a comment by the speaker on the form of the utterance (Molinier & Levrier, 2000, p. 66; our translation).

[14] Sketch Engine Token number: 45300794. https://www.cev.org.br/biblioteca/por-um-programa-investigativo-para-os-esportes-natureza/

5. Relation between style disjunctive (PS), verb-oriented (MV), and subject-oriented (MS) manner adverbs

In our previous analysis, we compared conjunctive and disjunctive style adverbs within a framework adapted from Molinier and Levrier (2000). This comparison led to the identification of additional multiword adverbs whose properties can be found in more than one subclass. In this section, we examine the relationships between subject-oriented adverbs (MS), manner adverbs (MV), and disjunctive style adverbs (PS).

As previously discussed, adverbs can be grouped into two broad categories: sentence-internal modifiers (M) and sentence-external modifiers (P). Within the sentence-internal modifier category, two relevant subclasses—manner adverbs (MV), which typically modify the verb by indicating how an action is performed, and subject-oriented adverbs (MS), which qualify not only the way the subject acts, but also qualify the human agent—share certain features with disjunctive style adverbs (PS). While PS adverbs primarily convey the speaker's stance or attitude toward the proposition, rather than the manner of the action itself, the boundaries between these categories are not always clear-cut. This section seeks to clarify these overlaps by outlining a set of distinguishing properties.

Depending on their syntactic position and the type of relation they establish—either with specific sentence constituents or with the sentence as a whole—adverbs typically classified as manner (MV) or subject-oriented (MS) may also function as disjunctive style adverbs (PS). Some illustrative examples include:

(i) MV (20) and PS (21) *em português claro* 'in plain Portuguese'

(20a) *Ela explicou em português claro todas as etapas do trabalho*
'She explained all the stages of the work in plain Portuguese'

(20b) **Em português claro, ela não explicou todas as etapas do trabalho*
'In plain Portuguese, she did not explain all the stages of the work'

(20c) *Foi em português claro que ela explicou todas as etapas do trabalho*
'It was in clear Portuguese that she explained all the stages of the work'

(21a) *Em português claro, este plano* (*não*) *vai dar certo*
= [*Eu digo*] *em português claro que este plano* (*não*) *vai dar certo*
'In plain Portuguese, this plan will (not) work'
'I say in plain Portuguese that this plan will (not) work'

(21b) **É em português claro que este plano não vai dar certo*.
'It is in clear Portuguese that this plan won't work.'

(ii) MS (22) and PS (23) *com todo o respeito* 'with all due respect'

(22a) *O jovem dirigiu-se ao presidente com todo o respeito*
'The young man addressed the president respectfully'

(22b) **Com todo o respeito, jovem não se dirigiu ao presidente*
'With all due respect, the young man did not address the president'

(22c) *Foi com todo o respeito que o jovem se dirigiu ao presidente*
'It was with all due respect that the young man addressed the president'

(23a) *Com todo o respeito, o presidente está se contradizendo*

'With all due respect, the president is contradicting himself'

(23b) *Com todo o respeito, o presidente* **não** *está se contradizendo*
'With all due respect, the president is not contradicting himself'

(23c) **É com todo o respeito que o presidente se está contradizendo*
'It is with all due respect that the president is contradicting himself'

(23d) *Eu digo com todo o respeito que o presidente está se contradizendo*
'I say with all due respect that the president is contradicting himself'

Additional examples of multi-word MV adverbs related to PS include: *em breves termos* 'in brief terms', *de brincadeira* 'playfully' *sem brincadeira*, *fora de brincadeira* 'no kidding', *na moral*, a sério 'seriously'. In the same way, MS adverbs with a PS parallel construction include: *sem meias palavras*, 'without mincing words' *sem mais delongas* 'without further ado'.

To avoid unnecessary duplication of entries, we adopted a strategy of adding these MV and MS adverbs, a feature that indicates they can also enter the PS construction.

6. PS adverbs on other performative metalinguistic operators

As previously noted, disjunctive style adverbs possess the capacity to modify performative verbs such as *eu digo* ('I say'), indicating how a statement is made. These verbs carry a metalinguistic function and can, in actual usage, be omitted—leaving the adverb in a sentence-initial, detached position to signal the speaker's stance independently. Our study has identified several PS adverbial expressions that may also be associated with other types of performative verbs beyond *eu digo*. These alternative verbs likewise operate metalinguistically, introducing how the speaker performs the speech act. Examples include imperative verbs such as *pedir* ('to ask/request'), *mandar* ('to order'), *suplicar* ('to beg'), and *implorar* ('to implore'), as illustrated in the following examples:

(24a) (*Eu peço-te*/*suplico*/*imploro-te*) *por*/*pelo amor de Deus*, *que não uses drogas!*[15]
'I beg/plead/implore you, for God's sake, do not use drugs'

(24b) *Por*/*pelo amor de Deus*, *não uses drogas!*
'For God's sake, don't use drugs!'

The same verbs above may be found with the following expressions:

- *por favor*, *se faz*/*fizer favor*, *por gentileza*[PB] 'please'
- *por*/*pelo amor de deus* 'for/by the love of god'[16]

[15] Instead of the European Portuguese *tu* , here used, in Brazilian Portuguese, the *você* form of subject would be preferable:
[BP]*Eu peço a você pelo amor de Deus que não use drogas! Pelo amor de Deus, não use drogas!*

[16] In the ptTenTen23 (PT), the expression *por deus* 'by God' is rare, and it seems either associated to *jurar* 'swear' (e.g., *Documentos?... Ou a fonte é "juro* por deus*" ??*) or an exclamative value (e.g., *Mas,* por deus*, que convertable era aquele?*). No imperative (request) example was found, In the BP veriety, the exclamative value seems to be more frequent, e.g., Por deus*, nunca pensei que eu estaria tão feliz como naquele momento.* Por deus*, o que ele tinha nas sacolas?* Por deus*, eu estou surtando. E,* por deus*, era tudo que ele queria fazer naquele momento.* The use with *jurar* 'swear'; e.g., *Juro por deus, ta lá,pode ver.* and the imperative use: *Pelo diabo ou por deus, manda às favas a gramática.* are all attested. This exclamative (multiword) expression is to be considered in the same way either as the exclamative adverb *fancamente*, mentioned above, or as an exclamative frozen sentence (class C0E).

- *por tudo o que há de mais sagrado* 'for all that is most sacred'
- *pela minha/tua* (*rica*) *saúde* 'for my/your (rich) health'
- *por alma de quem lá tens* 'for the soul of whoever you have there'

It is worth noticing that the anaphoric reference of the key elements within these expressions may vary. While in *por favor* 'please' one could establish a paraphrase with *tu fazes-me um favor a mim*, 'you do me a favor', where the addressee is the subject; in *pela minha/tua saúde* 'for my/your health' the noun *saúde* 'health' can refer both to the speaker and the addressee; a similar remark could be made regarding the variants of expression *por tudo o que tenho/tens de mais sagrado* 'for all that is most sacred'; on the contrary, in *por alma de quem lá tens* 'for the soul of those you have there(=heaven)' the noun *alma* 'soul' can only refer to the addressee; finally, in *por amor de deus* 'for the love of god' as the reference of *amor* is already explicitly encoded, no variation seems to be possible, e.g. **por amor do meu/teu deus* 'for the love of my/your god'.

Other constructions may require different types of performative verbs, such as *jurar* 'swear' and *prometer* 'promise':

(25a) (*Eu* ***juro***) *por tudo o que é/há de* (*mais*) *sagrado que* [*eu*] *não contei o seu segredo*
'(I swear) by all that is sacred, that I didn't tell your secret'

= (25b) *Por tudo o que é/há de* (*mais*) *sagrado, eu não contei o seu segredo*
By all that is sacred, that I didn't tell your secret'

(26a) (*Eu* ***prometo***) *por tudo o que é/há de* (*mais*) *sagrado que* [*eu*] *não contarei o seu segredo*
'(I promise) by all that is sacred, that I will not tell your secret'

= (26b) *Por tudo o que é/há de* (*mais*) *sagrado, eu não contarei o seu segredo*
'By all that is sacred, that I will not tell your secret'

The use of the past tense in (25a) and (25b) in *contar* 'tell' clearly indicates the pragmatic value of the performative *jurar* 'swear'. In (26a) and (26b), the performative *prometer* 'promise' requires the future tense, the past tense being unacceptable (**Eu prometo que contei* 'I promise I told you'). However, the verb *jurar* can also be used in the same way as prometer, and in that interpretation the future tense is perfectly acceptable (*Eu juro que contarei* 'I swear that I will tell' ≈ *Eu prometo que contarei* 'I promise that I will tell'). Still, there does not appear to be a clear-cut distinction in the distribution of these performative verbs in relation to the specialized adverbs, as both verbs seem to be used interchangeably.

7. Related work

This section compares key descriptive approaches to disjunctive adverbs in French and English to the emerging analysis of these adverbs in Portuguese. While the syntactic-semantic notion of disjunctive adverbs (PS) is recognized across languages, the criteria for classification, the range of adverbs, and the terminology show variation. References to these types of adverbs have been previously made in the literature, providing support for our analysis. These prior works provide a foundation. However, the specific behavior of Portuguese *-mente* adverbs and multi-word expressions remains under-described. This research aims to fill that gap.

As mentioned before, we adopt the syntactic classification leading to the constitution of the class of *disjunctive style adverbs* (PS) as they have been originally described by Molinier and Levrier (2000) in reference to French compound adverbs and single-word adverbs ending in *-ment* '-ly.' The corresponding expressions operate in Portuguese in a very similar way. More recently, Álvarez-Prendes (2021) described a subset of eight of these French adverbs and called them *adverbes d'enunciation*, while trying to describe their diachronic

evolution. Earlier on, Guimier (1996) described them as *adverbes illocutifs,* as the illocutionary act is an element of the speaker's intention. Guimier also expands on an interesting point about these adverbs not necessarily forming a class, as some of them are *polysemous* and may operate in more than one function, sometimes performing as manner/quantifier internal modifiers or as external illocutionary markers. He explains the polyfunctionality with the following examples of where the adverb *honnêment* 'honestly' is used (Guimier, 1996, p. 158):

An illocutionary adverb:

(27a) *Aujourd'hui encore, je ne peux, honnêment, me résoudre à trouver une diversion de la lâcheté* (Abellio)
'Even today, I can't honestly bring myself to find a cowardly diversion'

An adverb of manner, oriented towards the subject:

(27b) *J'aurais voulu mourir honnêment, comme mon père et ma mère* (Pourrat)
'I wanted to die honestly, like my father and mother'

An adverb modifying a verb:

(27c) *Vivre honnêment, qu'est-ce que ça veut dire?* (Abellio)
'Live honestly, what do you mean?

An adverb intensifying an adjective:

(27d) *Elle était touchante, la jolie Elodie, honnêment rondelette dans son blouson de tolie à manches courtes* (Sabatier)
'She was touching, pretty Elodie, honestly plump in her short-sleeved tolie jacket'

Irrespective of the precise interpretation assigned to some of these examples[17], it is clear that the illocutionary adverb exhibits a distinct meaning and syntactic behavior compared to other uses. Within the framework adopted for the description of such cases in Portuguese, we treat them as instances of *homonymy* — that is, distinct lexical-grammatical units with different meanings and constructions, despite sharing the same surface form.

Quirk *et al*. (1985, p. 615) describe *disjuncts* as having a superior role as they are syntactically more detached and seem to have a scope that extends over the sentence as a whole. *Style disjuncts* adverbs convey the speaker's comments and are separated into two types, namely:

i. Modality and Manner: *candidly*, *honestly*, *seriously*, *strictly*, *truly*, *truthfully*, *confidentially*

ii. Respect: *figuratively*, *generally*, *literally*, *personally*, *strictly*.

It is under the 'Respect' type that the authors include the longer phrases that include *generally speaking*, *strictly speaking,* and also other forms of corresponding *-ly* ending adverbs, which include examples such as *personally*: *to be personal*, *bluntly*, *if I may be blunt*, or *bluntly speaking*.

Style stance adverbials (Biber et al., 1999, p. 854), which correspond to style disjuncts, are one of the three major semantic classes of *stance adverbials* proposed by this author, that include *epistemic*, *attitude,* and *style*.

[17] In our view, the use of *honnêment* 'honestly' in examples (27b-c) is the same. The quantifying interpretation in (27d) is unclear.

Sentence-initial adverbials are the term used by Larsen-Freeman and Celce-Murcia (2016, pp. 516–517) for the group of adverbs related to speakers' attitude, while adopting four functional categories proposed by Halliday and Matthiessen (2014), referring to four aspects of speaker attitude:

i. Probability: *maybe*, *perhaps*, *certainly*, *surely*

ii. Usuality: *usually*, *generally*, *typically*, *occasionality*

iii. Presumption: *of course*, *obviously*, *clearly*, *evidently*

iv. Desirability: (*un*)*fortunately*, *luckily*, *hopefully*, *regrettably*

In European Portuguese, Costa (2008) defines the classes of *advérbios avaliativos orientados para o falante* 'speaker-oriented evaluative adverbs', supporting that although there are small differences between the attitude or the expression of a judgment on the part of the speaker, these adverbs may be grouped since they all convey a predication about the speaker. Raposo (2013, p. 1666), on the other hand, places this category of *advérbios de atitude do falante* 'speaker attitude adverbs' under a major group of *evaluative adverbs* that include other subclasses such as *evaluative of the situation*, *modal and intentional adverbs*, *adverbs restricting the domain of propositional content*, and *adverbs exhorting the listener* as they have the property of expressing the speaker's evaluations.

In Brazilian Portuguese, disjunctive adverbs of style ending in *-mente* have already been the object of a preliminary study (Fernandes, 2011), who listed several adverbs as PS from the 1,000 most frequent adverbs. These include the following (with their classification/comment):

concretamente 'concretely', PS

convincentemente 'convincingly', Not PS

decididamente 'definitely', PS

falsamente 'falsely', Not PS

honestamente 'honestly', PS

literalmente 'literally', PS

modestamente 'modestly', PS & MS

pessoalmente 'personally', PS

objetivamente 'objectively' PS

realmente 'really', Not PS

simplesmente 'simply', PS

sinceramente 'sincerely', PS

veementemente 'vehemently', Not PS

verdadeiramente 'truly', PS

Some of these adverbs cannot be considered PS as they do not enter the definitory construction with the performative *Eu digo* (or another performative metalinguistic operator) and the equivalent reduced form with the detached adverb, that is, they do not allow for the equivalence: *Eu digo Adv que isso é assim = Adv, isso é assim*. This is the case of *convincentemente*, *falsamente*, *pessoalmente*, *realmente*, and *veementemente*.

To sum up, and based on the descriptions above, we understand that PS adverbs in Portuguese exhibit shared features. They typically, or most frequently, appear detached at the beginning of the sentence and separated by a comma, although they can also appear at the end of the sentence, preceded by a comma; otherwise, they can appear detached by commas from the sentence they operate on. They have a scope on and modify the entire sentence. And, finally, they express the speaker's attitude towards the content of the utterance. Importantly, they contrast with manner adverbs, which directly modify verbs and do not comment on the utterance as a whole. This distinction helps refine the categorization of Portuguese adverbs within a broader typological framework. Our contribution, in this paper, lies mainly in the classification of multi-word adverbs in Portuguese (both European and Brazilian varieties). To the best of our knowledge, this is the first study to systematically compile a lexicon

of these multi-word adverbs, providing a clear and distinct syntactic and semantic classification within the framework of disjunctive style adverbs.

8. Conclusion and future work

In this paper, we examined the syntactic and semantic properties of a specific subset of Portuguese multiword adverbs, namely the disjunctive adverbs of style (PS). Within our current lexicon, which includes approximately 3,700 multiword expressions, we identified 87 expressions that fall into this category. For this paper, we also included 9 single-word adverbs ending in *-mente*. PS adverbs function as sentence-external modifiers, conveying the speaker's attitude toward the propositional content of the utterance. Some of these expressions may present other word senses, sharing features with conjunctive adverbs (PC), verb-oriented manner adverbs (MV), and subject-oriented manner adverbs (MS). We presented a set of syntactic and semantic criteria to distinguish PS adverbs from these overlapping categories. Furthermore, we proposed a tentative semantic classification of PS adverbs, organizing the main semantic types found within this class. This classification includes categories such as *request*, *manner of speaking*, *clarification*, *irony*, *validation*, *concessive*, *self-reference concessive*, and *conclusion*. These categories reflect the diverse ways in which speakers use PS adverbs to position themselves toward their utterances. Additionally, we identified performative verbs beyond *dizer* 'say', that appear in constructions with PS adverbial expressions and support their illocutionary and metalinguistic functions.

As part of future work, we intend to examine how these adverbs behave in authentic spoken language, using the *Roda Viva* corpus (Souza de Miranda *et al.,* 2024), an extensive collection of transcribed television interviews in contemporary Brazilian Portuguese. The spontaneous nature of these interviews provides an ideal environment to investigate the roles and discourse functions of PS adverbs in context. This corpus-based study will contribute to a deeper understanding of how these expressions operate in real-life communicative settings.

Acknowledgements

This work was partially supported by the Foundation for Science and Technology (FCT) (Project Reference DOI: 10.54499/UIDB/50021/2020) and by the Doctoral Program in Language Sciences of the University of Algarve.

Referências

Álvarez-Prendes, E. (2021). (Inter)subjectivisation et chaînes sémantiques dans les adverbes français en *–ment* : Analyse du passage de l'adverbe intégré à la proposition au marqueur discursif. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *66*(3), 346–373. https://doi.org/10.1017/cnj.2021.17

Antunes, D., Baptista, J., & Mamede, N. (2025). Processamento automático de expressões idiomáticas do português europeu. *Linguamática*, *17*(1), 55–73. https://doi.org/10.21814/lm.17.1.464

Baptista, J. (2005). *Sintaxe dos predicados nominais com* ser de. Fundação Calouste Gulbenkian/ Fundação para a Ciência e a Tecnologia.

Baptista, J. (2010) *Verba dicendi*: A structure looking for verbs. In T. Nakamura, É. Laporte, A. Dister, & C. Fairon (Eds.), *Les tables. La grammaire du français par le menu. Mélanges en hommage à Christian Leclère* (Cahiers du CENTAL 6, pp. 11–20). CENTAL.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.

Costa, J. (2008). *O advérbio em português europeu*. Colibri.

Fernandes, G. (2011). *Automatic disambiguation of* -mente *ending adverbs in Brazilian Portuguese* [Master's thesis, Universidade do Algarve & Universitat Autònoma de Barcelona). Sapientia – UAlg Scientific Repository. http://hdl.handle.net/10400.1/12191

Guimier, C. (1996). *Les adverbes du français : Le cas des adverbes en -ment*. Ophrys.

Gross, M. (1975) *Méthodes en syntaxe*. Hermann.

Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, *63*, 7–52.

Gross, M. (1986) *Grammaire transformationnelle du français : 3 - Syntaxe de l'adverbe*. ASSTRIL.

Gross, M. (1996) Lexicon-Grammar. In K. Brown & J. Miller (Eds.), *Concise encyclopedia of syntactic theories* (pp. 244–259). Pergamon.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Edward Arnold.

Harris, Z. S. (1976). *Notes du cours de syntaxe* (M. Gross, Trans.). Seuil.

Harris, Z. S. (1991). *A theory of language and information: A mathematical approach*. Clarendon Press.

Kilgarriff, A., Jakubíček, M., Pomikalek, J., Sardinha, T. B., & Whitelock, P. (2014). *PtTenTen*: A corpus for Portuguese lexicography. In T. Sardinha & T. Santos (Eds.), *Working with Portuguese corpora* (pp. 111–130). Bloomsbury.

Larsen-Freeman, D., & Celce-Murcia, M. (2016). *The grammar book: Form, meaning, and use for English language teachers* (3rd ed.). National Geographic Learning.

Molinier, C., & Levrier, F. (2000). *Grammaire des adverbes : Description des formes en* -ment. Librairie Droz.

Müller, I., Mamede, N., & Baptista, J. (2022). Bootstrapping a lexicon of multiword adverbs for Brazilian Portuguese. In *International Conference on Computational and Corpus-Based Phraseology* (pp. 160–174). Springer.

Müller, I., Mamede, N., & Baptista, J. (2023). Advérbios compostos do português do Brasil. *Revista da Associação Portuguesa de Linguística*, *1*(10), 230–250. https://doi.org/10.26334/2183-9077/rapln10ano2023a13

Müller, I., Baptista, J., & Mamede, N. J. (in press). *Conjunctive multiword adverbs in Portuguese*. In *Discourse markers in Romance languages: DISROM-8*.

Palma, C. (2009). *Expressões fixas adverbiais: descrição léxico-sintáctica e subsídios para um estudo contrastivo português-espanhol* [Master's thesis, Univresidade do Algarve]. Sapientia – UAlg Scientific Repository. http://hdl.handle.net/10400.1/428

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

Raposo, E. P. (2013). Advérbio e sintagma adverbial. In E. P. Raposo, M. Barbosa, M. A. Martins, A. M. Brito, I. Duarte, & I. Ribeiro (Eds.), *Gramática do português* (Vol. 2, pp. 1569–1675). Fundação Calouste Gulbenkian.

Silva, J. P. D. (2013). *Dicionário brasileiro de fraseologia* (preliminary online version, currently not available).

Souza de Miranda Jr., I., Wick-Pedro, G., Dias de Barros, C., & Vale, Oto A. (2024). *Roda Viva* boundaries: An overview of an audio-transcription corpus. In *3rd Workshop on Digital Humanities and Natural Language Processing* (pp. 165–169). Universidade de Santiago de Compostela. https://aclanthology.org/2024.propor-2.22.pdf

Annex

List of disjunctive adverbs of style in Portuguese

Adverbs exclusive of the European (PT) or Brazilian (BR) variety are marked as such.

concretamente 'concretely'
honestamente 'honestly'
francamente 'frankly'
literalmente 'literaly'
modestamente 'modestly'
resumidamente 'in short'
seriamente 'seriously'
sinceramente 'sincerely'
sumariamente 'briefly'
a bem da verdade 'truth be told'
a bem dizer 'strictly speaking'
a bem verdade[PB] 'truth be told'
a propósito 'by the way'
a rigor 'strictly speaking, technically'
aos meus olhos 'as I see it'
aqui entre nós 'between us, off the record'
cá entre nós 'between us, off the record'
com franqueza 'frankly'
com (*todo*) *o devido respeito* 'with all due respect'
com toda a sinceridade 'in all sincerity'
com toda a franqueza 'with complete frankness'
em abono da verdade 'in the interest of truth, to be fair'
em bom português 'in plain Portuguese'
em linhas gerais 'generally speaking'
em português claro 'in plain Portuguese'
em poucas palavras 'in a few words, to put it briefly'
em resumo 'in summary'
em rigor 'strictly speaking, technically'
em sentido estrito 'strictly speaking, in a strict sense'
em sentido restrito 'in a narrow sense, in a limited sense'
em síntese 'in synthesis'
em suma 'in sum'
em termos gerais 'in general terms'
em termos simples 'in simple terms'
em toda a extensão da palavra 'in the full sense of the word'
entre aspas 'in quotation marks, so to speak'
falando francamente 'speaking frankly'
fora de brincadeira 'no kidding, all jokes aside'
francamente falando 'frankly speaking'
modéstia à parte 'modesty aside'
mutatis mutandis 'with the necessary changes made'
na aceção da palavra[PE] 'in the sense of the word'
na acepção da palavra[PB] 'in the sense of the word'
na boa[PB] 'serioulsy/honestly'
na extensão da palavra[PB] 'in the extent of the word'
na moral[PB] 'serioulsy/honestly'
no sentido estrito da palavra 'in the strict sense of the word'
no sentido estrito do termo 'in the strict sense of the word'
no sentido restrito da palavra 'in the restricted sense of the word'
no sentido restrito do termo 'in the restricted sense of the term'
numa palavra 'in a word, to sum it up'
para abreviar razões[EP] 'to cut a long story short, to make it brief'
para começo de conversa 'to start with, for starters'
para dizer a verdade 'to tell the truth'
para dizer depressa e bem[EP] 'to put it quickly and clearly'
para encurtar a conversa 'to cut the conversation short'
para encurtar a história 'to cut the story short'
para ser sincero[PB] 'to be honest, to be sincere'
para seu conhecimento 'for your information'
para seu governo 'for your own awareness, just so you know'
para todos os efeitos 'for all intents and purposes'
para falar depressa e bem[EP] 'to speak quickly and well'
pelo amor de Deus 'for the love of God'
pelo mal dos meus/nossos pecados 'for the sin of my/our misdeeds, unfortunately'
pelo que me consta 'as far as I know, to my knowledge'
por amor de Deus 'for God's sake'
por assim dizer 'so to speak, as it were'
por bem dizer[PB] 'broadly speaking'
por Deus 'for God'
por especial favor 'as a special favor'
por favor 'please'
por favor especial 'as a special favor'
por gentileza 'kindly please'
por obséquio 'would you be so kind as to, please'
por outras palavras 'in other words'
por tudo o que há de mais sagrado 'by all that is sacred'
salvo engano 'unless I'm mistaken'
salvo erro 'unless I'm mistaken'
salvo erro de memória 'unless my memory fails me'
salvo erro ou omissão 'unless otherwise stated'
salvo indicação contrária 'unless otherwise indicated'
salvo indicação em contrário 'unless otherwise indicated'
salvo melhor juízo 'tentatively, pending better judgment'
salvo melhor entendimento 'unless otherwise interpreted'
salvo melhor opinião 'unless there is a better opinion'
salvo seja 'lit.: save be'
se a memória não me falha 'if memory serves me right'
se faz favor 'if you please'
se fizer favor 'if you please'
se não me falha a memória 'if memory serves me right'
sem brincadeira[PB] 'no joke, seriously'
sem exagero 'no exaggeration, honestly'
strictu sensu 'strictly speaking'
tanto quanto me lembro 'as far as I remember'
tanto quanto me recordo 'as far as I recall'
verdade seja dita 'truth be told'

Será o ChatGPT um bom divulgador científico em cosmetologia? Um estudo linguístico sobre textos de divulgação científica

Is ChatGPT a good popular science disseminator in cosmetology? A linguistic study on popular science texts

Ana Filipa Pacheco[1], *Nuno Guimarães*[2], *Ana Torres*[3,4], *Purificação Silvano*[5], *Isabel Almeida*[3,4]

[1]Faculdade de Letras da Universidade do Porto, INESC TEC – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
[2]Faculdade de Ciências da Universidade do Porto, INESC TEC – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
[3]UCIBIO – Unidade de Biociências Moleculares Aplicadas, Departamento de Ciências do Medicamento, Faculdade de Farmácia da Universidade do Porto
[4]Laboratório Associado i4HB – Instituto para a Saúde e Bioeconomia, Faculdade de Farmácia da Universidade do Porto
[5]Faculdade de Letras da Universidade do Porto, Centro de Linguística da Universidade do Porto, INESC TEC – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência

Abstract

The science popularisation texts are fundamental for disseminating scientific knowledge in an accessible and understandable way to a non-specialised audience and have a different structure and characteristics from scientific articles (e.g., Garcés-Conejos & Sanchéz-Macarro, 1998; Zamboni, 1998). Studies on the linguistic properties of science popularisation texts in European Portuguese are not abundant, the exception being the *Promoção da Literacia Científica* project (Gonçalves & Jorge, 2018). On the other hand, within the realm of producing content, the large language models (LLMs), namely OpenAI's GPT models, have gained widespread public attention in a short period of time. Since they are recent, there is still very little assessment of the linguistic quality of the texts produced. Bearing these premises in mind, this study aims to assess the linguistic and textual quality of the responses generated by ChatGPT (GPT-3.5) in the field of cosmetology, concerning cosmetic products, ingredients, safety and efficacy and regulation categories, to identify patterns that allow an understanding of the differences and/or similarities between the content generated by LLM and that produced by human experts on the *Portal infoCosméticos*. For this, twenty questions previously answered and published on the portal were selected, and subsequently, four different prompts with different degrees of complexity were created, which resulted in eighty answers generated by ChatGPT. The answers were then analysed according to the parameters defined in an evaluation grid consisting of 11 questions. The analysis produced different types of results: overall, the answers written by the experts produced slightly better results than those from ChatGPT; in terms of interphrasal cohesion, it was found that the texts produced by the experts use a reduced number of connectors, in contrast to the recurrent use of connectors in the ChatGPT texts; there is the use of non-explained scientific jargon and a macrostructure with the absence of a conclusion paragraph in the texts published on the portal; the texts generated by ChatGPT have a high frequency of repetitions and/or tautologies.

Keywords: popular science texts, linguistic features, LLMs, ChatGPT.

Resumo

Os textos de divulgação científica são fundamentais para a disseminação do conhecimento científico de forma acessível e compreensível junto do público não especializado, apresentando estrutura e características diferentes das dos artigos científicos (e.g., Garcés-Conejos & Sanchéz-Macarro, 1998; Zamboni, 1998). Os estudos sobre as propriedades linguísticas do texto de divulgação científica em português europeu não abundam, sendo a exceção o projeto *Promoção da Literacia Científica* (Gonçalves & Jorge, 2018). Por outro lado, no âmbito da produção de conteúdo, os grandes modelos de linguagem (LLM), nomeadamente os modelos GPT da OpenAI, ganharam, em pouco tempo, atenção generalizada do público. Sendo recentes, a avaliação da qualidade linguística dos textos produzidos é ainda muito reduzida. Tendo estas premissas em consideração, o presente estudo tem como objetivo avaliar a qualidade linguística e textual das respostas geradas pelo ChatGPT (GPT-3.5) no domínio da cosmetologia, no que respeita às categorias de produtos cosméticos, ingredientes, segurança e eficácia e regulamentação, visando identificar padrões que permitam compreender as diferenças e/ou semelhanças entre o conteúdo gerado pelo LLM e aquele produzido por especialistas humanos no *Portal infoCosméticos*. Para isso, foram selecionadas vinte questões previamente respondidas e publicadas no portal e, posteriormente, criados quatro *prompts* distintos com diferentes graus de complexidade, que deram origem a oitenta respostas geradas pelo ChatGPT. As respostas foram, de seguida, analisadas, de acordo com os parâmetros definidos numa grelha de avaliação composta por 11 perguntas. A análise produziu resultados de diferentes tipos: em termos globais, as respostas escritas pelos especialistas produzem resultados ligeiramente superiores às do ChatGPT; quanto à coesão interfrásica, constatou-se que os textos produzidos por especialistas usam um número reduzido de conectores, contrastando com o uso recorrentemente de conectores nos textos do ChatGPT; verifica-se o uso de jargão científico não explicado e uma estrutura composicional global com ausência do parágrafo da conclusão, nos textos publicados no portal; os textos gerados pelo ChatGPT apresentam uma frequência elevada de repetições e/ou tautologias.

Palavras-chave: textos de divulgação científica, características linguísticas, grandes modelos de linguagem, ChatGPT.

1. Introdução

Numa era em que os avanços científicos moldam grandemente a nossa sociedade, a capacidade de comunicar eficazmente conhecimentos complexos a públicos não especializados tornou-se essencial. A divulgação científica estabelece a ponte entre o discurso especializado e a forma como o público compreende e se relaciona com o conhecimento, transformando conceitos técnicos em conteúdo acessível.

Com o advento do acesso rápido e fácil ao conteúdo *online*, e à semelhança do que acontece noutras áreas do conhecimento, observa-se um crescimento significativo na procura de informação relacionada com a área da cosmetologia, por parte dos consumidores (cf. Ferreira et al., 2024). No entanto, esta busca por conhecimento nem sempre é acompanhada por uma avaliação crítica e cuidada das fontes consultadas. Este comportamento pode levar os consumidores a tomarem decisões menos informadas na escolha de produtos cosméticos e, consequentemente, a aumentarem o risco tanto de seleção como de utilização inadequadas de produtos cosméticos para as suas necessidades individuais, potenciando não só efeitos adversos e reações indesejadas, mas também a frustração face a resultados que não correspondem às expectativas.

Este cenário potencial sublinha a importância de uma comunicação clara e rigorosa na divulgação de informações sobre cosmética que garanta que o público não especializado tenha acesso a conteúdos fiáveis e cientificamente fundamentados.

Por outro lado, na corrida à automatização da comunicação, ao acesso à informação e à produção de conteúdo, os grandes modelos de linguagem (*large language models* – LLMs), nomeadamente os modelos GPT da OpenAI, entre os quais o ChatGPT (GPT-3.5) (Brown et al., 2020), ganharam, em pouco tempo, atenção

generalizada do público. Estes modelos, cujo conteúdo de saída se assemelha, aparentemente, aos resultados produzidos por humanos, vêm sendo progressivamente utilizados como um recurso de procura e produção de conteúdo, em vários domínios (e.g., Ray, 2023), designadamente na área da divulgação científica. Não obstante, ao facilitarem o acesso rápido a informações sobre, concretamente, cosmética – o objeto de estudo deste trabalho –, estas ferramentas introduzem novos desafios, como a verificação da precisão científica ou a potencial simplificação excessiva de conceitos complexos.

Sendo recentes, a avaliação da qualidade linguística e textual dos textos produzidos por LLMs é ainda pouco explorada. Contudo, tal análise pode determinar, por exemplo, se os textos gerados cumprem o seu objetivo comunicativo, na mesma medida em que os textos redigidos por humanos. Os dados fornecidos podem, assim, possibilitar contributos relevantes, tanto para a avaliação da adequação a diferentes contextos de comunicação dos textos gerados como para a identificação de padrões de erro, com consequente contributo para a compreensão de potenciais desafios linguísticos apresentados por linguagem gerada por Inteligência Artificial.

Tendo estas premissas em consideração, o presente trabalho pretende investigar o desempenho do ChatGPT (GPT-3.5) quanto à produção de conteúdo de divulgação científica, na área da cosmetologia, de modo a estabelecer padrões que possam identificar as diferenças e/ou semelhanças entre a produção de conteúdo sobre cosmética por humanos especialistas num portal de divulgação científica – o *Portal infoCosméticos* (P*i*C) – e por um LLM, no que diz respeito à qualidade linguística e textual.

Considerando este enquadramento, estabelecem-se as seguintes perguntas de investigação:

I. Quais são as principais diferenças e semelhanças linguísticas e textuais entre os textos gerados pelo ChatGPT e os textos produzidos por humanos especialistas no P*i*C?

II. Que fragilidades linguísticas e textuais se observam na produção textual do ChatGPT, quando comparada com a produção humana, no contexto da divulgação científica em cosmetologia?

Para responder a estas questões, o presente estudo procede à análise comparativa de um conjunto de textos de divulgação científica sobre cosmética, redigidos por especialistas no P*i*C e por um modelo de linguagem (ChatGPT, versão 3.5), com base numa grelha construída a partir de bibliografia selecionada.

2. Algumas considerações linguísticas sobre os textos de divulgação científica

Os textos de divulgação científica são fundamentais para a disseminação do conhecimento científico de forma acessível e compreensível junto do público em geral, não especializado, apresentando estrutura e características diferentes das dos artigos científicos (e.g., Garcés-Conejos & Sánchez-Macarro, 1998; Zamboni, 1998). Vários estudos destacam as propriedades linguísticas da comunicação científica dirigida ao público em geral que garantem a eficácia e a acessibilidade dos textos de divulgação científica.

Myers (1991), por exemplo, observa que a coesão lexical é marcada de forma mais explícita em textos de divulgação científica do que em artigos científicos pela necessidade de tornar mais claras as ligações lexicais a um público não especialista. Já Aubakir (2017) destaca a importância do uso de linguagem não especializada e da ausência de jargão científico, enfatizando o princípio da acessibilidade e da visibilidade da informação. Para o autor, a inclusão de exemplos desempenha um papel igualmente fundamental na compreensão e retenção dos conteúdos apresentados.

Por outro lado, o estudo de Molek-Kozakowska (2016) sobre os títulos da revista *New Scientist* revela tendências linguísticas específicas, tais como a redução do número de artigos definidos, bem como uma menor incidência de verbos, que, quando presentes, surgem maioritariamente no presente do indicativo. Formas verbais no passado são escassas, e a conjunção *and* é frequentemente utilizada para ligar ideias. Os advérbios tendem a ser supérfluos, surgindo apenas quando desempenham funções intensificadoras. Outros traços distintivos incluem o uso do pronome *your*, a presença dos verbos modais *may* e *could*, associados à modalidade

epistémica, e o uso do numeral ordinal *first*. Num estudo posterior, Molek-Kozakowska (2017) identifica determinados grupos de características essenciais no jornalismo de divulgação científica, como o uso de adjetivos na forma superlativa e comparativa e de adjetivos intensificadores para expressar a inovação, importância e destaque das descobertas e avanços científicos. A autora sublinha ainda a necessidade de adotar estilos comunicativos menos exigentes e mais envolventes, uma vez que o excesso de jargão e a complexidade excessiva dos textos podem afastar os leitores (Molek-Kozakowska, 2017).

Embora haja alguns estudos sobre as propriedades linguísticas do texto de divulgação científica, como aqueles mencionados acima, em português europeu os estudos não abundam, sendo a exceção o projeto *Promoção da Literacia Científica* (Gonçalves & Jorge, 2018). Nesse contexto, Filipe (2018) analisa algumas marcas do texto de divulgação científica e constata, por exemplo, uma maior predominância de conectores que atestam relações de adição/conjunção, seguidas das relações de temporalidade, contraste e finalidade/intencionalidade. Quanto aos tempos verbais, o autor destaca, igualmente, o uso do presente do indicativo seguido do pretérito perfeito e do futuro, sendo o pretérito imperfeito o que menos vezes ocorre.

Não é objetivo do presente estudo analisar exaustivamente alguns dos aspetos identificados por Molek-Kozakowska (2016, 2017) e por Filipe (2018). No entanto, considerou-se relevante partilhá-los, uma vez que contribuem para a contextualização dos textos de divulgação científica e para uma melhor compreensão das marcas linguísticas que os caracterizam.

Para além das propriedades linguísticas já mencionadas, a organização global do texto também constitui um fator essencial para a eficácia comunicativa da informação veiculada. Uma das dimensões avaliadas nas respostas que constituem o corpus deste trabalho prende-se com a estrutura composicional global, tendo sido verificada a existência (ou não) de uma organização tripartida (introdução, desenvolvimento e conclusão) do conteúdo temático. Esta divisão, embora não formalizada teoricamente por Bronckart (1997), aproxima-se da sua noção de plano geral do texto, que diz respeito à organização global do conteúdo temático e é passível de ser apreendida pelo leitor durante o processo de leitura. Por sua vez, Adam (2005) propõe que os textos se organizem em unidades hierárquicas complexas, que formam estruturas globais relativamente autónomas. Assim, a presença de uma estrutura tripartida pode ser interpretada como um tipo de configuração funcional que decorre da organização dessas unidades superiores e contribui para a progressão temática e coerência global. Em contextos de divulgação científica, esta estrutura revela-se particularmente relevante, na medida em que favorece a orientação do leitor e a construção progressiva do significado.

Contudo, Gonçalves & Magalhães (2019) destacam a heterogeneidade inerente à divulgação científica, apontando para a diversidade e natureza dos formatos textuais, que não se reduzem a somente um género textual estabilizado, mas a um conjunto diversificado de práticas comunicativas.

Apesar de se registarem algumas investigações desenvolvidas neste domínio, estudos em português europeu ainda são limitados, reforçando-se, assim, a necessidade de mais contribuições para a análise das características linguísticas e estruturais dos textos de divulgação científica, contributo que este trabalho pretende dar.

3. Algumas considerações sobre o desempenho dos LLMs na geração de textos de divulgação científica

Estudos sobre o desempenho dos LLMs em diferentes contextos revelam um conjunto de potencialidades, mas também desafios que podem influenciar a geração de textos em domínios como os da divulgação científica. Em Onder et al. (2024), por exemplo, é avaliada a fiabilidade e a legibilidade das respostas do ChatGPT-4 sobre hipotiroidismo na gravidez, e os resultados mostram fiabilidade de moderada a boa e qualidade alta, adequada ao público em geral. No entanto, a legibilidade tende a ser reduzida, sugerindo que os textos gerados exigem um nível universitário para serem compreendidos, o que pode dificultar a sua acessibilidade no domínio da divulgação científica.

Zhang et al. (2025) analisam a capacidade dos LLMs para gerar conteúdo (perguntas, concretamente) a partir de um dado contexto, comparando-o com conteúdo formulado por humanos. Os resultados sugerem que os LLMs tendem a produzir perguntas com um nível de cobertura equilibrado, mas com preferência por questões que exigem respostas descritivas e extensas. Esta característica pode indicar detalhe e extensão nos textos gerados pelos modelos.

Zhou et al. (2024) alertam para uma tendência preocupante: à medida que os modelos aumentam de escala, a sua fiabilidade tende a diminuir. Os erros persistentes, mesmo em tarefas simples, e as respostas incorretas a perguntas complexas constituem desafios que devem ser considerados na utilização de LLMs na procura de informação do domínio da divulgação científica. O estudo salienta, ainda, que a supervisão humana é muitas vezes ineficaz na correção destes erros, uma vez que os utilizadores tendem a confiar demasiado nos modelos e podem não reconhecer os resultados incorretos.

A análise de Moraes et al. (2024) reforça outro aspeto crítico relacionado com a forma como modelos como o ChatGPT lidam com ambiguidades linguísticas em português do Brasil, que vão desde as ambiguidades semânticas e sintáticas às lexicais. Os resultados indicam que, apesar dos avanços, os modelos ainda apresentam limitações na interpretação de textos com ambiguidades, o que pode comprometer a clareza e a precisão das informações geradas.

Embora os estudos pareçam mostrar que os LLMs apresentam potencialidades a considerar para a geração de textos com grau de especificidade mais elevado, como os textos de divulgação científica, o seu desempenho é influenciado por limitações que podem estar relacionadas, por exemplo, com a legibilidade, a fiabilidade das informações geradas e a interpretação de ambiguidades.

4. O estudo

O presente estudo tem como objetivo avaliar a qualidade linguística e textual das respostas geradas pelo ChatGPT (GPT-3.5) no domínio da cosmetologia, no que respeita às categorias de produtos cosméticos, ingredientes, segurança e eficácia e regulamentação. A análise visa identificar padrões que permitam compreender as diferenças e/ou semelhanças entre o conteúdo gerado pelo LLM e aquele produzido por especialistas humanos num portal de divulgação científica – o *Portal infoCosméticos* (P*i*C).

4.1. Contextualização do estudo

O P*i*C é um projeto que assenta numa prática pedagógica de aprendizagem colaborativa na criação de conhecimento para divulgação científica no âmbito da promoção e educação para a saúde e envolve três unidades orgânicas da Universidade do Porto: a Faculdade de Farmácia, a Faculdade de Letras e a Faculdade de Belas Artes (Almeida & Silvano, 2021). O portal facilita, igualmente, o acesso a informação validada cientificamente, educando e informando os consumidores e profissionais na área da cosmética através de linguagem acessível e de infografias para facilitar a compreensão.

O processo metodológico para a criação de conteúdo de divulgação científica para o P*i*C inicia-se com a identificação de uma questão relevante na área da cosmética, sendo a redação da resposta atribuída a um estudante do Mestrado Integrado em Ciências Farmacêuticas da Faculdade de Farmácia da Universidade do Porto. O colaborador deve seguir um guião para a elaboração das respostas, que são acompanhadas de infografias, desenhadas por estudantes da Faculdade de Belas Artes. Os textos e as infografias passam, então, por uma edição linguística antes da sua submissão para validação científica. Após esta etapa, procede-se à reformulação e validação da versão final da resposta por um revisor, seguida de uma revisão linguística final. Posteriormente, o conteúdo é harmonizado pela coordenadora do P*i*C, sendo finalmente publicado, num período que pode variar entre cinco e oito meses.

Dada a crescente utilização de LLMs na geração automática de conteúdo, este estudo propõe-se a comparar a qualidade linguística e textual das respostas produzidas pelos colaboradores do P*i*C com as respostas geradas

pelo ChatGPT, um dos modelos mais avançados de processamento de linguagem natural. Para esse efeito, recorreu-se ao ChatGPT-3.5, uma versão otimizada do modelo GPT-3, lançada pela OpenAI em março de 2023 (Brown et al., 2020). Este modelo introduz melhorias na geração de texto, nomeadamente ao nível da coerência e da compreensão contextual, tendo sido desenhado para produzir respostas semelhantes às de um ser humano com base no *input* recebido (Ray, 2023; Ye et al., 2023).

4.2. **Processo metodológico**

Numa primeira fase, procedeu-se ao desenho de uma grelha (cf. Anexo 1) de análise das propriedades linguísticas e textuais dos textos de divulgação científica, com base em estudos como Myers (1991), Bronckart (1997), Gárces-Conejos & Sánchez-Macarro (1998), Aubakir (2017) e Gonçalves & Jorge (2018). Esta grelha, composta por 11 perguntas, permitiu examinar os seguintes elementos: eficiência comunicativa, estrutura composicional global, clareza, coesão, coerência, (presença/ausência de) jargão científico, inclusão de exemplos, erros de natureza morfossintática, lexical, semântica e/ou ortográfica e repetições/tautologias.

A seleção das perguntas e respostas utilizadas no estudo foi realizada pela equipa das Ciências Farmacêuticas, tendo sido escolhidas vinte questões previamente respondidas e publicadas no P*i*C pelos colaboradores. As perguntas pertencem a quatro categorias: produtos cosméticos, ingredientes, segurança e eficácia e regulamentação.

Posteriormente, para cada uma das perguntas, foram criados, por um membro da equipa de Ciência de Computadores, quatro *prompts* distintos (cf. Tabela 1) com diferentes graus de complexidade que permitiram a geração das respostas por parte do ChatGPT e garantiram uma avaliação abrangente do desempenho do modelo: (i) o *prompt* 1 – *zero-shot*: pergunta ou instrução sem exemplos prévios; simula o comportamento do utilizador do LLM; (ii) o *prompt* 2 – *expert 1*: é feita uma contextualização sobre a tarefa a realizar de forma a orientar o *output* do modelo; (iii) o *prompt* 3 – *expert 2*: refinamento à contextualização usada para o *prompt* 2, providenciando-se mais contexto; e (iv) o *prompt* 4 – *few-shot*: pergunta acompanhada de exemplo de resposta desejada de forma a aumentar a precisão e adequação da resposta ao contexto (a resposta selecionada foi retirada do P*i*C e corresponde a uma pergunta não usada na amostra).

O processo de avaliação das respostas foi conduzido por uma investigadora, com formação em Linguística e conhecimento não especializado, mas informado, em cosmética e cosmetologia, o que permitiu a aplicação crítica e metodológica da grelha de critérios linguísticos.

A análise foi realizada manualmente, com recurso ao visualizador de documentos do sistema operativo macOS – o Preview –, e baseou-se na leitura e análise de cada resposta, seguida do preenchimento da grelha linguística previamente definida. Foram tomadas notas durante o processo de análise, com registo imediato das pontuações atribuídas a cada critério.

A recolha e análise dos conectores presentes nas respostas foram realizadas pela mesma investigadora, tendo sido posteriormente revistas por dois especialistas na área da Linguística, que orientaram e validaram todo o processo. Esta verificação garantiu a consistência da identificação e categorização dos conectores utilizados em cada um dos dois grupos de respostas.

Dado que a análise foi realizada por um único elemento, não foi possível calcular métricas de acordo interanotador. Reconhece-se esta limitação, comum em estudos exploratórios de pequena escala, e sugere-se a replicação do presente trabalho com múltiplos avaliadores, de modo a reforçar a robustez metodológica e a fiabilidade dos resultados.

Tabela 1

Prompts *Usados no Processo de Geração de Respostas pelo ChatGPT*

Tipo de *prompt*	Princípio	Exemplo
***zero-shot* (*prompt* 1)**	Pergunta direta no ChatGPT – simula o comportamento do utilizador	É mais eficaz um protetor solar com proteção UVA ou UVB?
***expert 1* (*prompt* 2)**	Contextualização sobre a tarefa a realizar	És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Questão: É mais eficaz um protetor solar com proteção UVA ou UVB? A resposta deverá ter entre 250 e 300 palavras.
***expert 2* (*prompt* 3)**	Contextualização sobre a tarefa a realizar + refinamento	Estás a produzir um texto de divulgação científica na área da cosmética. Tens de explicar, de uma forma simples, mas rigorosa, conceitos aos consumidores dos produtos dessa área e aos profissionais em cosmética. Questão: É mais eficaz um protetor solar com proteção UVA ou UVB? A resposta deverá ter entre 250 e 300 palavras.
***few-shot* (*prompt* 4)**	Contextualização + disponibilização de exemplo do tipo de resposta esperada	És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Aqui estão uma pergunta e resposta que servem como exemplo do que é pretendido. Pergunta: O que são champôs sem sulfatos? Resposta: [resposta publicada no *Portal infoCosméticos*: https://portalinfocosmeticos.pt/ingredientes/o-que-sao-champos-sem-sulfatos/] O aparecimento no mercado dos produtos capilares de champôs rotulados com menção 'champô sem sulfatos' surgiu na sequência de notícias que colocam em causa a segurança destes ingredientes. Os champôs contêm tensioativos que atuam como agentes lavantes por apresentarem caraterísticas físico-químicas que lhes conferem capacidade para envolver a gordura e sujidade de forma a serem arrastadas com a água. Existem vários grupos destes ingredientes, sendo que os mais representativos pertencem à classe dos tensioativos aniónicos, entre os quais se incluem os sulfatos como o sulfato de laurilo e sódio (Sodium Lauryl Sulfate) ou sulfato de laurilo e amónio (Ammonium Lauryl Sulfate) e os seus éteres derivados como o lauril éter sulfato de sódio (Sodium Laureth Sulfate) ou lauril éter sulfato de amónia (Ammonium Laureth Sulfate). Por norma, os tensioativos aniónicos são dotados de elevado poder detergente e bom poder espumante, podendo, contudo, remover lípidos essenciais à qualidade da haste capilar. Assim, um champô com certos tipos de sulfatos, poderá deixar o cabelo fragilizado, opaco e emaranhado. Dado o baixo tempo de contacto com estes ingredientes sujeitos a posterior enxaguamento, o couro

	cabeludo deverá ficar limpo, mas a sua qualidade não deverá ser afetada. A partir do ano 2000, promoveu-se o desenvolvimento de formulações com menor teor em sulfatos, compensados ou substituídos por agentes tensioativos menos espumantes, menos desengordurantes e, por conseguinte, mais suaves. Assim sendo, formulações sem sulfatos não contêm os ingredientes acima referidos, apesar de estes terem já sido estabelecidos como seguros nas condições de utilização previstas. Os sulfonatos, sarcosinatos e sulfosucinatos são tensioativos aniónicos alternativos com razoável poder espumante e maior suavidade. Pergunta: É mais eficaz um protetor solar com proteção UVA ou UVB?

Os *prompts* apresentados na Tabela 1 deram, portanto, origem a oitenta respostas geradas pelo ChatGPT e recolhidas em janeiro de 2024, as quais, juntamente com as vinte respostas às vinte perguntas selecionadas a partir do P*i*C, constituem o *corpus* deste trabalho. Este *corpus* totaliza 25398 palavras: 6449, referentes ao *corpus* constituído pelos textos de divulgação científica do P*i*C, e 18949, correspondentes aos textos gerados pelo ChatGPT.

Por último, e após a geração das respostas pelo ChatGPT, procedeu-se à análise com base nos resultados das avaliações linguísticas e textuais e da qualidade científica[1], conduzidas a partir das grelhas desenvolvidas, das respostas do P*i*C e do ChatGPT.

4.3. **Resultados**

A análise produziu resultados de diferentes tipos. Em primeiro lugar, permitiu detetar diferenças linguísticas e textuais entre os textos produzidos por especialistas na área da cosmetologia e aqueles gerados pelo LLM. Essas diferenças revelaram-se importantes para a reflexão sobre as características dos textos de divulgação científica, escritos em português europeu, contribuindo para a sua caracterização.

Em termos globais (cf. Tabela 2), os resultados revelam que as respostas escritas pelos especialistas produzem resultados ligeiramente melhores do que as do ChatGPT. Os textos do P*i*C obtiveram uma média global de 13,6/20 pontos, enquanto aqueles gerados pelo ChatGPT apresentaram uma média ligeiramente inferior de 13,43/20 pontos. Ao nível da categoria de ingredientes, o ChatGPT apresenta valores acima daqueles obtidos pelo P*i*C (14 vs. 13,2 pontos, respetivamente) e inferiores em relação à categoria de segurança (12,15 vs. 14 pontos, respetivamente).

Ao analisar-se os dados do ponto de vista dos diferentes *prompts* utilizados para gerar respostas do ChatGPT, os resultados demonstram uma variação considerável no desempenho. O *prompt* 3, criado tendo em consideração o refinamento da contextualização sobre a tarefa a realizar, gerou as respostas com a melhor pontuação média (15,15/20 pontos), enquanto o *prompt* 1, que simula o comportamento do utilizador do LLM, apresentou a menor (11/20 pontos), o que sugere que a formulação do pedido influencia significativamente a qualidade da resposta obtida, tal como Brown et al. (2020) referem. Na categoria de segurança, a discrepância é ainda mais evidente, ficando clara a diferença de resultados entre a análise das respostas do P*i*C (14/20 pontos) e aquelas geradas pelo ChatGPT (9,8/20 pontos).

Quanto à variabilidade dos resultados (desvio padrão, σ), o P*i*C apresenta menor dispersão (σ global = 1,9) em comparação com o ChatGPT (σ global = 2,5), indicando uma maior consistência na produção dos textos de divulgação científica. Nas respostas do ChatGPT, a variabilidade é particularmente elevada nos resultados

[1] Neste artigo, só são apresentados os resultados da avaliação linguística e textual. A análise da qualidade científica das respostas é apresentada em Torres et al. (2024).

globais do *prompt* 1 (σ = 2,79) e, especificamente, na categoria de produtos (σ = 4,34), sugerindo que a qualidade das respostas pode oscilar dependendo da formulação da pergunta e do tópico solicitado. Por outro lado, o *prompt* 4 apresentou o menor desvio padrão (σ = 1,09), o que pode indicar uma maior uniformidade na geração das respostas com este formato de *prompt*, para o qual é fornecido um exemplo de resposta desejada.

A Tabela 2 apresenta os dados estatísticos detalhados das respostas do P*i*C e do ChatGPT, por *prompt* e por resultado global. De forma a facilitar a leitura comparativa, os dados da Tabela 2 estão representados graficamente na Figura 1.

Tabela 2

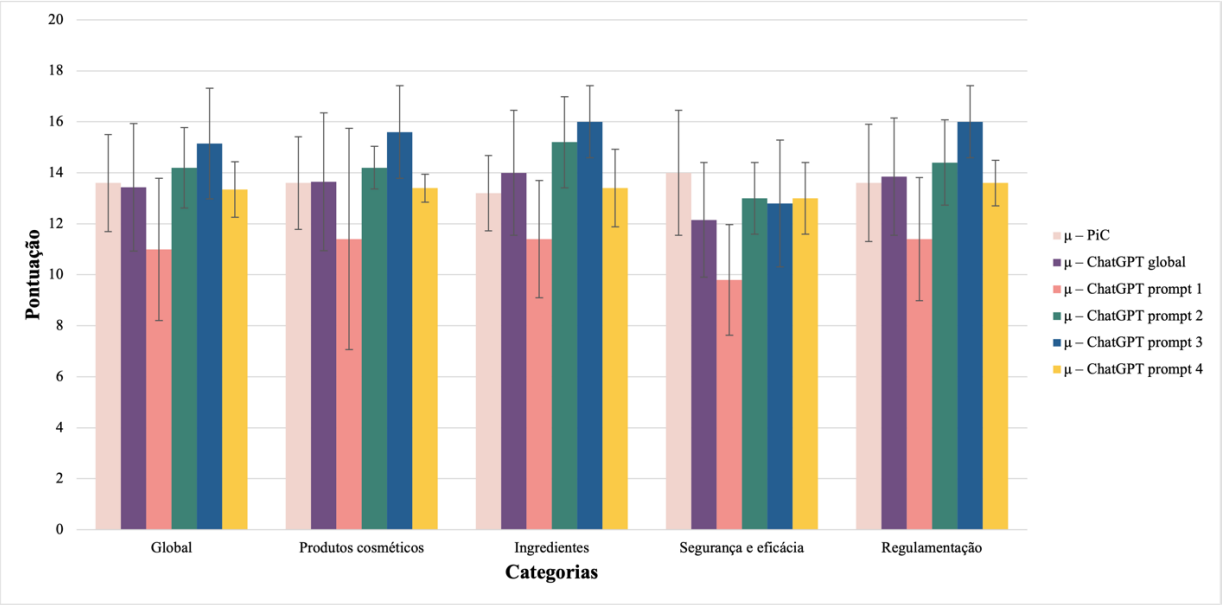
Dados Estatísticos Referentes à Análise Linguística das Respostas do PiC e do ChatGPT

	Global	**Produtos cosméticos**	**Ingredientes**	**Segurança e eficácia**	**Regulamentação**
μ – P*i*C	13,6	13,6	13,2	14	13,6
μ – ChatGPT global	13,43	13,65	14	12,15	13,85
μ – ChatGPT *prompt* 1	11	11,4	11,4	9,8	11,4
μ – ChatGPT *prompt* 2	14,2	14,2	15,2	13	14,4
μ – ChatGPT *prompt* 3	15,15	15,6	16	12,8	16
μ – ChatGPT *prompt* 4	13,35	13,4	13,4	13	13,6
σ – P*i*C	1,9	1,82	1,48	2,45	2,3
σ – ChatGPT global	2,5	2,7	2,45	2,25	2,3
σ – ChatGPT *prompt* 1	2,79	4,34	2,3	2,17	2,41
σ – ChatGPT *prompt* 2	1,58	0,84	1,79	1,41	1,67
σ – ChatGPT *prompt* 3	2,17	1,82	1,41	2,49	1,41
σ – ChatGPT *prompt* 4	1,09	0,55	1,52	1,41	0,89

Os resultados parecem demonstrar que, embora o ChatGPT possa produzir respostas cujos resultados da análise aparentam ser comparáveis ou até superiores aos produzidos pelo P*i*C em algumas categorias, o seu desempenho varia consideravelmente de acordo com o tipo de *prompt* utilizado e a categoria em análise.

Figura 1

Representação Gráfica dos Dados Estatísticos



Paralelamente, foi possível identificar alguns problemas linguísticos e composicionais nos textos produzidos pelos especialistas humanos e pelo LLM, cuja análise é apresentada de seguida. Para este estudo, centrar-nos-emos nas questões em que foram observados mais problemas: a estrutura composicional global, a coesão interfrásica, as repetições e/ou tautologias e o jargão científico.

No que diz respeito à **estrutura composicional global dos textos**, a análise procurou caracterizar a estrutura dos textos produzidos pelos especialistas humanos e pelo LLM, verificando a presença, ou não, de uma estrutura tripartida (cf. Tabela 3, na qual as três partes da estrutura se encontram marcadas a negrito verde [introdução], rosa [desenvolvimento] e laranja [conclusão]). Nos textos do P*i*C foi possível constatar a ausência frequente do parágrafo de conclusão (cf. Tabela 4), o que pode comprometer a clareza e a estrutura discursiva, dificultando a perceção global da mensagem por parte do leitor não especializado. A solução para este problema é, no entanto, relativamente simples e consiste na inclusão sistemática de um parágrafo final que sintetize os pontos essenciais abordados e forneça um encerramento adequado ao texto. Esta recomendação foi já comunicada à equipa responsável pela produção de conteúdos do portal, garantindo que futuros textos integrem esse elemento fundamental para a coesão e eficiência comunicativa dos textos.

Tabela 3

Estrutura Composicional Global das Respostas do PiC e do ChatGPT

Pergunta: É mais eficaz um protetor solar com proteção UVA ou UVB?	***Prompt* 1 – *zero-shot*: É mais eficaz um protetor solar com proteção UVA ou UVB?**
PiC	**ChatGPT**
A pele está permanentemente exposta à radiação solar, que inclui a radiação ultravioleta (UV). A radiação UV subdivide-se em radiação UVA, que corresponde a cerca de 90-95% da radiação UV que atinge a superfície terrestre, a radiação UVB, que constitui aproximadamente 5-10%, e a radiação UVC. Destas, as duas primeiras requerem especial atenção, uma vez que a radiação UVC não atravessa a camada de ozono e não atinge a superfície da Terra. **[introdução]** A radiação UVA, com comprimentos de onda compreendidos entre os 320-400 nanómetros (nm), tem a capacidade de atingir não só a epiderme (camada mais superficial da pele), mas também a derme (camada mais profunda da pele), sendo também responsável por promover o bronzeamento direto. Este tipo de radiação provoca stress oxidativo, contribuindo, assim, para um envelhecimento prematuro da pele com consequente formação de rugas e flacidez cutânea. É importante salientar que a radiação UVA apresenta outros efeitos nocivos, como o aumento do risco de cancro da pele e indução de reações de fotossensibilidade. **[desenvolvimento]** A radiação UVB, por sua vez, caracteriza-se por comprimentos de onda entre os 290-320 nm e atinge apenas a epiderme. Esta radiação está associada a queimaduras solares e ao risco de desenvolvimento do cancro de pele, como, por exemplo, o melanoma. Uma das respostas do nosso organismo à exposição de radiação UVB é o bronzeamento devido à promoção da síntese de mais melanina, a qual confere alguma proteção a este tipo de radiação. **[desenvolvimento]** Tendo em consideração os efeitos nocivos de ambos os tipos de radiação para o nosso organismo, na hora de escolher um protetor solar, é importante optar por um que ofereça proteção simultânea contra as radiações UVA e UVB. Desta forma, o protetor, ao ser aplicado em períodos regulares, evita as queimaduras solares, previne o envelhecimento precoce da pele e protege de outros efeitos prejudiciais. **[conclusão]**	Ambos os tipos de proteção são importantes para garantir uma proteção eficaz contra os danos causados pelo sol. Os raios UVA são responsáveis pelo envelhecimento da pele e podem penetrar nas camadas mais profundas, causando danos a longo prazo, como rugas, manchas e perda de elasticidade. Eles também podem contribuir para o desenvolvimento de câncer[2] de pele. **[introdução]** Os raios UVB são responsáveis pelas queimaduras solares e são mais intensos durante o verão e em altitudes mais elevadas. Eles também são um fator importante no desenvolvimento de câncer de pele. **[desenvolvimento]** Portanto, é importante escolher um protetor solar que ofereça proteção contra ambos os tipos de raios. Procure por produtos que tenham um amplo espectro de proteção, indicado na embalagem como "FPS" (fator de proteção solar) seguido de um número e "UVA" dentro de um círculo. Isso indica que o produto oferece proteção contra os raios UVA e UVB. **[conclusão]**

Quanto ao ChatGPT, a presença da estrutura tripartida foi consistente na maior parte dos textos gerados através dos *prompts* 2, 3 e 4. Os problemas mais relevantes registaram-se face a respostas produzidas com recurso ao *prompt* que simula o comportamento do utilizador – o *prompt* 1. Nestes casos, foi recorrente a ausência total de uma estrutura com introdução, desenvolvimento e conclusão, como é exemplo o texto gerado

[2] Um problema recorrente observado nos textos produzidos pelo LLM é o uso do Português do Brasil, sobretudo ao nível lexical, como evidencia este exemplo.

como resposta à pergunta "*Os cosméticos para bebé têm "perfume"?*" (cf. Tabela 4). Para o mesmo *prompt*, ocorrências de uma estrutura incompleta foram igualmente frequentes, o que evidencia uma aparente fragilidade do modelo na geração de conteúdo especializado perante instruções menos detalhadas e/ou contextualizadas. Este resultado sugere que a qualidade estrutural dos textos gerados pelo ChatGPT depende significativamente da formulação do *prompt*, sendo que instruções vagas ou que simulem um utilizador não especializado tendem a resultar em respostas desorganizadas e menos completas, com baixo valor qualitativo.

Tabela 4

Estrutura Composicional Global das Respostas do PiC e do ChatGPT – Problemas

Pergunta: Os cosméticos para bebé têm "perfume"?	***Prompt* 1 – *zero-shot*: Os cosméticos para bebé têm "perfume"?**
P*i*C	**ChatGPT**
Os cosméticos para bebé são produtos destinados a bebés e crianças com menos de três anos de idade e são especialmente formulados para serem suaves e não irritantes, pelo que os ingredientes utilizados são criteriosamente selecionados. Fragrâncias, designadas frequentemente por "perfume", são combinações complexas de substâncias naturais e/ou artificiais que são utilizadas numa grande variedade de produtos cosméticos para transmitir um odor agradável e/ou mascarar o odor de outros ingredientes, melhorando a experiência de utilização do produto. Para efeitos de rotulagem dos produtos cosméticos estas substâncias são identificadas na lista de ingredientes pelos termos genéricos "perfume", "parfum" ou "aroma", mantendo-se a confidencialidade dos compostos odoríferos utilizados, com a exceção das substâncias enumeradas no Anexo III do Regulamento europeu aplicável aos produtos cosméticos. Neste anexo, estão identificadas 26 fragrâncias que devem ser obrigatoriamente mencionadas na lista de ingredientes quando a sua concentração é superior ao valor limite legislado (0,001 % nos produtos não enxaguados e 0,01 % nos produtos enxaguados) devido ao seu potencial alergénico, mesmo que sejam introduzidas no produto cosmético como parte de um ingrediente complexo, sendo identificadas de acordo com a Nomenclatura Internacional dos Ingredientes Cosméticos (INCI). A pele das crianças, apesar de ser semelhante à dos adultos, em termos estruturais, é potencialmente mais sensível. Desta forma, tal como os produtos destinados a adultos, os cosméticos destinados a bebés podem conter perfume. No entanto, é utilizada menor quantidade de perfume e/ou é evitada a incorporação de fragrâncias com potencial sensibilizante. A utilização de fragrâncias em produtos cosméticos é sujeita a restrições de forma a garantir a proteção da saúde e a segurança do consumidor. Contudo, em caso de preocupação com o potencial sensibilizante das fragrâncias, o consumidor pode optar por produtos	Sim, muitos cosméticos para bebé têm perfume. No entanto, é importante escolher produtos com fragrâncias suaves e hipoalergénicas, especialmente formuladas para a pele sensível dos bebés. É recomendado evitar produtos com fragrâncias fortes, uma vez que podem irritar a pele delicada do bebé. **[não estruturado em introdução, desenvolvimento e conclusão]**

cosméticos sem perfume. Para os identificar é necessário verificar a lista de ingredientes e garantir que não são referidos os termos "perfume" ou "aroma", nem são referidas as substâncias odoríferas que se encontram referidas no Anexo III da Legislação Europeia tais como, o "eugenol", "limonene", "benzyl alcohol", "benzyl benzoate", "linalool", "citronellol", "benzyl salicylate" entre outros. **[falta conclusão]**	

Como mencionado anteriormente, um outro problema observado durante a análise realizada diz respeito à **coesão interfrásica**. A este nível, constatámos que os textos produzidos por especialistas usam um número muito reduzido de conectores, o que pode dificultar o estabelecimento, pelo público, de relações entre as ideias apresentadas e comprometer a clareza e a eficiência da mensagem. Neste trabalho, seguimos a proposta de Lopes & Carapinha (2013) para os conectores "de natureza adverbial, preposicional ou até verbal que exercem funções de articulação textual, sem dar origem a produtos frásicos" (Lopes & Carapinha, 2013, p. 89). Adotamos ainda na nossa análise a proposta das autoras quanto à categorização destes conectores sistematizada na Tabela 5.

Tabela 5

Categorização dos Conectores (Lopes & Carapinha, 2013)

Categorias de conectores		**Valor**	**Exemplos do *corpus***
Reformuladores	de paráfrase	Clarificação da informação	*ou seja*
	de retificação	Correção da informação	-----
	de síntese	Síntese da informação	*em resumo, em suma, em conclusão*
Especificativos		Exemplificação ou particularização da informação	*por exemplo, como, tais como, assim como*
Reforçativos		Reforço da informação com valor aditivo	*além disso, de facto*
Estruturadores		Ordenação da informação	*por outro lado, em seguida, em primeiro lugar, por fim*
Consequenciais		Expressão de consequência ou resultado	*assim, desta forma, consequentemente, portanto, assim sendo, como tal, por isso, como resultado*
Contrastivos		Expressão de contraste	*no entanto, contudo, por sua vez, porém*

Os resultados da análise da frequência de conectores, apresentados na Tabela 6, revelam diferenças significativas tanto na distribuição global quanto na densidade de uso dentro de cada conjunto. No total, 19,71% de conectores ocorre nos textos de divulgação científica do P*i*C, enquanto 80,29% ocorre nos textos gerados pelo ChatGPT. Estes resultados mostram claramente que o ChatGPT inclui uma elevada frequência de conetores em comparação com o P*i*C.

No entanto, considerando-se a frequência relativa dentro de cada conjunto, observa-se que 1,46% das palavras dos textos do P*i*C correspondem a conectores, enquanto nos textos do ChatGPT, essa percentagem sobe para 2,02%. Assim, além de gerar mais conectores em termos absolutos, o ChatGPT também apresenta uma maior densidade de conectores no conjunto de textos produzidos pelo modelo de linguagem.

Tabela 6

Frequência dos Conectores

Categorias de conectores		P*i*C	ChatGPT
Reformuladores	de paráfrase	1	2
	de retificação	-----	-----
	de síntese	2	30
Especificativos		61	168
Reforçativos		5	69
Estruturadores		2	12
Consequenciais		14	44
Contrastivos		15	68

Nos textos do P*i*C, os conectores mais usados são os especificativos (cf. Tabela 7 e (1)), os contrastivos (cf. Tabela 7 e (2)) e os consequenciais (cf. Tabela 7 e (3)). Destaca-se o facto de os conectores especificativos "como" e "por exemplo" surgirem frequentemente de forma contígua, como mostra o exemplo (4).

(1) A segurança dos ingredientes dos produtos cosméticos é analisada por organismos especializados, **tais como** a organização Cosmetic Ingredient Review.

(2) **Contudo**, a legislação europeia não define nenhum valor para as concentrações máximas em que estes ingredientes poderão ser encontrados em produtos cosméticos.

(3) **Assim**, é importante evitar a exposição a altas temperaturas e o frasco deve ficar devidamente fechado.

(4) Esta radiação está associada a queimaduras solares e ao risco de desenvolvimento do cancro de pele, **como**, **por exemplo**, o melanoma.

Já os textos do ChatGPT, por outro lado, fazem uso frequente dos conectores, mas de um modo pouco diversificado dentro do mesmo texto e quando comparados os vários textos. Um caso que ilustra esta observação diz respeito, por exemplo, às repetições de "no entanto" e "além disso" 66 e 69 vezes, respetivamente, ao longo dos textos. No *corpus*, os conectores mais utilizados são os conectores os especificativos, como "por exemplo" e "como" (cf. (5) e (6)), como nos textos do P*i*C, os reforçativos, como "além disso" (cf. Tabela 7 e (7)), com valor aditivo (cf. Filipe (2018)), os contrastivos, como "no entanto" (cf. Tabela 7), os consequenciais, como "portanto" (cf. Tabela 7), e os reformuladores de síntese, como "em resumo" (cf. Tabela 7).

(5) **Por exemplo**, em casos em que a segurança de um ingrediente é desconhecida ou quando há uma preocupação legítima com a saúde humana, é permitido realizar testes em animais.

(6) Estas alternativas incluem conservantes naturais, **como** óleos essenciais e extratos de plantas, bem como conservantes sintéticos, **como** o fenoxietanol.

(7) **Além disso**, é fundamental realizar o procedimento com um profissional qualificado, que saiba avaliar o estado do cabelo e aplicar os produtos corretamente, evitando danos aos fios.

As respostas à pergunta "*É mais eficaz um protetor solar com proteção UVA ou UVB?*" dos textos de divulgação científica apresentados na Tabela 7 ilustram estas diferenças. No texto do P*i*C, são utilizados apenas três conectores, "por sua vez", "desta forma", "por exemplo", ou seja, de acordo com Lopes e Carapinha (2013), um conector contrastivo, um consequencial e um especificativo, respetivamente, enquanto no texto gerado pelo ChatGPT são usados cinco: dois contrastivos, um consequencial, um reforçativo e um reformulador de síntese.

Tabela 7

Uso de Conectores nas Respostas do PiC e do ChatGPT

Pergunta: É mais eficaz um protetor solar com proteção UVA ou UVB?	***Prompt* 2 – *expert 1*: És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Questão: É mais eficaz um protetor solar com proteção UVA ou UVB? A resposta deverá ter entre 250 e 300 palavras.**
PiC	**ChatGPT**
A pele está permanentemente exposta à radiação solar, que inclui a radiação ultravioleta (UV). A radiação UV subdivide-se em radiação UVA, que corresponde a cerca de 90-95% da radiação UV que atinge a superfície terrestre, a radiação UVB, que constitui aproximadamente 5-10%, e a radiação UVC. Destas, as duas primeiras requerem especial atenção, uma vez que a radiação UVC não atravessa a camada de ozono e não atinge a superfície da Terra. A radiação UVA, com comprimentos de onda compreendidos entre os 320-400 nanómetros (nm), tem a capacidade de atingir não só a epiderme (camada mais superficial da pele), mas também a derme (camada mais profunda da pele), sendo também responsável por promover o bronzeamento direto. Este tipo de radiação provoca stress oxidativo, contribuindo, assim, para um envelhecimento prematuro da pele com consequente formação de rugas e flacidez cutânea. É importante salientar que a radiação UVA apresenta outros efeitos nocivos, como o aumento do risco de cancro da pele e indução de reações de fotossensibilidade. A radiação UVB, **por sua vez**, caracteriza-se por comprimentos de onda entre os 290-320 nm e atinge apenas a epiderme. Esta radiação está associada a queimaduras solares e ao risco de desenvolvimento do cancro de pele, como, **por exemplo**, o melanoma. Uma das respostas do nosso organismo à exposição de radiação UVB é o bronzeamento devido à promoção da síntese de mais melanina, a qual confere alguma proteção a este tipo de radiação. Tendo em consideração os efeitos nocivos de ambos os tipos de radiação para o nosso organismo, na hora de escolher um protetor solar, é importante optar por um que ofereça proteção simultânea contra as radiações UVA e UVB. **Desta forma**, o protetor, ao ser aplicado em períodos regulares, evita as queimaduras solares, previne o envelhecimento precoce da pele e protege de outros efeitos prejudiciais.	Ambos os raios UVA e UVB são prejudiciais para a pele e podem causar danos a longo prazo, como envelhecimento precoce, rugas, manchas e até mesmo câncer de pele. **No entanto**, é importante entender a diferença entre eles para escolher o protetor solar mais eficaz. Os raios UVA são os principais responsáveis pelo envelhecimento da pele, pois penetram profundamente na derme, a camada mais interna da pele. Eles são menos intensos do que os raios UVB, mas são mais abundantes e estão presentes durante todo o ano, mesmo em dias nublados. Os raios UVA também podem penetrar através do vidro, o que significa que você ainda está exposto a eles mesmo quando está dentro de casa ou no carro. Os raios UVB, **por outro lado**, são mais intensos e são a principal causa de queimaduras solares. Eles penetram apenas na camada mais externa da pele, a epiderme, mas são mais fortes do que os raios UVA. Os raios UVB são mais comuns durante os meses de verão e são mais fortes durante o meio do dia. **Portanto**, um protetor solar eficaz deve oferecer proteção contra ambos os tipos de raios. Procure por um protetor solar de amplo espectro, que protege contra os raios UVA e UVB. **Além disso**, certifique-se de que o protetor solar tenha um fator de proteção solar (FPS) de pelo menos 30. Isso significa que o protetor solar protege a pele contra 97% dos raios UVB. Lembre-se de aplicar o protetor solar generosamente e reaplicá-lo a cada duas horas, especialmente se estiver nadando ou suando. Além disso, use roupas de proteção, como chapéus de aba larga e óculos de sol, e evite a exposição solar direta durante as horas de pico. **Em resumo**, tanto os raios UVA quanto os UVB são prejudiciais para a pele e um protetor solar eficaz deve oferecer proteção contra ambos. Escolha um protetor solar de amplo espectro com FPS 30 ou superior e siga as recomendações de aplicação e reaplicação para garantir uma proteção adequada contra os danos causados pelo sol.

Outro dos problemas identificados nos textos gerados pelo ChatGPT prendeu-se com a presença de **repetições e/ou tautologias**, praticamente inexistentes nas respostas publicadas no P*i*C, com exceção de um caso único observado. As repetições e/ou tautologias presentes em metade das respostas do *corpus*

comprometem a clareza e a eficiência comunicativa dos textos, podendo mesmo dificultar a leitura e consequente compreensão da informação. Do mesmo modo, poder-se-á argumentar que as repetições excessivas tendem a reforçar a perceção de falta de rigor do modelo. Na Tabela 8, abaixo, apresentam-se marcados a negrito vermelho alguns exemplos de repetições e/ou tautologias presentes nas respostas geradas pelo ChatGPT.

Tabela 8

Exemplos de Repetições e/ou Tautologias nas Respostas Geradas pelo ChatGPT

Pergunta	Resposta
És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Questão: O que é o alisamento de queratina? A resposta deverá ter entre 250 e 300 palavras.	(...) O procedimento de alisamento de queratina consiste na aplicação de uma solução contendo queratina e outros ingredientes hidratantes nos cabelos, seguida de uma etapa de selagem térmica, onde os fios são alisados com o auxílio de uma prancha de calor. **A queratina presente na solução penetra nos fios, preenchendo as áreas danificadas e fortalecendo a estrutura capilar.** O calor da prancha ajuda a selar a queratina nos fios, proporcionando um efeito liso e duradouro. O alisamento de queratina é indicado para pessoas que desejam reduzir o volume e o frizz dos cabelos, além de obter um efeito liso. **É especialmente recomendado para cabelos danificados, pois a queratina ajuda a reparar os danos e fortalecer os fios.** (...)
Estás a produzir um texto de divulgação científica na área da cosmética. Tens de explicar, de uma forma simples, mas rigorosa, conceitos aos consumidores dos produtos dessa área e aos profissionais em cosmética. Questão: Os cosméticos para bebé têm "perfume"? A resposta deverá ter entre 250 e 300 palavras.	(...) Existem diferentes tipos de perfumes, alguns dos quais podem ser irritantes para a pele sensível dos bebés. No entanto, **os fabricantes de cosméticos para bebé têm o cuidado de escolher perfumes suaves e seguros para a pele dos bebés**. Os perfumes utilizados em cosméticos para bebé são geralmente formulados com ingredientes suaves e não irritantes. Estes perfumes são testados dermatologicamente **para garantir que são seguros para a pele sensível dos bebés. Além disso, os fabricantes de cosméticos para bebé seguem regulamentações rigorosas para garantir a segurança dos seus produtos**. (...)
És um perito em produtos cosméticos. O teu objetivo é esclarecer, de uma forma simples, mas rigorosa, os consumidores destes produtos e os profissionais em cosmética. Questão: O que significa a alegação "não-comedogénico" nos produtos cosméticos? A resposta deverá ter entre 250 e 300 palavras.	**A alegação "não-comedogénico" nos produtos cosméticos refere-se à capacidade do produto de não obstruir os poros da pele**, evitando assim a formação de comedões, também conhecidos como pontos negros ou cravos. Os comedões são pequenas obstruções dos poros da pele causadas pelo acúmulo de sebo, células mortas e sujidade. Quando os poros ficam obstruídos, as bactérias presentes na pele podem proliferar, levando à inflamação e ao aparecimento de acne. **Os produtos cosméticos que são rotulados como "não-comedogénicos" são formulados de forma a minimizar o risco de obstrução dos poros.** (...)

Por último, destaca-se a ocorrência frequente de **jargão científico não explicado** nos textos produzidos pelo P*i*C, fenómeno muito menos frequente nas respostas geradas pelo ChatGPT (cf. (8) e (9)). Verificou-se que os exemplos apresentados na Tabela 9 não são explicados nem no corpo do texto nem através de hipertexto

para outras respostas publicadas no portal ou para conteúdo externo ao portal que defina e/ou contextualize o termo em questão. A presença de jargão científico não explicado pode dificultar grandemente a compreensão da mensagem por parte de um leitor não especializado no domínio da cosmetologia e comprometer a acessibilidade e a visibilidade da informação, princípios fundamentais dos textos de divulgação científica, como asseveram Molek-Kozakowska (2017) e Aubakir (2017). Não obstante, esta questão pode ser facilmente resolvida através da explicação da terminologia no corpo do texto ou da disponibilização de hiperligações (assegurando-se a sua constante atualização). Como anteriormente, esta estratégia foi, do mesmo modo, sugerida à equipa responsável pela produção científica do portal, visando-se melhorar a acessibilidade e a clareza da informação.

Tabela 9

Exemplos de Jargão Científico Não Explicado nas Respostas do PiC Analisadas

Pergunta	Resposta
Os nanomateriais são permitidos em produtos cosméticos?	(...) Entre as vantagens da utilização de nanomateriais em produtos cosméticos salientam-se: a melhoria da **cosmeticidade** das formulações (textura, facilidade de aplicação, estética, conservação), controlo da libertação, aumento da eficácia dos ingredientes, otimização da fotoproteção e do desempenho dos produtos de maquilhagem. (...)
Como se calcula o Fator de Proteção Solar?	(...) O FPS pode ser calculado através da razão entre a **Dose Eritematosa Mínima** (DEM) da pele protegida por um protetor solar, e a DEM da pele não protegida. Assim, se o produto cosmético apresenta um valor de FPS igual a 30, significa que é necessária uma exposição solar 30 vezes maior para produzir **eritema**, comparativamente a uma situação em que este indivíduo não estaria a usar protetor solar. Quanto maior for o FPS, maior será a proteção do produto cosmético, sendo recomendado, pelo menos, um FPS de 30. De uma forma geral, para determinar o FPS, um grupo de voluntários com um **fototipo entre I e III** é submetido à exposição de doses de radiação UV progressivas, emitidas por uma fonte de luz artificial. (...)
Quais são os constituintes típicos de um verniz de unhas?	(...) De forma a garantir a distribuição uniforme destes constituintes, pode ser adicionado um **aditivo reológico**, como as bentonites, sílicas, celuloses, acrilatos, uretanos, entre outros. Podem ainda ser utilizados outros ingredientes, como os filtros de radiação ultravioleta (UV), com o objetivo de proteger o produto, evitando o desvanecimento da cor.

(8) Portanto, é essencial seguir as instruções de uso do produto e realizar um teste de **patch** antes de usar qualquer produto cosmético.

(9) Os corantes diretos são moléculas de cor que penetram na cutícula do cabelo e se depositam no **córtex**, conferindo cor ao cabelo.

5. Considerações finais

O presente estudo permitiu investigar o desempenho do ChatGPT quanto à produção de conteúdo dentro dos textos de divulgação científica na área da cosmética, tendo como objetivo a análise e o estabelecimento de padrões linguísticos e textuais que permitissem identificar as diferenças e/ou semelhanças entre a produção de conteúdo científico especializado da área da cosmética por humanos especialistas num portal de divulgação científica – o *Portal infoCosméticos* (P*i*C) – e por um LLM.

A análise linguística e textual produziu resultados de diferentes tipos e permitiu identificar problemas em ambos os textos analisados, o que levou à reflexão sobre as características dos textos de divulgação científica, escritos em português europeu, contribuindo para a caracterização destes textos.

Assim, por um lado, as respostas geradas pelo ChatGPT, destacaram-se pela estrutura tripartida, garantindo uma estrutura composicional global bem demarcada e sem lacunas estruturais evidentes, nos textos gerados a partir dos *prompts* 2, 3 e 4. Além disso, a ausência de jargão científico contribuiu para uma maior acessibilidade da informação, tornando a mensagem compreensível para um público não especializado. No entanto, registaram-se limitações significativas, nomeadamente quanto ao uso frequente de repetições e/ou tautologias, que comprometem a clareza e a eficiência comunicativa dos textos. Verificou-se, também, ausência de exemplos, o que afetou a resposta ao tópico que a pergunta pretendia responder, não sendo, por isso, tão eficiente do ponto de vista comunicativo. No que diz respeito à coesão interfrásica, observou-se uma elevada frequência de conectores ao longo dos vários textos, apesar da baixa diversidade, sendo mais frequentes os conectores reforçativos, especificativos, contrastivos, consequenciais e reformuladores de síntese. A avaliação global do desempenho linguístico do ChatGPT situou-se nos 13,43 pontos em 20, com uma variabilidade dos resultados expressa por uma dispersão de $\sigma = 2{,}5$.

As respostas apresentadas no P*i*C, por outro lado, obtiveram um desempenho médio global ligeiramente superior (13,6/20 pontos) face à pontuação obtida pelas respostas geradas pelo ChatGPT, apresentando uma dispersão de $\sigma = 1{,}9$. As respostas demonstram ausência quase total de repetições e/ou tautologias e incluem consistentemente exemplos, tanto inseridos no corpo de texto como apresentados sob a forma de hipertexto, o que favoreceu a clareza da informação apresentada. Contudo, a estrutura composicional global revelou algumas fragilidades, nomeadamente no que diz respeito à ausência frequente do parágrafo de conclusão. Simultaneamente, o uso frequente de jargão científico não explicado pode dificultar a compreensão da mensagem por parte de um leitor não especializado no domínio da cosmética. Por fim, ao nível da coesão interfrásica, a baixa frequência de conectores explícitos pode, igualmente, comprometer a compreensão do texto, dado que funcionam como pistas de como o leitor deve interpretar o texto. Os conectores mais frequentes são os especificativos, contrastivos e consequenciais. Estas questões podem ser facilmente resolvidas, tendo sido apresentadas soluções à equipa responsável pela produção de conteúdos do portal.

Com base nestes resultados, futuros trabalhos poderão permitir aprofundar os mecanismos de coesão interfrásica nos textos gerados pelo ChatGPT, analisando detalhadamente a diversidade e adequação do uso de conectores por parte dos LLMs. Além disso, a rápida evolução dos modelos de inteligência artificial torna relevante a necessidade de reavaliações periódicas no que ao desempenho de versões atualizadas do ChatGPT diz respeito. Novas iterações do estudo podem permitir verificar potenciais melhorias na construção textual, na coesão e na eficácia comunicativa dos textos de divulgação científica, por exemplo.

No âmbito do P*i*C, por sua vez, este estudo contribuirá para a melhoria do guião orientador da redação de textos de divulgação científica, visando a consistência ao nível da estrutura composicional global, mas também a uniformização quanto ao uso de conectores e a sensibilização para a utilização de jargão científico explicado. Concomitantemente, a dinamização de oficinas de escrita destinadas aos colaboradores do P*i*C na área das Ciências Farmacêuticas poderá contribuir para o aperfeiçoamento das práticas de redação, reforçando a clareza, a coesão, a coerência e a eficiência comunicativa dos textos produzidos.

Referências

Adam, J.-M. (2005). *La linguistique textuelle*. Armand Colin.

Almeida, I. M. & Silvano, P. (2022). Portal infoCosméticos: Um projeto colaborativo FFUP/ FLUP de divulgação científica. In *Livro de atas do 7.º Congresso Nacional de Práticas Pedagógicas no Ensino Superior* (pp. 288–294). UA Editora. https://doi.org/10.48528/yhzq-cp97

Aubakir, N. (2017). Distinctive features of popular science discourse. *World Science, 2*(11), 32–35.

Bronckart, J.-P. (1997). *Activité langagière, textes et discours : pour un interactionisme socio-discursif*. Delachaux et Niestlé.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). *Language models are few-shot learners*. arXiv. https://arxiv.org/abs/2005.14165

Ferreira, M. S., Cuboia, N., Azevedo, L. F. R., Magalhães, O. E., Almeida, I. M. de, & Ribeiro, V. M. (2024). Quality of information about cosmetic products in social media: One-year analysis of Portuguese influential Instagram profiles. *Journal of Consumer Health on the Internet, 28*(3), 203–223. https://doi.org/10.1080/15398285.2024.2381281

Filipe, L. (2018). O artigo de divulgação científica – algumas marcas de género. In M. Gonçalves & N. Jorge (Org.), *Literacia científica na escola* (pp. 79–89). NOVA FCSH-CLUNL.

Garcés-Conejos, P., & Sanchéz-Macarro, A. (1998). Scientific discourse as interaction: Scientific articles vs. popularizations. In A. Sanchez-Macarro & R. Carter (Eds.), *Linguistic choice across genres* (pp. 173–190). John Benjamins.

Gonçalves, M. & Jorge, N. (Org.) (2018). *Literacia científica na escola*. NOVA FCSH-CLUNL.

Gonçalves, M. & Magalhães, M. (2019). Corpus e géneros textuais nas práticas de divulgação de ciência ou as novas hierarquias na construção do conhecimento. *Revista da Associação Portuguesa de Linguística*, *5*, 145–157. https://doi.org/10.26334/2183-9077/rapln5ano2019a1

Onder, C. E., Koc, G., Gokbulut, P., Taskaldiran, I., & Kuskonmaz, S. M. (2024). Evaluation of the reliability and readability of ChatGPT-4 responses regarding hypothyroidism during pregnancy. *Scientific Reports, 14*(1), 243. https://doi.org/10.1038/s41598-023-50884-w

Myers, G. (1991). Lexical cohesion and specialized knowledge in science and popular science texts. *Discourse Processes*, *14*, 1–26.

Molek-Kozakowska, K. (2016). Stylistic analysis of headlines in science journalism: a case study of New Scientist. *Public Understanding of Science*, 1–14.

Molek-Kozakowska, K. (2017). Communicating environmental science beyond academia: stylistic patterns of newsworthiness in popular science journalism. *Discourse & Communication, 11*(1), 69–88.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations, and future scope. *Internet of Things and Cyber-Physical Systems*, *3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Moraes, L. C., Silvério, I. C., Marques, R. A. S., Anaia, B. C., Paula, D. F., Faria, M. C. S., Cleveston, I., Correia, A. S., & Freitag, R. M. K. (2024). *Análise de ambiguidade linguística em modelos de linguagem de grande escala (LLMs)*. arXiv. https://arxiv.org/abs/2404.16653

Lopes, A. C. M. & Carapinha, C. (2013). *Texto, coesão e coerência*. Almedina.

Torres, A., Pacheco, A. F., Guimarães, N., Silvano, P., Ferreira, A. & Almeida, I. F. (2024, 10-11 julho). *How reliable is ChatGPT information regarding cosmetic products and ingredients?* [Comunicação oral]. Cosmetinnov 2024, Lisboa, Portugal.

Ye, J., Chen, X., Xu, N., Zu, C., Shao, Z., Liu, S., Cui, Y., Zhou, Z., Gong, C., Shen, Y., Zhou, J., Chen, S., Gui, T., Zhang, Q. & Huang, X. (2023). *A comprehensive capability analysis of GPT-3 and GPT-3.5 series models*. arXiv. https://arxiv.org/abs/2303.10420

Zamboni, L. M. S. (1998). Heterogeneidade e subjetividade no discurso da divulgação científica. *Sínteses – Revista dos Cursos de Pós-Graduação, 3*, 389–395.
Zhang, Y., Liu, X., Sun, Y., Alharbi, A., Alzahrani, H., Alomair, B. & Song, D. (2025). *Can LLMs design good questions based on context?* arXiv. https://arxiv.org/abs/2501.03491
Zhou, L., Schellaert, W., Martínez-Plumed, F., Moros-Daval, Y., Ferri, C. & Hernández-Orallo, J. (2024). Larger and more instructable language models become less reliable. *Nature*, *634*, 61–68. https://doi.org/10.1038/s41586-024-07930-y

Anexo

Tabela 1

Grelha de avaliação linguística e textual

Análise linguística			
1. A resposta está de acordo com as características dos textos de divulgação científica?			
Não	Parcialmente	Quase completamente	Sim
0	**1**	**2**	**3**
2. A resposta é eficiente do ponto de vista comunicativo, respondendo ao tópico?			
Não	Parcialmente	Quase completamente	Sim
0	**1**	**2**	**3**
3. A resposta está bem estruturada?			
Não estruturada [a resposta não inclui introdução, desenvolvimento e conclusão]	Semiestruturada [a resposta inclui introdução, desenvolvimento ou conclusão]	Quase bem estruturada [a resposta inclui introdução, desenvolvimento e conclusão, com lacunas]	Bem estruturada [a resposta inclui introdução, desenvolvimento e conclusão]
0	**1**	**2**	**3**
4. A resposta é clara?			
Não clara [a resposta não é concisa, direta ou simples]	Algo clara [a resposta é clara em certa medida]	Clara [a resposta é concisa, direta e simples]	
0	**1**	**2**	
5. A resposta é coesa?			
Não coesa [a resposta inclui poucos mecanismos de coesão referencial, temporal e interfrásica]	Algo coesa [a resposta inclui alguns mecanismos de coesão referencial, temporal e interfrásica, com lacunas]	Coesa [a resposta inclui os mecanismos de coesão referencial, temporal e interfrásica adequados]	
0	**1**	**2**	
6. A resposta é coerente?			
Não coerente [as ideias não estão encadeadas logicamente]	Algo coerente [as ideias estão encadeadas logicamente, com lacunas]	Coerente [as ideias estão encadeadas logicamente]	
0	**1**	**2**	
7. Quando presente, o jargão científico é explicado?			
Não		Sim	
0		**1**	
8. A resposta inclui exemplos?			
Não		Sim	
0		**1**	
9. A resposta apresenta erros ortográficos?			
Sim		Não	
0		**1**	
10. A resposta apresenta erros linguísticos (morfossintáticos, lexicais ou semânticos)?			
Sim		Não	
0		**1**	
11. A resposta apresenta repetições/tautologias?			
Sim		Não	
0		**1**	
Pontuação global: -/20			

Um estudo exploratório em Pragmática Cross-cultural sobre o ato de fala de desmarcar compromissos de última hora em italiano, português europeu e português brasileiro

An exploratory study in Cross-cultural Pragmatics on the speech act of last-minute cancellation in Italian, European Portuguese and Brazilian Portuguese

Silvia Pallone[1]

[1] Sapienza Università di Roma / Università Roma Tre / Universidade de São Paulo

Abstract

This exploratory study investigates the discourse strategies employed by European Portuguese, Brazilian Portuguese, and Italian speakers when cancelling appointments, particularly at the last minute, in informal contexts. The analysis is based on 147 cancellations collected via messaging applications, supplemented by participants' metapragmatic reflections. Two main patterns were identified: (1) an explicit structure centred on the cancellation sub-act and (2) an indirect strategy relying on justifications and apologies or offers of repair. Findings reveal a convergence in the use of both positive and negative politeness strategies, with cultural differences in the acceptance of last-minute cancellations. The results suggest a stronger orientation towards negative politeness in European Portuguese and Italian, contrasting with a positive politeness orientation in Brazilian Portuguese, which aligns with pragmatic patterns observed in other Latin American cultures. The theoretical framework of this study is grounded in research on Verbal Politeness and Cross-cultural Pragmatics, particularly the works of Brown and Levinson (1987), Blum-Kulka et al. (1989), Nuzzo and Cortés Velásquez (2020), Santoro (2017) and Seara (2024).

Keywords: cross-cultural pragmatics, last-minute cancellation, Portuguese, Italian, linguistic politeness.

Resumo

Este estudo exploratório investiga as estratégias discursivas utilizadas por falantes de português europeu, português brasileiro e italiano ao desmarcar compromissos, nomeadamente de última hora, em contextos informais. A análise baseia-se em 147 cancelamentos recolhidos por meio de aplicações de mensagens e complementados por reflexões de natureza metapragmática por parte dos participantes. Identificámos dois esquemas principais para a realização do ato: (1) uma estrutura explícita centrada no sub-ato de cancelamento e (2) uma estratégia indireta baseada em justificações e pedidos de desculpa ou propostas de remarcação. Observou-se uma convergência no uso de estratégias de cortesia positiva e negativa, com diferenças culturais na aceitação do cancelamento tardio. Os resultados sugerem uma maior orientação para a cortesia negativa em português europeu e italiano, face a uma orientação positiva no português brasileiro, que se alinha com padrões pragmáticos de outras culturas latino-americanas. O enquadramento teórico deste estudo tem como base as investigações no âmbito da Cortesia verbal e da Pragmática Cross-cultural, com destaque para os trabalhos de Brown e Levinson (1987), Blum-Kulka et al. (1989), Nuzzo e Cortés Velásquez (2020), Santoro (2017) e Seara (2024).

Palavras-chave: pragmática cross-cultural, desmarcar compromissos de última hora, português, italiano, cortesia linguística.

1. Introdução

A partir do final dos anos oitenta do século passado, tem-se assistido à proliferação de estudos empíricos no domínio da investigação da Pragmática Cross-cultural[1] (cf. Blum-Kulka et al., 1989; Verschueren, 2016; Wierzbicka, 1985, 2003), cujo principal objetivo é a realização de "synchronic comparisons between two, or a small number of, co-existent language usage systems" (Verschueren, 2016, p. 3). Em particular, neste âmbito, os investigadores têm-se dedicado à comparação dos atos linguísticos em diferentes línguas e culturas, partindo da Teoria dos Atos de Fala (Austin, 1962; Searle, 1969) e da Teoria da Cortesia (Brown & Levinson, 1987), e foi observado que existe uma grande variabilidade entre as comunidades de falantes. A razão subjacente a esta variação nos usos linguísticos reside no facto de estes serem fortemente influenciados por normas sociopragmáticas que variam significativamente consoante a cultura. A este respeito, afirma Santoro (2017, p. 180, tradução nossa)[2]:

> A variabilidade é um dos conceitos-chave sobre os quais se funda a pragmática linguística, que estuda os efeitos do contexto sobre o que fazemos ao falar e aborda a língua do ponto de vista dos utilizadores. Este princípio é central para a pragmática cross-cultural ou contrastiva, que se dedica a analisar como as "normas" da língua mudam de cultura para cultura, não apenas quando o sistema linguístico utilizado é diferente, mas também quando, apesar de ser formalmente o mesmo, na prática não o é, uma vez que a cultura de referência é outra.

Assim sendo, o respeito destas normas por parte dos falantes revela-se essencial para o sucesso comunicativo numa determinada cultura, uma vez que dominar uma língua implica também o facto de saber "conduzir procedimentos de ação reconhecidos, reconhecíveis e partilhados por um determinado grupo social"[3], conforme esclarece Caffi (2017, p. 19).

A comparação entre línguas e culturas na realização dos atos de fala permite evidenciar diferenças que podem levar a mal-entendidos. Investigar estas diferenças mesmo entre falantes de línguas e culturas afins, como no presente estudo, torna-se ainda mais relevante, pois o seu desconhecimento pode aumentar a probabilidade de subestimá-las e de incorrer em incompreensões na comunicação intercultural (Santoro et al., 2019, p. 170). Com efeito, estas falhas interacionais, definidas por Thomas (1983) como *pragmatic failures*, podem ser erroneamente associadas ao caráter do indivíduo ou do grupo étnico a que pertence, contribuindo para a criação ou o reforço de estereótipos e preconceitos (Nuzzo & Santoro, 2017). Este fenómeno torna-se particularmente evidente no uso de uma segunda língua, dado que, frequentemente, os falantes tendem a refletir a sua própria língua e cultura no seu agir linguístico e nas suas expectativas relativamente ao comportamento do interlocutor, independentemente do seu nível de competência linguístico-comunicativa (cf. Cortés Velásquez & Nuzzo, 2021; Nuzzo & Santoro, 2017). Além disto, no caso da comunicação mediada por computador (CMC), nomeadamente em aplicações de mensagens instantâneas, o risco de mal-entendidos pode manifestar-se com maior frequência do que na comunicação presencial. Tal ocorre devido à ausência de simultaneidade interacional e de elementos proxémicos, que limitam as possibilidades de negociação do significado (Brocca, 2021; Pallotti, 2019). Portanto, compreender a existência de uma variabilidade situacional e cultural e o seu impacto no discurso e nas estratégias interacionais de comunicação permite aos falantes desenvolver uma consciência metapragmática mais apurada, contribuindo para a minimização de conflitos (Santoro, 2017).

Nesta perspetiva, foi realizado um estudo exploratório sobre o ato de fala de desmarcar compromissos, produzido por falantes nativos de italiano e português, nas variedades europeia e brasileira. Mais concretamente,

[1] Neste trabalho, esta área de investigação é assim definida por derivação da expressão inglesa *cross-cultural pragmatics*.
[2] No original, "Quello della variabilità è uno dei concetti chiave su cui si fonda la pragmatica linguistica che studia gli effetti del contesto su ciò che facciamo parlando e si occupa della lingua dal punto di vista degli utenti. Questo principio è centrale per la pragmatica cross-culturale o contrastiva che si dedica ad analizzare come le "norme" della lingua cambiano da cultura a cultura, non solo nel caso in cui il sistema linguistico utilizzato è diverso, ma anche quando, pur essendo lo stesso da un punto di vista formale, nella pratica non lo è, visto che è un'altra la cultura di riferimento" (Santoro, 2017, p. 180).
[3] No original, "saper condurre procedure di azione riconosciute, riconoscibili, e condivise da un dato gruppo sociale" (Caffi, 2017, p. 19).

o objetivo desta investigação é conduzir uma análise qualitativa das estratégias pragmalinguísticas identificadas num *corpus* naturalístico[4] de 147 cancelamentos de compromissos, produzidos de forma espontânea através de mensagens escritas enviadas em aplicações de mensagens. A análise foi realizada a partir de estudos empíricos fundamentados na Teoria dos Atos de Fala (cf. Blum-Kulka et al., 1989; Brocca et al., 2023; Cortés Velásquez, 2017; Cortés Velásquez & Nuzzo, 2017; Félix-Brasdefer, 2003; Nuzzo & Cortés Velásquez, 2020), e foi posteriormente complementada por reflexões de natureza metapragmática por parte dos participantes, obtidas através de entrevistas.

2. O ato de fala de desmarcar compromissos na perspetiva cross-cultural

Os primeiros estudos que se focalizaram no ato de fala de desmarcar compromissos, em particular no último momento (*last-minute cancellations*), a partir da perspetiva cross-cultural, foram realizados por Cortés Velásquez (2017), Nuzzo e Cortés Velásquez (2020) e Brocca et al. (2023) e compararam as produções de falantes nativos de italiano, espanhol colombiano e alemão austríaco. Nestes estudos, tal ato é identificado como uma manifestação discursiva distinta, ainda que correlacionada ao ato de recusa. Nomeadamente, trata-se de uma das estratégias discursivas possíveis de rejeição ou não aceitação empregadas pelos falantes em resposta a um convite. Segundo os autores mencionados, recusas e cancelamentos de compromissos são atos comissivos (Searle, 1976), uma vez que vinculam o falante ao cumprimento de uma determinada ação, e desempenham a mesma função: ou seja, informar o interlocutor de que não será possível participar no evento que ele propõe. Todavia, foi observado que estes atos diferem em dois aspetos, em relação ao momento em que são realizados e em termos de cortesia linguística. Relativamente ao primeiro, a recusa é um ato reativo, realizado em resposta a um convite, pedido ou proposta; enquanto o cancelamento é um ato proativo, realizado pela pessoa convidada, que comunica a impossibilidade de participar no evento para o qual o convite do interlocutor tinha sido previamente aceite (Nuzzo & Cortés Velásquez, 2020, p. 335). No que diz respeito ao âmbito da cortesia linguística, ambos os atos pertencem à categoria dos *Face-Threatening Act*s (FTAs), os atos ameaçadores para a *face*, na terminologia de Brown e Levinson (1987). O conceito de "face", introduzido por Goffman (1959, 1967) e retomado posteriormente por Brown e Levinson (1978, 1987), indica a imagem pública que cada indivíduo possui e compõe-se de duas vertentes: a face positiva e a face negativa. A face *positiva* representa a imagem que cada indivíduo expõe à sociedade, esperando que esta seja aceite e aprovada pelos outros, e, consequentemente, a necessidade de ser apreciado e sentir-se parte de um determinado grupo; já a face *negativa* é o desejo do indivíduo de ter o seu próprio território e a sua liberdade de ação protegidos e respeitados.

Neste contexto, segundo Nuzzo e Cortés Velásquez (2020, pp. 335–336), o ato da recusa e, em particular, recusar um convite assim que este é feito, representa uma ameaça à face positiva do convidado, porque indica o não cumprimento do desejo de quem fez o convite de partilhar tempo e afinidade com o convidado (falta de solidariedade). Por conseguinte, o falante pode empregar estratégias para minimizar esta ameaça e, entre elas, existe também a opção de não realizar o ato, ou adiá-lo para um momento posterior, por meio de um cancelamento. Neste caso, um cancelamento de última hora é potencialmente ameaçador para a face negativa, pois interfere nos planos da pessoa que convida (invasão do território do alocutário), mas, ao mesmo tempo, pode salvar tanto a face positiva como a negativa porque o convidado atribui a sua ausência a uma circunstância imprevista e inevitável, fora ou contra da sua vontade. Segundo os investigadores, a preferência pelo emprego de um ou o outro ato de fala pode variar de acordo com as regras das diferentes comunidades e culturas. A este propósito, Cortés Velásquez (2017) e Nuzzo e Cortés Velásquez (2020) recolheram dados sobre a realização e a perceção do ato de fala de desmarcar compromissos por parte de italianos e colombianos por meio de um *Discourse Completion Task* escrito (DCT)[5] e observaram um fenómeno interessante: para os colombianos, o

[4] Na terminologia inglesa "*naturally occurring data*" (cf. Golato, 2017).

[5] Instrumento de elicitação de dados frequentemente adotado no campo da pragmática cross-cultural que consiste em breves diálogos escritos, geralmente precedidos de uma série de indicações contextuais, em que um turno de fala é omitido e o informante deve completá-

cancelamento de última hora após um convite é considerado uma prática comum e parece ser muito mais raro recusar imediatamente; ao passo que, para os italianos, seria preferível uma recusa imediata em resposta a um convite. Investigar este aspeto assume um papel relevante, uma vez que pode ter efeitos sobre as estratégias de cortesia linguísticas adotadas pelos falantes para atenuar a força ilocutória do enunciado (Cortés Velásquez, 2017). Nomeadamente, em termos de cortesia, os resultados destes estudos apontam que os italianos demonstraram maior preocupação com a proteção da face negativa do interlocutor do que os colombianos, que privilegiaram a valorização da face positiva, enfatizando o apreço pelo convite.

Apesar de diversos estudos terem abordado este âmbito de investigação, parece faltar uma análise mais aprofundada do contexto lusófono, assim como uma comparação entre este e o contexto italiano, em particular na área da comunicação mediada por computador.

Ainda assim, o italiano, o português europeu e o português brasileiro constituem candidatos adequados para a comparação entre línguas aparentadas. Embora pertençam ao domínio das línguas românicas e partilhem certas afinidades culturais, é plausível supor que apresentem diferenças nos usos pragmáticos. Essa hipótese é corroborada por estudos de natureza cross-cultural que já compararam o italiano e o português brasileiro (cf. Benine & Santoro, 2022; Rocha et al., 2015; Santoro, 2017). A este respeito, Benine e Santoro (2022, p. 113) afirmam que "no caso de estudos cross-culturais, podem ter um peso as diferenças culturais e o fato de que a mesma situação pode ser percebida de formas distintas dependendo da cultura de cada falante". Nesta perspetiva, o Português parece representar um bom exemplo de língua materna partilhada por comunidades de diferentes culturas, como refere Mateus (2003, p. 87), e cujas normas sociais podem diferir significativamente.

3. Os atos de fala no *instant messaging*

Hoje em dia, uma grande parte da nossa interação diária ocorre em ambientes virtuais, nas redes sociais (Facebook, Instagram, Twitter/X), fóruns de discussão, blogs, e-mail e aplicações de mensagens (WhatsApp, Telegram, Messenger), através de dispositivos eletrónicos (Derks et al., 2008; Herring, 2004). Nas últimas décadas, a comunicação mediada por computador (CMC) tem sido objeto de estudos de diferentes áreas que enfatizaram as suas características idiossincráticas (cf. Androutsopoulos, 2006; Crystal, 2001; Dresner & Herring, 2010; Herring, 2004; Locher, 2010; Simone, 2012). O presente trabalho foca-se numa tipologia de CMC, o *instant messaging* (IM), na qual os interlocutores interagem por meio da linguagem verbal, geralmente digitada num teclado e lida como texto num ecrã (de telemóvel, computador ou tablet), e onde os corpos físicos (e suas ações) são tecnicamente inexistentes (Herring, 2004). É evidente que a CMC se tem vindo a tornar cada vez mais multimodal e a interação *online* prevê outros sistemas semióticos, como emojis, imagens, vídeos ou mensagens de voz, mas a linguagem escrita ainda constitui um componente essencial da comunicação *online* (Crystal, 2001; Herring, 2004; Yus, 2011).

Em relação à linguagem escrita em formato digital, diversos linguistas apontaram um fenómeno de hibridismo e de contaminação entre elementos retirados da língua escrita e falada (cf. Flores-Salgado & Castineira-Benitez, 2018; Sánchez-Moya & Cruz-Moya, 2015; Seara, 2024; Yus, 2011). Por exemplo, os autores que analisaram conversas escritas na aplicação WhatsApp mencionaram o uso de uma linguagem bastante próxima do discurso falado. De facto, o discurso digital tem sido definido como texto resultante do processo de transferência do discurso oral para a forma escrita, muitas vezes com ícones (emojis, emoticons, GIFs, etc.) usados tanto no momento de reforçar algumas ações comunicativas quanto para comunicar sentimentos difíceis de transmitir num estilo escrito (Dresner & Herring, 2010; Flores-Salgado & Castineira-Benitez, 2018; Herring, 2004). Na verdade, se olharmos para a forma como usamos pontuação, maiúsculas e emojis, vemos como a linguagem escrita se adaptou para substituir os elementos suprassegmentais típicos da comunicação oral (Baron & Ling, 2011; Longo, 2020; Sánchez-Moya & Cruz-Moya, 2015). A este propósito,

lo por escrito a partir do que diria, caso estivesse em contextos semelhantes aos propostos (cf. Blum-Kulka et al., 1989; Félix-Brasdefer, 2010; Leech, 2014; Ogiermann, 2018; Schneider, 2018).

para o contexto italiano, Antonelli (2014, p. 543) justifica este hibridismo textual afirmando que o italiano escrito digitalmente criou as condições para a afirmação de um italiano escrito informal. A impressão de proximidade com a oralidade que se tem na escrita telemática deve-se em grande parte a essa informalidade, que leva os utilizadores da rede a sobrepor a "gramática informal" e a "gramática da fala". Segundo o autor, várias formas de "neo-epistolaridade tecnológica" (e-mail, SMS e mensagens instantâneas) causaram um importante regresso à comunicação escrita no dia a dia, e isto determina a dissolução das coordenadas que há séculos caracterizavam e condicionavam a escrita. A escrita transita para o domínio do efémero; à medida que escrevemos com tanta frequência, o ato de escrever torna-se um gesto quotidiano, extremamente distante da solenidade que sempre o envolveu.

Apesar de a crítica se ter debruçado com grande interesse sobre a CMC, no âmbito de pesquisa da pragmática linguística continuam a faltar estudos empíricos que tenham analisado as estratégias de cortesia adotadas pelos falantes ao enviar uma mensagem instantânea. Um dos poucos trabalhos que têm como foco a análise das estratégias pragmalinguistícas no *instant messaging* é o de Flores-Salgado e Castineira-Benitez (2018). As investigadoras estudaram o ato de fala do pedido em interações espontâneas no WhatsApp entre falantes de espanhol do México: a investigação centrou-se na escolha das estratégias de cortesia, revelando que os participantes recorreram predominantemente a estratégias convencionalmente indiretas, acompanhadas de um uso significativo de modificação sintática. Além disso, verificou-se que as sequências de abertura e fecho estavam presentes em todas as interações analisadas. Em particular, as saudações foram amplamente preferidas em detrimento de formas de tratamento deferentes, o que pode indicar uma estratégia de construção de solidariedade entre os interlocutores. No que diz respeito à variação diamésica, salientou-se que os pedidos escritos e enviados por meio de mensagens instantâneas podem ser formulados em vários turnos, frequentemente introduzidos por um elemento preparatório, e podem ser atenuados através do uso de pictogramas, como emoticons ou emojis.

No caso em análise, da produção do ato de fala de desmarcar compromissos no IM, Brocca (2021) investigou este ato de fala no âmbito do italiano como língua segunda. Foi recolhido um *corpus* de escrita digital através de DCTs realizados no WhatsApp por aprendentes germanófonos de italiano e por falantes nativos. Os questionários tinham como objetivo elicitar os atos de fala do pedido e do cancelamento em contextos de crescente distância social. A análise, conduzida numa perspetiva contrastiva, revelou diferenças significativas na realização do ato de cancelamento entre os dois grupos: os falantes nativos produziram cancelamentos compostos por vários sub-atos, enquanto os aprendentes apresentaram um número reduzido de sub-atos e um uso limitado do sub-ato de justificação. Estes resultados sublinham a importância de sensibilizar os falantes para as suas escolhas comunicativas e para as estratégias sociopragmáticas a adotar em função do contexto linguístico-cultural em que estão inseridos.

Como afirma Locher (2010), é particularmente relevante investigar de que forma as questões interpessoais de cortesia e descortesia são negociadas na interação *online* e como estas se articulam com a construção da identidade e a negociação da face. Ambos os aspetos estão ligados ao conceito de cortesia linguística e transmitem-se durante o discurso, através das escolhas linguísticas feitas pelos falantes (cf. Amossy, 2010; Caffi, 2017; Locher, 2010; Maingueneau, 2005; Spencer-Oatey, 2007). Por outras palavras, o enunciador constrói e negocia a própria identidade e imagem social com o seu interlocutor, a partir do discurso e da interpretação do papel que vão desempenhando em cada situação comunicativa (Caffi, 2017, p. 56). Nesta perspetiva dinâmica, de "co-construção" da imagem do locutor e do alocutário no momento da enunciação, os falantes podem optar por utilizar estratégias de cortesia na realização dos atos de fala seguindo os critérios de adequação metapragmática e metacomunicativa reconhecidos e partilhados pela sua própria comunidade (2017, p. 56).

É fundamental salientar que a comunicação digital é assíncrona, embora os tempos de resposta possam variar em rapidez. Isto implica que os interlocutores podem estar presentes ou ausentes em qualquer momento, uma vez que possuem a liberdade de entrar e sair da rede conforme a sua conveniência (Cabral, 2014, p. 516). Como consequência, as mensagens não são necessariamente recebidas e respondidas em tempo real, sendo

comum que haja um intervalo até à resposta do interlocutor. Estas dinâmicas intensificam tanto as expectativas quanto as perceções dos utilizadores, sejam elas positivas ou negativas, ressaltando, assim, a centralidade da linguagem verbal e das estratégias de cortesia na interação digital (Cabral, 2014, p. 516).

Atualmente, como sublinham Flores-Salgado e Castineira-Benitez (2018), a investigação sobre os atos de fala em IM continua a ser uma área pouco estudada, não existindo diretrizes linguísticas para a redação de mensagens neste meio. Consequentemente, os utilizadores podem sentir insegurança quanto às estratégias de cortesia e às formas linguísticas mais adequadas consoante a situação linguístico-comunicativa, especialmente no caso de contextos multiculturais.

Apenas se registam algumas tentativas para investigar empiricamente o ato de fala do cancelamento de última hora após um convite, apesar de se tratar de uma situação comunicativa bastante comum. Até à data, o contexto lusófono parece carecer ainda de um estudo mais aprofundado sobre a realização deste ato no IM. Este estudo preliminar pretende colmatar esta lacuna e constituir uma base para o desenvolvimento de futuras investigações sobre o uso dos atos de fala em aplicações de mensagens.

4. Metodologia

Nesta secção, será apresentada a metodologia escolhida para esta investigação de caráter exploratório: apresentaremos o processo de constituição do *corpus*, a tipologia de informantes que constituiu a amostra e a taxonomia utilizada para a análise do ato de fala de desmarcar compromissos.

4.1. Objetivos

O objetivo principal do presente trabalho é analisar qualitativamente o ato de fala de desmarcar compromissos, produzido por falantes nativos de italiano e português, nas variedades europeia e brasileira, em ambientes virtuais. Por outras palavras, o estudo visa investigar as estratégias discursivas empregadas na CMC pelos informantes para a realização de cancelamentos de compromissos, em particular cancelamentos de última hora, e as suas perceções quanto à possibilidade e necessidade da sua realização.

Mais concretamente, questionámo-nos sobre se o cancelamento de última hora após um convite é considerado uma prática comum por falantes portugueses e brasileiros, conforme já observado por Cortés Velásquez (2017) e por Cortés Velásquez e Nuzzo (2021) no contexto linguístico-cultural colombiano, ou se, pelo contrário, esse comportamento é percebido como indelicado, à semelhança do que ocorre no contexto linguístico-cultural italiano. De acordo com Gomes (2020), os brasileiros teriam mais dificuldade em formular recusas diretas em resposta a um convite, pedido ou proposta, ou seja, a dizer "não" explicitamente. Consequentemente, podemos assumir que sejam mais propensos a recorrer a cancelamentos de última hora do que portugueses e italianos.

Em suma, a presente investigação propõe-se, neste sentido, a:

i. Identificar estratégias pragmalinguísticas usadas pelos falantes para a realização de cancelamentos de compromissos em interações virtuais entre amigos;

ii. Investigar as perceções de portugueses, brasileiros e italianos sobre este ato de fala em relação aos seus hábitos socioculturais.

4.2. Constituição do *corpus*

Conforme expusemos anteriormente, esta investigação pretende considerar fenómenos linguístico-discursivos relacionados com o ato de fala de desmarcar compromissos em contextos informais na comunicação mediada por computador, através da análise de dados empíricos produzidos, de forma espontânea, por falantes

nativos de italiano, português europeu e português brasileiro. Para o efeito, foi conduzido um estudo exploratório a partir de uma análise qualitativa de um *corpus* naturalístico de 147 cancelamentos de compromissos, realizados em conversas de amigos por WhatsApp. Nomeadamente, 48 cancelamentos foram realizados por brasileiros (em particular, do Estado de São Paulo), 43 por portugueses (das regiões de Lisboa e Porto) e 56 por italianos (da cidade de Roma).

O *corpus* foi recolhido por meio de capturas de ecrã enviadas pelos próprios participantes entre dezembro de 2023 e junho de 2024. Os informantes que participaram neste estudo totalizam 128 indivíduos com idades entre os 20 e os 35 anos, amigos ou conhecidos da própria investigadora, que concordaram previamente em partilhar as suas conversas (chats) por WhatsApp e autorizaram o uso dos seus dados para fins de investigação e de divulgação científica. A maioria é composta por estudantes universitários e trabalhadores.

Posteriormente, para compreender as perceções dos falantes sobre o ato de fala objeto da investigação, foram recolhidas reflexões metapragmáticas por meio de entrevistas, tendo sido entrevistados 76 informantes para o estudo dos conhecimentos socioculturais dos falantes da língua e cultura-alvo (Félix-Brasdefer & Hasler-Barker, 2017).

Em conformidade com o Regulamento Geral sobre a Proteção de Dados (RGPD), todos os dados foram anonimizados ou pseudonimizados. As conversas recolhidas foram numeradas e transcritas num documento Microsoft Word e, posteriormente, inseridas numa tabela Excel.

As interações no WhatsApp constituem um *corpus* privilegiado para a análise sociopragmática do comportamento linguístico *online*, porque permitem investigar os atos de fala em conversas quotidianas em ambientes virtuais com base em dados naturais (Félix-Brasdefer, 2007b; Flores-Salgado & Castineira-Benitez, 2018; Golato, 2017). Para os efeitos do presente trabalho, foram assim considerados instrumentos e critérios de análise pertencentes à área da CMC (cf. Androutsopoulos, 2006; Crystal, 2001; Herring, 2004), da sociopragmática e da pragmática cross-cultural (cf. Blum-Kulka et al., 1989; Félix-Brasdefer, 2003; Flores-Salgado & Castineira-Benitez, 2018; Nuzzo & Cortés Velásquez, 2020).

4.3. **Categorias de análise**

A taxonomia utilizada para a análise do ato de fala de desmarcar compromissos é inspirada nos trabalhos de Nuzzo e Cortés Velásquez (2020), Cortés Velásquez e Nuzzo (2021) e Brocca et al. (2023). Os autores dividem cada ato em unidades ilocucionárias mínimas definidas como 'sub-atos'. Os sub-atos encontrados no *corpus* foram categorizados de acordo com o seu valor ilocucionário específico. Esta perspetiva, diferentemente da classificação adotada pela maioria das investigações no âmbito da pragmática cross-cultural (cf. Beebe et al., 1990; Blum-Kulka et al., 1989; Félix-Brasdefer, 2003; Santoro et al., 2021), em vez de identificar um ato principal (núcleo) e os atos de suporte, classifica os diferentes sub-atos sem atribuição hierárquica. Tal abordagem, segundo os autores, simplifica o processo de etiquetagem uma vez que nem sempre é possível distinguir o ato principal das estratégias acessórias, especialmente quando o ato de desmarcar ocorre de forma indireta (por exemplo, apenas por meio de uma justificação). Por conseguinte, todos os sub-atos podem igualmente contribuir para transmitir a principal força ilocutória do ato de fala.

Durante a análise dos dados, essa taxonomia foi ajustada aos nossos *corpora* através de uma abordagem orientada pelos dados ("*data-driven*"), garantindo que a classificação final contemplasse exclusivamente as estratégias discursivas efetivamente atestadas nos dados.

Portanto, a partir da observação do *corpus* e da revisão da literatura de referência, conseguimos assim identificar 11 categorias de sub-atos[6]:

[6] Para evitar potenciais confusões ou dúvidas, as categorias de análise mantém a nomenclatura em inglês, original nos trabalhos de Nuzzo e Cortés Velásquez (2020), Cortés Velásquez e Nuzzo (2021) e Brocca et al. (2023).

1. *Alerter*: o locutor chama a atenção do interlocutor para começar a interação (a conversa pode ser aberta com saudações afetuosas);
2. *Appeal to empaty*: o locutor apela à empatia do interlocutor para que tenha uma atitude compreensiva;
3. *Cancellation*: o locutor comunica o cancelamento, ou seja, que não poderá comparecer ao evento para o qual tinha anteriormente aceite o convite;
4. *Explanation*: o locutor explica por que razão vai cancelar, referindo-se a um impedimento específico ou a um imprevisto genérico;
5. *Farewell*: o locutor despede-se para encerrar a conversa;
6. *Gratitude*: agradece o convite, manifestando o seu apreço pela proposta recebida;
7. *Offer of repair*: propõe uma alternativa ou promete não faltar da próxima vez, ou, no caso de eventos em grupo, pede para não ser desmarcado o compromisso, apesar da sua ausência;
8. *Preparator*: prepara o terreno dizendo que tem notícias para dar;
9. *Remedial move*: expressa descontentamento e/ou pede desculpa;
10. *Willingness*: expressa o seu desejo de participar apesar do cancelamento;
11. *Wishes:* deseja que o interlocutor se divirta e expressa o seu voto de sucesso para o evento.

Para ilustrarmos as categorias de análise, na Tabela 1 utilizámos alguns exemplos previamente retirados dos três *corpora*, italiano (ITA), português brasileiro (PB) e português europeu (PE).

Tabela 1

Taxonomia dos Sub-atos

SUB-ATO	**ITA**	**PB**	**PE**
Alerter	*Buongiorno!; Raga*	*Oi, meninas!; Mano*; *Gente*	*Amiga linda!!; Alô coração!*; *Amigas*
Appeal to empathy	*Non mi uccidete per favore*	*Não me odeie*; *Eu sou a pior pessoa do mundo*	
Cancellation	*Io stasera non riesco a venire*	*Não vou mais conseguir ir* 😳	*Tenho que cancelar amanhã* 🤦‍♀️
Explanation	*Non mi sento tanto bene, ho il raffreddore*	*Eu confesso que tô muito cansada* 🤩	*Estou com covid* 😭 😭
Farewell	*Ci sentiamo presto*	*Beijos!; Abraço*	*Um beijinho! Fim de semana feliz!*
Gratitude	*Grazie mille per l'invito e auguri!!* 🤗 🥳	*Agradeço de coração pelo convite*	
Offer of repair	*Rimandiamo, ok?*	*Acha que podemos jogar pra quinta ou sexta?*	*Havemos de combinar para uma altura mais tranquila*
Preparator	*Ho un problema*; *Cattive notizie* 😳		*Péssimas notícias…*
Remedial move	*Mi dispiace tantissimo*	*Me perdoe* 😭	*Mil perdões pelo trastorno!*
Willingness		*Queria ir hoje… mas…*	*Estava tanto a apetecer me ir*
Wishes	*Divertiti!*	*Bom almoço pra vcs!*	*Mas divirtam-me muito!!!*

Nem todos os 11 sub-atos foram identificados nos três *corpora*; no entanto, optou-se por manter todos os que foram registados em cada *sub-corpus*, como *Appeal to Empathy*, *Gratitude*, *Preparator* e *Willingness*.

Por último, três categorias de modificadores internos foram identificadas:

1. *Evaluation*: introduz uma avaliação do locutor sobre o estado de coisas descrito no sub-ato em que ocorre;
2. *Intensifier*: aumenta a força ilocutória do sub-ato em que ocorre;

3. *Term of endearment*: termo de afeição que qualifica a relação entre o locutor e o interlocutor.

Trata-se de modificadores lexicais e morfossintáticos utilizados pelos falantes para atenuar ou intensificar a força ilocutória do ato do cancelamento, como exemplificado na Tabela 2 (os modificadores estão a negrito). Emoticons e emojis também foram anotados, e a sua função foi determinada dependendo do contexto, principalmente como modificadores proxémicos (Brocca et al., 2023).

Tabela 2

Taxonomia dos Modificadores

MODIFICADOR	**ITA**	**PB**	**PE**
Evaluation	***Purtroppo*** *mi è salita la febbre;* ***Mi sa che*** *vi devo dare buca oggi*	*não vou conseguir ir ver a tua apresentação* ***infelizmente****!* 😞; ***Acho que*** *não vou* 😳	*Olha* ***infelizmente*** *não consigo ir ao jantar* 😕*;* ***Creio que*** *já não vou conseguir ir*
Intensifier	*Mi dispiace* ***proprio****;* *mi dispiace* ***tantissimo****…*	***Mil*** *desculpas;*	*Eu queria* ***bue*** *mas não consigo* 😭 😭 😭*;* ***Mil*** *perdões pelo trastorno!*
Term of endearment	***ciccine*** *ciao*❤️ *Ciao* ***carissima***	*Po* ***irmão****;* *Bom dia* ***amiga****, tudo bem?; Nossa,* ***gato****!*	***Amigas*** 😭*;* ***Meus amores****;* *Alô* ***coração****!*

5. Análise das estratégias discursivas

Nesta secção, após a anotação manual dos *corpora*, foi possível realizar uma análise ilustrativa das estratégias pragmalinguísticas utilizadas para a realização do ato de desmarcar compromissos a partir de um estudo de caso. Em particular, foram extraídos exemplos dos três *sub-corpora* (italiano, português brasileiro e português europeu) sobre cancelamentos de compromissos entre amigos (por exemplo passeios, jantares, festas, etc.), enviados por IM. Os resultados são apresentados com base tanto nas estratégias empregadas nos sub-atos quanto nos modificadores. Nomeadamente, são analisados o conteúdo e a ordem das estratégias pragmáticas. De referir, no entanto, que não foi possível estabelecer a frequência de uso dessas estratégias, devido ao facto de ser um estudo em desenvolvimento e de a dimensão da amostra não permitir fazer uma análise quantitativa. Futuramente, esperamos aprofundar a investigação, apontando nesta direção.

5.1. **O ato de fala de desmarcar compromissos entre amigos no *instant messaging***

5.1.1. **Os sub-atos que compõem o ato de desmarcar**

Os exemplos[7] aqui apresentados consideram interações caracterizadas por um alto nível de informalidade. A este respeito, os dados naturais recolhidos até ao momento foram enviados/recebidos através da aplicação de mensagens instantâneas WhatsApp, que, devido à sua natureza, é habitualmente utilizada em contextos informais e/ou semi-informais, exigindo, portanto, um grau médio-baixo de distância social e de poder relativo (na terminologia de Brown & Levinson, 1987). Além disso, as interações espontâneas, obtidas nas línguas e/ou variedades objeto do estudo, referem-se principalmente a atividades de convívio entre amigos e colegas (de trabalho ou da faculdade). Isto implica uma maior proximidade entre os interlocutores e, por conseguinte, o uso de uma linguagem coloquial, com expressões informais, e de elementos típicos da oralidade.

Considerando os pressupostos teórico-metodológicos delineados nos parágrafos anteriores, foi possível identificar, nos três *sub-corpora*, dois modelos principais para a realização do ato de fala de desmarcar compromissos: em diversos casos, a configuração do ato do cancelamento apresenta uma estrutura centralizada no sub-ato *Cancellation*, ou seja, uma asserção explícita de não participação, em combinação com outros sub-atos; no entanto, em outros casos, o ato de desmarcar é expresso por meio de sub-atos distintos de *Cancellation*, implicando, assim, o recurso a estratégias comunicativas indiretas.

Quando o ato do cancelamento é constituído pela combinação do sub-ato *Cancellation* com outros sub-atos, tal configuração manifesta-se principalmente através do sub-ato *Cancellation*, em coocorrência com os sub-atos *Alerter* e *Explanation*, como é possível observar nos três exemplos que seguem (1) – (3), em italiano, português brasileiro e português europeu, respetivamente.

(1) ITA_49:
Raga [*Alerter*] Molto probabilmente una mia coinquilina ha il Covid [*Explanation*] Quindi non sarò dei vostri **temo** [*Cancellation*]
*Malta Muito provavelmente uma das minhas colegas de casa tem Covid Por isso **receio** que não possa juntar-me a vocês*

(2) PB_08:
Oi, meninas! [*Alerter*] Acordei hoje com muita dor de garganta [*Explanation*], **acho que** não será uma boa sair hoje 😩 [*Cancellation*]

(3) PE_43:
Meus amores [*Alerter*], surgiu-me uma coisa [*Explanation*] e **creio que** já não vou conseguir ir [*Cancellation*], mas divirtam-me muito!!! 🥹 🫶🏻 ✨ [*Wishes*]

Em relação às estratégias discursivas, os três falantes optam por fazer uma referência explícita ao facto de estarem a cancelar e de não poderem ir ao evento. No entanto, ao desmarcarem, incluem construções com verbos de opinião e sentimento, *acho que*, *creio que*, em português brasileiro e europeu, e *temo* ('receio') em italiano, para atenuar a força ilocutória do enunciado. Desta forma, espera-se que o interlocutor interprete a mensagem como rejeição, mesmo não sendo algo definitivo. Este fenómeno, que foi encontrado nos três *sub-corpora*, é

[7] Para a apresentação dos exemplos, foram omitidos os nomes dos participantes e foram aplicados códigos identificativos para cada ato de fala produzido. No caso de mensagens em que os falantes mencionaram nomes próprios (por exemplo, nas expressões vocativas), foram mantidas apenas as iniciais. Os exemplos retirados do *corpus* são apresentados exatamente como foram escritos e enviados pelos informantes via WhatsApp. Portanto, nenhum erro gramatical, de pontuação ou sintático, assim como abreviação, foi corrigido durante o processo de transcrição e tradução dos enunciados. Os textos das mensagens originais em italiano foram traduzidos para o português pela investigadora com a máxima fidelidade possível.

corroborado pelos resultados evidenciados em outras investigações sobre estratégias de cortesia e atenuação[8], tanto no contexto italiano como no contexto do português (cf. Carreira, 2004; Cortés Velásquez & Nuzzo, 2021; González Berrio et al., 2021; Seara, 2017). Na mesma linha, foi identificada a presença do verbo "conseguir" nos *corpora* de informantes brasileiros e portugueses para atenuar o ato de cancelar (*não vou conseguir ir*), exemplos (3), (11) e (13); e no corpus de italiano foi utilizada uma expressão equivalente por meio do verbo pronominal "farcela" (*non ce la faccio*), veja-se o exemplo (8).

Nos exemplos (2) e (3), em PB e PE, também foram utilizados emojis com função de modificadores internos nos sub-atos *Cancellation* e *Wishes* para expressar tristeza e frustração como estratégia de aproximação ao alocutário.

Os participantes consideraram igualmente necessário fornecer uma explicação ao cancelar em cima da hora, referindo-se a um impedimento específico, em (1) e (2), ou a um imprevisto genérico, em (3). Ao atribuírem o cancelamento a um elemento independente da sua vontade que impede, justamente, a participação no evento, os locutores podem salvar tanto a face positiva como a negativa dos seus alocutários. De facto, como já foi observado em estudos anteriores que investigaram este ato de fala em italiano, espanhol colombiano e alemão austríaco (Brocca et al., 2023; Nuzzo & Cortés Velásquez, 2020), ao apresentar motivos imperiosos para desmarcar, o falante pode dar a entender que, normalmente, não ia ameaçar a face positiva do amigo que o convidou, ao não cumprir o seu desejo de partilhar tempo juntos; e, por outro lado, também não ia ameaçar a face negativa ao interferir na execução dos seus planos.

Por fim, em (3) o falante nativo do PE, ao contrário dos falantes italiano e brasileiro, abre a interação com uma saudação afetuosa "meus amores", interpretada como uma estratégia de intensificação[9] que reforça a cortesia positiva (ou seja, a valorização da imagem do alocutário), criando empatia com o interlocutor. De facto, em Macário (2019, pp. 231–232), refere-se que os possessivos seguidos por nomes ou adjetivos (nominalizados), denotam tipicamente, em contexto de vocativo, uma proximidade afetiva entre os interlocutores. Segundo a autora, estes elementos são marcadores de dêixis emocional, entendida como expressão, por parte do falante, de proximidade ou distância afetiva relativamente à entidade referida. No entanto, o mesmo falante em (3), prefere usar uma justificação vaga em que se limita a mencionar um acontecimento imprevisto, ao passo que os outros optam por adicionar detalhes; e conclui a mensagem desejando que os interlocutores se divirtam.

No que diz respeito ao segundo esquema de realização do ato de desmarcar encontrado no *corpus*, o falante não diz expressamente que não estará presente no evento para o qual aceitou o convite. Neste caso, a força ilocutória do cancelamento é expressa através da combinação de diferentes sub-atos, que não inclui o sub-ato de *Cancellation*. O ato de desmarcar ocorre, assim, de forma indireta, por exemplo por meio de uma justificação junto à necessidade de remarcar, como em (4). Explicar apenas o motivo pelo qual alguém não pode comparecer a um evento pode ser entendido pelo interlocutor como um cancelamento implícito (Brocca et al., 2023). Vejam-se alguns exemplos:

[8] Sobre o conceito de atenuação, Briz (2006, p. 228) afirma: "Se trata de un mecanismo más complejo, incluso aparentemente contradictorio, en tanto me alejo del mensaje, suavizándolo, evitando algo de lo que digo o hago, para acercarme o no alejarme demasiado del otro y de ese modo evitar posibles tensiones, malentendidos, amenazas a la imagen propia y, sobre todo, ajena. Una estrategia, así pues, de distancia lingüística a la vez que una estrategia de acercamiento social".

[9] A intensificação, como sublinha Briz (Briz, 2017, p. 37), é: "una estrategia pragmática que colabora en el ámbito retórico-argumentativo como refuerzo de lo dicho o del punto de vista propio o ajeno, ya sea coincidente o contrario, y en el ámbito social como mecanismo de refuerzo, ya sea valorizador o amenazante de la imagen propia o ajena".

(4) ITA_41:
Raga [*Alerter*] io so mezza influenzata 😳 [*Explanation*] Mi sa che dobbiamo rimandare [*Offer of repair*]
Malta *eu estou meio doente* *Acho que temos de remarcar*

(5) PB_48:
Gente, [*Alerter*] desculpa [*Remedial move*], surgiu um compromisso para mim também… vou ter que acompanhar minha mãe em uma consulta… [*Explanation*] Tentem marcar vocês. [*Offer of repair*] Beijos! [*Farewell*]

(6) PE_25:
Esqueci me que o meu pai faz anos hj haha [*Explanation*]

Nestes exemplos, o cancelamento é comunicado por meio do uso de outros sub-atos, entre os quais se destaca a *Explanation*, que pode ser utilizada isoladamente, como no exemplo (6), ou em combinação com outros sub-atos, como em (4) e (5). Em relação a estes últimos dois casos, podemos observar que em (4) o esquema adotado pelo falante italiano para a realização deste ato foi: *Alerter* + *Explanation* + *Offer of repair*. Primeiro, o falante abre a interação utilizando a expressão vocativa *raga* ("malta", "pessoal") para chamar a atenção dos interlocutores; depois disso, explica que está doente e, por fim, sugere remarcar o jantar. No exemplo em PB, foi utilizada uma estrutura mais elaborada em torno ao sub-ato da *Explanation*: *Alerter* + *Remedial move* + *Explanation* + *Offer of repair* + *Farewell*. Também neste caso, o falante chama a atenção do interlocutor por começar a interação com o uso de um vocativo, *gente*; em seguida, pede desculpa e explica por que motivo não pode participar, referindo-se a um impedimento específico; depois, contrariamente ao exemplo em italiano, o falante não propõe uma alternativa, mas pede aos interlocutores para marcarem, apesar da sua ausência (*tentem marcar vocês*); e, por último, despede-se (*Beijos!*).

Em termos de cortesia linguística, importa salientar que, no exemplo (5), o falante brasileiro, ao cancelar em cima da hora, recorre a um pedido de desculpa (*Remedial move*), com o intuito de atenuar um ato potencialmente ameaçador e de apelar à compreensão do interlocutor. Ao pedir desculpa por interferir nos planos do interlocutor, o convidado manifesta preocupação com a sua face negativa. Com efeito, neste sentido, o falante sugere então que os elementos do grupo mesmo assim se encontrem, com o objetivo de não comprometer os seus planos.

Resumindo, nos exemplos (4) a (6), os informantes preferem não mencionar explicitamente a impossibilidade de comparecer ao evento, mas empregam outras estratégias para o interlocutor perceber que a pessoa convidada não poderá estar presente.

5.1.2. **Os modificadores internos**

Como foi referido nos parágrafos anteriores, o ato de fala de desmarcar compromissos pode ser modificado por elementos lexicais ou gramaticais para atenuar a sua força ilocutória. Estes elementos de modificação, definidos como modificadores internos, são opcionais e podem ser incorporados à estrutura do enunciado. Os modificadores encontrados no *corpus* foram categorizados a partir da taxonomia elaborada por Nuzzo & Cortés Velásquez (2020): *Evaluation*, *Intensifier* e *Term of endearment*. Dentro dos três *sub-corpora*, italiano, português brasileiro e europeu, foram identificados e analisados modificadores pertencentes a estas três categorias. Começando pelo *sub-corpus* do italiano, observamos o uso de modificadores por parte dos falantes, em particular para os sub-atos *Alerter*, *Cancellation*, *Explanation* e *Remedial move*. Vejam-se os exemplos seguintes:

(7) ITA_07:
Ciao, **purtroppo** mi è salita la febbre e domani non potrò **proprio** esserci
Olá, ***infelizmente*** *a minha febre subiu e amanhã não vou poder* ***mesmo*** *estar aí*

(8) ITA_31:
Mi sa che domani non ce la faccio perché oggi non sono riuscita ad aprire libro
Mi dispiace **tantissimo** darti buca 😭
Acho que *amanhã não vou conseguir, porque não consegui pegar num único livro*
Lamento ***imenso*** *por deixar-te pendurado/a* 😭

(9) ITA_56:
ciccine ciao ❤️ scusateci ma oggi non ce la sentiamo di fare niente stanotte siamo stati male forse ci ha dato fastidio la cena ci sentiamo un po' degli straccetti 😅😂🥲🥲🥲
Queridinhas*, olá* ❤️ *desculpem-nos, mas hoje não estamos com disposição para fazer nada esta noite passámos mal talvez o jantar não nos tenha caído bem estamos uns farrapos* 😅😂🥲🥲🥲

Conforme ilustrado a negrito nos exemplos (7) e (8), os sub-atos produzidos foram objeto de atenuação ou intensificação através do emprego de modificadores que se enquadram nas categorias de *Evaluation* e *Intensifier*. Para a primeira categoria (*Evaluation*), foram utilizados modificadores lexicais para expressar tristeza (*purtroppo mi è salita la febbre*, "infelizmente a minha febre subiu"), no sub-ato *Explanation* em (7), ou dúvidas em relação à não participação no evento (*mi sa che domani non ce la faccio*, "acho que amanhã não consigo"), no sub-ato *Cancellation* em (8), com o intuito de atenuar a força ilocutória do ato do cancelamento. No caso da segunda categoria (*Intensifier*), foram utilizados não só modificadores lexicais para transmitir com mais intensidade os sentimentos do falante ao comunicar que não estará presente (*domani non potrò proprio esserci,* "amanhã não vou poder mesmo estar aí") em (7); como também modificadores morfossintáticos (*mi dispiace tantissimo darti buca* 😭, "lamento imenso deixar-te pendurado/a"), com o emprego do grau superlativo absoluto através do sufixo "-issimo" no adjetivo "tanto" em (8). Neste último exemplo, o locutor recorreu a um emoji de choro que contribui, por sua vez, para reforçar o ato ilocutório. No exemplo (9), no sub-ato *Alerter*, foi usada a expressão vocativa *ciccine ciao* ("queridinhas, olá") acompanhada do emoji do coração, interpretada como *Term of endearment*. Nomeadamente, ao usar um adjetivo de avaliação positiva, o locutor mostra preocupar-se com a proteção da face positiva do alocutário. De facto, neste caso, há uma tentativa de valorizar a relação que existe entre ele e o seu destinatário e apelar à sua solidariedade. Este fenómeno foi encontrado igualmente no *sub-corpus* de português europeu, no exemplo (3), e no *sub-corpus* de português brasileiro, que passamos a apresentar por meio de alguns exemplos:

(10) PB_23:
Pessoal, **infelizmente**, amanhã tenho uma banca de qualificação

(11) PB_45:
Poo, n vou conseguir ir com vcs
Desculpa **msm**
Queria

(12) PB_27:
Po **irmão**, fiquei com um contratempo pra resolver

Nestes três exemplos em PB, encontrámos semelhanças com os exemplos em italiano no uso dos modificadores internos: em (10) o falante brasileiro também utilizou um modificador lexical pertencente à categoria *Evaluation*, para expressar descontentamento no sub-ato *Explanation* (*infelizmente*). Em (11) aparece o modificador lexical da categoria *Intensifier* para intensificar a força do sub-ato *Remedial move* (*desculpa msm*). É interessante notar que, neste caso, o modificador não é usado apenas para manifestar descontentamento ou tristeza, mas também para reforçar um pedido de desculpa. No exemplo (12), o vocativo *irmão*, identificado como *Term of endearment*, sinaliza uma proximidade afetiva e, em particular, uma relação de amizade entre

dois jovens. Esta forma de tratamento pode ser considerada uma estratégia de cortesia positiva (ou seja, de solidariedade), pois, como já foi evidenciado em outros casos, trata-se de termos afetuosos utilizados principalmente no sub-ato *Alerter* para abrir a interação e predispor positivamente o interlocutor.

No que diz respeito aos modificadores encontrados no *sub-corpus* do português europeu, estes foram usados como estratégia de atenuação ou intensificação em particular nos sub-atos de *Alerter, Cancellation*, *Remedial move*, *Explanation*. Nos exemplos (13) e (15), novamente, foram usados modificadores da categoria *Term of endearment* (vocativos afetuosos e emojis) como estratégias de cortesia valorizadora para chamar a atenção do interlocutor e apelar à sua solidariedade, *amiga linda* e *bom dia* 😘. Cabe destacar que, no exemplo (13), também é usado um *Term of endearment* no sub-ato de *Cancellation* para atenuar a força ilocutória do cancelamento: *meu amor*, acompanhado do emoticon de choro :"(. Adicionalmente, os sub-atos *Cancellation* e *Explanation* são frequentemente objeto de modificação, através do emprego de modificadores pertencentes às categorias de *Evaluation* (*não sei se…*; *infelizmente*) e *Intensifier* (*mesmo*), como demonstrado nos exemplos (14) e (15). No caso do sub-ato *Remedial move*, pelo contrário, este parece incorporar apenas modificadores da categoria *Intensifier* (*mil perdões*), veja-se o exemplo (13). Tal fenómeno aparece igualmente nos dados sobre o italiano e o português brasileiro e encontra confirmação também na literatura de referência.

(13) PE_42:
Amiga linda!! Como estás, coisa boa? 🥹 💗
Afinal, não vou conseguir ir ao filminho amanhã, **meu amor :"(Mil** perdões!

(14) PE_31:
M., eu **não sei se** é boa ideia eu ir amanhã, eu estou **mesmo** a falecer

(15) PE_24:
Bom dia 😘 tudo bem?
Olha **infelizmente** não consigo ir ao jantar 😕 vou fazer noite nesse dia, só vou sair depois das 23h.
Espero que se divirtam muito 🫶🏻

Tendo em conta que foi realizada uma análise à linguagem escrita em ambientes virtuais, vale ressaltar que, além da combinação de diferentes sub-atos e modificadores lexicais ou gramaticais, existem outros recursos que contribuem para a atenuação da força ilocutória do ato de fala produzido. Isto é, elementos conversacionais típicos da CMC, como emoticons, onomatopeias de gargalhada e exclamações, que permitem expressar sentimentos e ações difíceis de transmitir através da escrita. A título de exemplo, o emprego da onomatopeia do riso (*haha*) pode constituir um recurso que auxilia os participantes na gestão de momentos interacionais particularmente delicados (Petitjean & Morel, 2017), como em (6). Na mesma linha, encontra-se o ponto de exclamação, que, segundo Baron e Ling (2011), pode funcionar como um *courtesy symbol* e *conversational softener*, sendo, portanto, utilizado para tornar o tom da mensagem mais amigável e sincero. Este sinal de pontuação, bem como o sinal de reticências, parece então assumir um papel neutro ou mesmo "positivo" na marcação do final de uma unidade textual, funcionando como alternativa ao ponto final, que, nesses mesmos contextos, tende a adquirir conotações negativas (Longo, 2020, p. 48), como se pode observar nos exemplos (5) e (13). Da mesma forma, emojis e emoticons representam estratégias da escrita digital destinadas a recuperar os aspetos não verbais da comunicação face a face, nomeadamente a expressão facial e a gestualidade (Antonelli, 2011), e também ajudam a atenuar a força ilocutória e, consequentemente, a potencial "ameaça" do ato que está a ser realizado, como em (2), (3), (4), (8), (9), (13) e (15).

5.2. **Comentários dos participantes a partir de uma entrevista**

Nesta secção, serão apresentados comentários dos participantes recolhidos através de entrevistas *online* com o intuito de obter dados sobre as suas perceções em relação ao ato de fala objeto deste estudo. As entrevistas

foram realizadas na língua materna dos participantes para possibilitar uma expressão livre e espontânea, prevenindo assim eventuais obstáculos à comunicação.

O acesso aos conhecimentos e às interpretações dos informantes representa um recurso essencial ao investigador para se orientar no processo de análise dos dados empíricos e das práticas sociolinguísticas da comunidade que quer estudar. De facto, o uso de instrumentos introspetivos ou retrospetivos na pragmática cross-cultural constitui uma metodologia válida para o estudo dos conhecimentos socioculturais dos falantes da língua e cultura-alvo e da perceção de determinados atos de fala em contextos formais e informais, bem como das noções de cortesia, de diretividade e indiretividade (Félix-Brasdefer & Hasler-Barker, 2017, p. 33). Portanto, ainda que esta constitua uma amostra reduzida (76 entrevistados: 30 brasileiros, 21 portugueses e 25 italianos) os informantes forneceram pistas acerca das suas expectativas ao receberem um cancelamento. Mais precisamente, responderam às seguintes perguntas:

i. Como consideras o ato de desmarcar um compromisso, em particular em cima da hora?

ii. Normalmente, qual é a tua reação ao receber um cancelamento por parte de um amigo?

iii. É algo que faz parte do teu quotidiano e da tua língua e cultura?

A partir destas perguntas surgiram diversas considerações nas três comunidades de fala em estudo; seguem-se alguns exemplos[10]:

(16) ITA_23 [F-27-Roma]:
Se per un imprevisto vero e proprio, può capitare a tutti. Se è un'abitudine, è poco carino ed educato. Se so già che non potrò andare rifiuto subito, però, se mi capita di cancellare, avviso appena so.
Se for por causa de um imprevisto real, pode acontecer a qualquer um. Se é um hábito, não é muito simpático e educado. Se eu já sei que não vou poder ir recuso-me imediatamente. Mas se tiver de desmarcar algo aviso assim que souber.

(17) ITA_25 [F-26-Roma]:
Tendenzialmente, quando mi si propone una cosa e già non mi va, dico subito no, cioè dico subito "no, ho da fare". Se annullo il giorno stesso o un'ora prima, vuol dire che è veramente successo qualcosa (che sto male, che ho un imprevisto, comunque può succedere). Se è il giorno stesso, non annullo mai solo perché non mi va.
Normalmente, quando recebo um convite e não tenho vontade de ir, digo logo que não, ou seja, digo imediatamente "não, estou ocupado". Se eu cancelar no mesmo dia ou uma hora antes, significa que algo realmente aconteceu (que estou doente, que tenho um imprevisto, etc.). Se for no mesmo dia, nunca cancelo só porque não me apetece.

(18) PB_13 [F-26-Interior São Paulo]:
No geral, a gente não tem muito a cultura de falar não. É mais educado cancelar e dar um motivo / desculpa do que falar não. E o que eu faço às vezes é esperar que todo mundo esqueça. Então não confirmo porque aí as pessoas vão lembrar.

(19) PB_29 [M-22-São Paulo]:
Normalmente deixamos em aberto a proposta e se não for possível ir recusamos em cima da hora. Você pode não ter certeza do que vai fazer naquele dia então aceita e chega no dia fica com preguiça

[10] Para garantir o anonimato dos entrevistados, colocámos códigos de identificação, compostos por um número sequencial, sexo, idade e cidade de proveniência do participante.

de ir e arruma desculpa, ou acaba tendo que fazer outra coisa de fato. Mas acontece bastante sim

(20) PE_12 [F-23-Porto]:
Não tenho muitos exemplos de pessoal a cortar-se à última da hora, só a dizerem me que não podem de imediato.
Eu prefiro que me digam não logo de início do que permitirem que eu marque e me oriente, para depois dizerem que não. Desmarcar em cima da hora, a menos que seja por um motivo muito válido, eu encaro como falta de respeito e apreciação.

(21) PE_17 [M-24-Porto]:
Isso depende muito da pessoa. Mas considero que o ideal é a pessoa dizer logo ou que não está interessada ou que não pode, ou então avisar com antecedência que não vai poder. Eu das pessoas que conheço a tendência acho que é ou não aceitar de imediato com uma justificação (ou aceitar e ir). Cancelar no próprio dia é mais raro.

Por meio desta metodologia, foi possível observar não apenas a forma como os falantes consideram o ato de fala de desmarcar compromissos de última hora, mas também a sua perceção quanto à possibilidade e necessidade da sua realização. Estes aspetos fornecem dados relevantes para a análise metapragmática das três comunidades de fala em estudo.

Sucintamente, nestes comentários, os italianos e os portugueses enfatizam a importância de respeitar os compromissos assumidos e, segundo eles, devem ser cancelados apenas em casos de emergência; sentem portanto que tal ato implica um grau elevado de ameaça para a face negativa; ao passo que os brasileiros, mesmo tendo uma visão negativa semelhante dos cancelamentos de última hora, referem que, no contexto linguístico-cultural brasileiro, há uma preferência por aceitar um convite quando o recebem, para não frustrar as expectativas de quem fez o convite e, possivelmente, tendem a cancelá-lo posteriormente.

Importa sublinhar que a amostra selecionada não é representativa das populações inteiras das três comunidades, sendo que as considerações metapragmáticas que surgiram durante as entrevistas revelam a necessidade de expandir o *corpus* para o aprofundamento da investigação.

6. Considerações finais e perspetivas futuras

Neste estudo exploratório, procurámos investigar alguns aspetos sobre as estratégias de realização do ato de fala de desmarcar compromissos, em particular de última hora, por parte de três diferentes comunidades de falantes (italianos, portugueses e brasileiros). Pretendemos, assim, categorizar e analisar uma amostra de 147 cancelamentos de compromissos enviados por aplicações de mensagens pelos informantes em contextos reais e depois partilhados com a própria investigadora. Cada ato de fala foi segmentado em diferentes sub-atos a partir da taxonomia elaborada por Nuzzo e Cortés Velásquez (2020) e, durante o processo de análise dos dados, foi adaptada aos nossos *corpora* por meio de procedimentos *data-driven*, de modo que a classificação final incluísse exclusivamente as estratégias discursivas atestadas nos dados. Esta análise foi completada com dados introspetivos de tipo metapragmático obtidos por meio de entrevistas, que nos permitiram cruzar dados empíricos sobre a produção e a perceção dos falantes em relação ao ato de fala objeto do estudo.

Para exemplificarmos os fenómenos encontrados no *corpus*, optamos por apresentar um estudo de caso sobre cancelamentos de compromissos em contextos informais entre amigos e, ao observar as realizações e as perspetivas dos atos produzidos pelos falantes das três línguas e/ou variedades escolhidas, evidenciaram-se várias convergências, mas também divergências que cabe ressaltar. Por um lado, a análise pragmalinguística e sociopragmática dos dados recolhidos permitiu identificar, nos três grupos de falantes, dois esquemas principais para a realização do ato de desmarcar: (1) uma configuração centrada no sub-ato *Cancellation*, isto é, uma asserção explícita de não participação, combinada com outros sub-atos; (2) uma estrutura composta por sub-

atos distintos de *Cancellation*, implicando, assim, o recurso a estratégias comunicativas indiretas (por exemplo, apenas por meio de fórmulas reparadoras como pedidos de desculpa e justificações, ou propostas de remarcação). Além disto, em termos de cortesia linguística, os resultados apontam uma semelhança também no que concerne à proteção da face do interlocutor no momento de cancelar um compromisso em cima da hora, porque os informantes adotaram tanto estratégias de cortesia positiva (saudações afetuosas, expressões de descontentamento, tentativas de remarcar) como de cortesia negativa (pedidos de desculpa, justificações). É importante realçar que os locutores tendem também a atribuir a sua ausência a uma circunstância imprevista e inevitável, contra a sua vontade, e isto pode salvar tanto a face positiva como a face negativa. De facto, ao justificar o cancelamento com motivos imperiosos, o falante mostra que normalmente não comprometeria os planos do quem o convidou (face negativa), como também não frustraria as suas expectativas (face positiva).

Por outro lado, detetamos uma diferença no que diz respeito às perceções dos falantes sobre a possibilidade de realização deste ato de fala nos contextos linguístico-culturais em que se encontram. Em fase de entrevista, os falantes portugueses e italianos manifestaram a necessidade de ver respeitada a própria face negativa e a do interlocutor (ou seja, evitar a invasão do território: a recusa imediata em resposta a um convite parece ser mais aceitável e, inversamente, o cancelamento de última hora seria uma opção menos preferível). Pelo contrário, os falantes brasileiros, mesmo tendo uma opinião negativa sobre este ato, consideram-no um fenómeno bastante generalizado na própria cultura, em que se tende a não frustrar as expectativas alheias aquando da receção de um convite.

Face à carência de investigações empíricas quantitativas de maior robustez, podemos apenas avançar a hipótese de que a língua e cultura brasileira apresentam uma orientação predominantemente positiva em termos de cortesia – isto é, privilegiam a valorização da imagem do alocutário e a aprovação dos seus desejos no momento em que faz uma proposta ou um convite, através da adoção de estratégias de solidariedade – enquanto a língua e cultura italiana e a portuguesa parecem ser mais orientadas para a cortesia negativa. Esta perspetiva, no caso do português brasileiro, está em consonância com o que foi constatado noutras culturas da América Latina (Cortés Velásquez, 2017; Nuzzo & Cortés Velásquez, 2020; Placencia & Garcia-Fernandez, 2007), e, relativamente ao italiano, é corroborada pelos resultados dos trabalhos de Brocca et al. (2023) e Nuzzo e Cortés Velásquez (2020).

Por último, é importante salientar que, neste trabalho, a recolha de dados foi baseada em material autêntico que é estruturalmente limitado e não perfeitamente equilibrado, sobretudo no que diz respeito à representação cultural e de género, fator esse que poderá ter influência nos resultados. Além disso, tratando-se de um estudo em desenvolvimento, estes resultados não são definitivos e não podem ser generalizados, mas são indicadores relevantes de práticas linguísticas que merecem ser investigadas numa escala maior. Assim, para a prossecução do estudo, consideramos necessária a ampliação do *corpus* de dados naturalísticos, "*naturally occurring data*" (Golato, 2017), ou seja, produzidos de forma espontânea. Este tipo de dados, contudo, apresenta limites: os *corpora* de dados autênticos podem não disponibilizar uma quantidade suficiente de dados para identificar tendências e ocorrências elevadas do fenómeno pragmático em análise, o que os torna pouco adequados para a realização de estudos de natureza comparativa (Kasper, 2008). Para superar essa limitação, consideramos essencial complementar o nosso *corpus* com dados elicitados por meio de instrumentos como o *Discourse Completion Task* (DCT) e questionários de perceção. Trata-se de metodologias amplamente utilizadas no âmbito da pragmática cross-cultural (Ogiermann, 2018), que podem proporcionar ao investigador diversas vantagens, entre as quais o controlo das variáveis sociolinguísticas, a recolha de grandes quantidades de dados e a obtenção de resultados comparáveis e replicáveis, aspeto que os distingue dos *corpora* de dados naturais (Félix-Brasdefer, 2007a). Como referido anteriormente, o DCT é um questionário baseado na descrição de situações comunicativas, no qual se solicita ao informante que redija a resposta que produziria caso se encontrasse nos contextos apresentados. Por conseguinte, optámos por criar um DCT *online*, a partir do *corpus* de dados espontâneos analisado até ao momento, que permitirá captar as perceções de um amplo conjunto de falantes sobre o comportamento linguístico considerado mais adequado para cancelar um compromisso, consoante as situações apresentadas. Além disso, este questionário será aplicado em ambientes virtuais,

possibilitando a recolha de dados relevantes para a análise da comunicação mediada por computador, e os resultados obtidos serão comparados com o *corpus* de dados espontâneos. Desta forma, será possível conduzir uma análise contrastiva mais aprofundada sobre o ato de fala de desmarcar compromissos, relacionando dados quantitativos e qualitativos sobre o uso das estratégias de cortesia linguística e a consciência metapragmática dos falantes nativos de italiano, português europeu e português brasileiro.

Referências

Amossy, R. (2010). *La présentation de soi. Ethos et identité verbale*. https://doi.org/10.3917/puf.amoss.2010.01

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, *10*(4), 419–438. https://doi.org/10.1111/j.1467-9841.2006.00286.x

Antonelli, G. (2011). Lingua. In A. Afribo & E. Zinato (Eds.), *Modernità italiana. Cultura, lingua e letteratura dagli anni Settanta a oggi* (pp. 15–52). Carocci. https://www.academia.edu/43974550/Lingua

Antonelli, G. (2014). L'e-taliano: Una nuova realtà tra le varietà linguistiche italiane? In E. Garavelli & E. Suomela-Härmä (Eds.), *Dal manoscritto al web: Canali e modalità di trasmissione dell'italiano* (Vol. 2, pp. 537–556). Franco Cesati Editore. https://iris.unicas.it/handle/11580/36582

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Baron, N. S., & Ling, R. (2011). Necessary smileys & useless periods: Redefining punctuation in electronically-mediated communication. *Visible Language*, *45*(1–2), 45–67.

Beebe, L., Takahashi, T., & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. Scarcella, E. S. Andersen & S. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 55–73). Newbury House Publishers.

Benine, A. B. G., & Santoro, E. (2022). O ato de fala do pedido em português brasileiro e italiano: Um estudo sobre a avaliação do contexto por parte dos falantes. *Revista de Italianística*, *44*, 105–134. https://doi.org/10.11606/issn.2238-8281.i44p105-134

Blum-Kulka, S., House, J., & Kasper, G. (1989). *Cross-cultural pragmatics: Requests and apologies*. Ablex Publishing.

Briz, A. (2006). Atenuación y cortesía verbal en la conversación coloquial: Su tratamiento en la clase de ELE. In *Actas del Programa de formación para profesorado de Español como Lengua Extranjera 2005-2006* (pp. 227–255). Instituto Cervantes de Munich. http://cvc.cervantes.es/ensenanza/biblioteca_ele/publicaciones_centros/PDF/munich_2005-2006/02_briz.pdf

Briz, A. (2017). Otra vez sobre las funciones de la intensificación en la conversación coloquial. *Boletín de Filología*, *52*(2), pp. 37–58. https://doi.org/10.4067/S0718-93032017000200037

Brocca, N. (2021). Ladder: Un corpus di scritture digitali per l'insegnamento della pragmatica in L2. Un esempio di analisi di disdette in Whatsapp. *Italiano LinguaDue*, *13*(1), 241–259. https://doi.org/10.13130/2037-3597/15867

Brocca, N., Nuzzo, E., Velásquez, D. C., & Rudigier, M. (2023). Linguistic politeness across Austria and Italy: Backing out of an invitation with an instant message. *Journal of Pragmatics*, *209*, 56–70. https://doi.org/10.1016/j.pragma.2023.02.018

Brown, P., & Levinson, S. C. (1978). Universals in language usage: Politeness phenomena. In E. N. Goody (Ed.), *Questions and politeness: Strategies in social interaction* (pp. 56–311). Cambridge University Press.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Cabral, A. L. T. (2014). Interações verbais em ambientes virtuais: Cortesia, descortesia e mal-entendido. In I. Seara (Ed.), *Cortesia: Olhares e (re) invenções* (Vol. 1, pp. 497–518). Chiado Editora.

Caffi, C. (2017). *Pragmatica. Sei lezioni*. Carocci.

Carreira, M. H. A. (2004). *Plus ou moins?! L'atténuation et l'intensification dans les langues romanes*. Université Paris 8 Vincennes Saint-Denis.
Cortés Velásquez, D. (2017). Cancelar una cita como estrategia de rechazo postergado: Resultados e implicaciones didácticas de un estudio transcultural. *EuroAmerican Journal of Applied Linguistics and Languages*, *4*(2), 115–134. https://doi.org/10.21283/2376905X.7.101
Cortés Velásquez, D., & Nuzzo, E. (2017). Disdire un appuntamento: Spunti per la didattica dell'italiano L2 a partire da un corpus di parlanti nativi. *Italiano LinguaDue*, *9*(1), 17–36. https://doi.org/10.13130/2037-3597/8761
Cortés Velásquez, D., & Nuzzo, E. (2021). Il progetto DISDIR: Origine, sviluppo e direzioni di ricerca. In E. Santoro, L. A. da Silva, & M. Zulma Kulikowski (Eds.), *Estudos em pragmáticas: Atos de fala em português, italiano, espanhol e* inglês (pp. 35–52.). Universidade de São Paulo.
Crystal, D. (2001). *Language and the internet*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164771
Derks, D., Bos, A. E. R., & von Grumbkow, J. (2008). Emoticons and online message interpretation. *Social Science Computer Review*, *26*(3), 379–388. https://doi.org/10.1177/0894439307311611
Dresner, E., & Herring, S. C. (2010). Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Communication Theory*, *20*(3), 249–268. https://doi.org/10.1111/j.1468-2885.2010.01362.x
Félix-Brasdefer, J. C. (2003). Declining an invitation: A cross-cultural study of pragmatic strategies in American English and Latin American Spanish. *Multilingua-journal of Cross-cultural and Interlanguage Communication*, *22*(3), 225–255. https://doi.org/10.1515/mult.2003.012
Félix-Brasdefer, J. C. (2007a). Natural speech vs. elicited data: A comparison of natural and role play requests in Mexican Spanish. *Spanish in Context*, *4*(2), 159–185. https://doi.org/10.1075/sic.4.2.03fel
Félix-Brasdefer, J. C. (2007b). Pragmatic development in the Spanish as a FL classroom: A cross-sectional study of learner requests. *Intercultural Pragmatics*, *4*(2), 253–286. Scopus. https://doi.org/10.1515/IP.2007.013
Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 41–56). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.26.03fel
Félix-Brasdefer, J. C., & Hasler-Barker, M. (2017). Elicited data. In A. Barron, Y. Gu, & G. Steen (Eds.), *The Routledge handbook of pragmatics* (pp. 27–40). Routledge.
Flores-Salgado, E., & Castineira-Benitez, T. A. (2018). The use of politeness in *WhatsApp* discourse and move 'requests'. *Journal of Pragmatics*, *133*, 79–92. https://doi.org/10.1016/j.pragma.2018.06.009
Goffman, E. (1959). *The presentation of self in everyday life*. Doubleday.
Goffman, E. (1967). *Interaction ritual: Essays in face-to-face behavior*. Doubleday-Anchor.
Golato, A. (2017). Naturally occurring data. In A. Barron, G. Steen, & Y. Gu (Eds.), *The Routledge handbook of pragmatics* (pp. 21–26). Routledge.
Gomes, A. P. Q. (2020). Variação pragmática: Cortesia linguística e intercompreensão cultural entre brasileiros e portugueses. *Web Revista SOCIODIALETO*, *10*(30), 306–316.
González Berrio, S., Martín Leralta, S., & Aparecida Rocha, N. (2021). "Me es imposible": Elección del género discursivo y estrategias pragmáticas del rechazo en la comunicación académica digital de españoles y brasileños. *Pragmalingüística*, *29*, 190–210. https://doi.org/10.25267/Pragmalinguistica.2021.i29.10
Herring, S. (2004). Computer-mediated discourse analysis: An approach to researching online communities. Em S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for virtual communities in the service of* learning (pp. 338–376). Cambridge University Press. https://doi.org/10.1017/CBO9780511805080.016
Kasper, G. (2008). Data collection in pragmatics research. In H. Spencer-Oatey (Ed.), *Culturally speaking: Managing rapport through talk across* cultures (pp. 279–303). Continuum.

Leech, G. (2014). Methods of data collection: Empirical pragmatics. In G. Leech (Ed.), *The pragmatics of politeness* (pp. 248–261). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195341386.003.0009

Locher, M. A. (2010). *Introduction: Politeness and impoliteness in computer-mediated communication*. *6*(1), 1–5. https://doi.org/10.1515/jplr.2010.001

Longo, F. (2020). La punteggiatura nella comunicazione mediata dal computer in lingua italiana: Stato dell'arte. *Lingue e culture dei media*, *4*(1), 41–62. https://doi.org/10.13130/2532-1803/14062

Macário Lopes, A. C. (2019). A deixis emocional em português europeu contemporâneo: Alguns contributos. *Revista da Associação Portuguesa de Linguística*, *5*, 225–235. https://doi.org/10.26334/2183-9077/rapln5ano2019a16

Maingueneau, D. (2005). Ethos, cenografia, incorporação. In R. Amossy (Ed.), *Imagens de si no discurso: A construção do ethos* (pp. 69–92). Contexto.

Mateus, M. H. M. (2003). Se a língua é um factor de identificação cultural, como se compreende que a mesma língua identifique culturas diferentes? *Revista de Letras*, *1*(25), 9–21. http://periodicos.ufc.br/revletras/article/view/2246

Nuzzo, E., & Cortés Velásquez, D. (2020). Canceling last minute in Italian and Colombian Spanish: A cross-cultural account of pragmalinguistic strategies. *Corpus Pragmatics*, *4*(3), 333–358. https://doi.org/10.1007/s41701-020-00084-y

Nuzzo, E., & Santoro, E. (2017). Apprendimento, insegnamento e uso di competenze pragmatiche in italiano L2/LS: La ricerca a partire dagli anni Duemila. *EuroAmerican Journal of Applied Linguistics and Languages*, *4*(2), 1–27. https://doi.org/10.21283/2376905X.7.116

Ogiermann, E. (2018). Discourse completion tasks. In A. Jucker, K. P. Schneider, & W. Bublitz (Eds.), *Methods in pragmatics* (pp. 229–256). De Gruyter. https://doi.org/10.1515/9783110424928-009

Pallotti, G. (2019). Assessing tasks: The case of interactional difficulty. *Applied Linguistics*, *40*(1), 176–197. https://doi.org/10.1093/applin/amx020

Petitjean, C., & Morel, E. (2017). "Hahaha": Laughter as a resource to manage *WhatsApp* conversations. *Journal of Pragmatics*, *110*, 1–19. https://doi.org/10.1016/j.pragma.2017.01.001

Placencia, M. E., & Garcia-Fernandez, C. (Eds.). (2007). *Research on politeness in the Spanish-speaking world*. Routledge. https://doi.org/10.4324/9781315088822

Rocha, B., Torre, M., & Mello, H. (2015). Estratégias de cortesia no ato de fala da recusa a pedidos: Uma investigação do português brasileiro e do italiano. *Revista de Italianística*, *30*, 61–79. https://doi.org/10.11606/issn.2238-8281.v0i30p61-79

Sánchez-Moya, A., & Cruz-Moya, O. (2015). "Hey there! I am using WhatsApp": A preliminary study of recurrent discursive realisations in a corpus of WhatsApp statuses. *Procedia - Social and Behavioral Sciences*, *212*, 52–60. https://doi.org/10.1016/j.sbspro.2015.11.298

Santoro, E. (2017). Richieste e attenuazione: Un confronto tra italiano e portoghese brasiliano. *Normas*, *7*(2), 179–204. https://doi.org/10.7203/Normas.v7i2.11173

Santoro, E., Cortés Velásquez, D., & Nuzzo, E. (2019). Tra consapevolezza pragmatica e intercomprensione: Uno studio esplorativo sull'atto linguistico della disdetta con italiani e brasiliani. n E. Nuzzo & I. Vadder (Eds.), *Lingua in contesto: La prospettiva pragmatica* (pp. 169–184). Associazione Italiana di Linguistica Applicata.

Santoro, E., Silva, L. A. da, & Kulikowski, M. Z. (Orgs.). (2021). *Estudos em Pragmática: Atos de fala em português, italiano, espanhol e inglês*. Portal de Livros Abertos da USP. https://doi.org/10.11606/9786587621661

Schneider, K. P. (2018). Methods and ethics of data collection. In A. Jucker, K. P. Schneider, & W. Bublitz (Eds.), *Methods in pragmatics* (pp. 37–94). De Gruyter Mouton. https://doi.org/10.1515/9783110424928-002

Seara, I. R. (2017). A atenuação como estratégia verbal de aproximação e de cortesia. In A. Ciama, M. Enachescu, & A. Gebaila (Eds.), *L'expression de l'imprécision dans les langues romanes* (pp. 306–318). Ars Docendi - Universitatea din Bucaresti. http://hdl.handle.net/10400.2/9395

Seara, I. R. (2024). Poder, persuasão, exibição: Análise de mensagens de "influenciadoras" na rede social Instagram. *Linha D'Água*, *37*(1), 128–148. https://doi.org/10.11606/issn.2236-4242.v37i1p128-148

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, *5*(1), 1–23. https://doi.org/10.1017/S0047404500006837

Simone, R. (2012). *Presi nella rete: La mente ai tempi del web*. Garzanti.

Spencer-Oatey, H. (2007). Theories of identity and the analysis of face. *Journal of Pragmatics*, *39*(4), 639–656. https://doi.org/10.1016/j.pragma.2006.12.004

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, *4*(2), 91–112. https://doi.org/10.1093/applin/4.2.91

Verschueren, J. (2016). Contrastive pragmatics. In J. Östman & J. Verschueren (Eds.), *Handbook of pragmatics* (Vol. 20, pp. 1–45). John Benjamins.

Wierzbicka, A. (1985). Different cultures, different languages, different speech acts: Polish vs. English. *Journal of Pragmatics*, *9*(2), 145–178. https://doi.org/10.1016/0378-2166(85)90023-2

Wierzbicka, A. (2003). *Cross-Cultural Pragmatics: The Semantics of Human Interaction* (2.ª ed.). De Gruyter Mouton. https://doi.org/10.1515/9783110220964

Yus, F. (2011). *Cyberpragmatics. Internet-mediated communication in context*. John Benjamins Publishing Company. https://doi.org/10.1075/pbns.213

“Mas és português ou brasileiro?” O efeito do contacto entre variedades do português na fala de imigrantes brasileiros em Portugal

“But are you Portuguese or Brazilian?” The effect of contact between varieties of Portuguese in the speech of Brazilian immigrants in Portugal

Ronan Pereira[1], *Catarina Rosa*[1], *Mariana Silva*[1]

[1] Centro de Linguística da Universidade NOVA de Lisboa

Abstract

This study investigates the perception of Brazilian Portuguese native speakers' accent by native European Portuguese speakers in a context of contact between these dialects in Portugal, crossing these results with individual factors that are traditionally considered in second dialect acquisition (SDA), given the expressive variability in the outcomes of this process (Siegel, 2010). Audio snippets of 30 immigrants who had been living in continental Portugal for over six years (as well as of 8 Brazilian non-immigrants and 12 Portuguese who served as control groups) were evaluated by 47 EP natives, who determined if the speakers sounded Brazilian or Portuguese and how certain they were of their evaluation by means of a Likert-like scale from 1 (Brazilian: I'm sure) to 6 (Portuguese: I'm sure). The evaluators easily identified Brazilian (M = 1.17) and Portuguese (M = 5.95) controls; Brazilian immigrants, on the other hand, were, in general, identified as Brazilians, but with a significantly lower level of certainty (M = 1.88, $p < .001$) comparing with non-immigrants. An analysis of individual factors showed that age of arrival and the will to be perceived as Portuguese were statistically significant ($ps \leq .039$). Nonetheless, individual analyses do not indicate many effects of linearity between variables and evaluations. This suggests that the weight of each variable at an individual level will be different, which helps to explain the different outcomes seen in SDA.

Keywords: bidialectalism, phonology, accent rating, individual factors.

Resumo

Este estudo investiga a perceção do sotaque de falantes nativos do português brasileiro por falantes nativos do português europeu num contexto de contacto entre estes dialetos em Portugal, cruzando estes resultados com fatores individuais que são tradicionalmente considerados na aquisição de segundo dialeto (AD2), dada a expressiva variabilidade nos desfechos deste processo (Siegel, 2010). Trechos de áudios de 30 imigrantes que residiam em Portugal continental há mais de seis anos (bem como de 8 brasileiros não imigrantes e 12 portugueses que serviram de grupos de controlo) foram avaliados por 47 nativos do português europeu, que determinaram se os falantes soavam como brasileiros ou portugueses e o quão certos estavam da sua avaliação por meio duma escala de Likert de 1 (brasileiro: tenho a certeza) a 6 (português: tenho a certeza). Os avaliadores facilmente identificaram os controlos brasileiros (M = 1,17) e portugueses (M = 5,95); já os imigrantes brasileiros foram, em geral, identificados como brasileiros, mas com um grau de certeza significativamente mais baixo (M = 1,88, $p < 0{,}001$) comparativamente aos brasileiros não imigrados. Uma análise dos fatores individuais para uma maior probabilidade de serem avaliados como portugueses determinou que a idade de chegada e a vontade de ser percecionado como português foram estatisticamente significativos ($ps \leq 0{,}039$). Todavia, análises individuais apontam para poucos efeitos de linearidade entre as variáveis e as avaliações. Isto sugere que o peso de cada variável será diferente a nível individual, o que ajuda a explicar os diferentes desfechos observados na AD2.

Palavras-chave: bidialetalismo, fonologia, avaliação de sotaques, fatores individuais.

1. Considerações iniciais

A aquisição de segundo dialeto (AD2) é o processo de aquisição de um sistema linguístico adicional que envolve variedades da mesma língua. É vista como um tipo especial de aquisição de língua segunda (L2), pois, para fins comunicativos, dado o provável alto grau de inteligibilidade mútua entre o dialeto nativo (D1) e o segundo dialeto (D2), não será necessário, à partida, que os falantes "aprendam" o D2 (Siegel, 2010, p. 1). Em geral, a AD2 ocorre em diferentes contextos que envolvem o contacto entre dois (ou mais) dialetos, como, por exemplo, em zonas onde dois dialetos co-existem historicamente (e.g., zonas da Alemanha onde um dialeto local coexiste com o alemão padrão) ou quando indivíduos saem da sua zona dialetal e estabelecem residência noutra zona e passam a ter contacto mais intenso com o D2 ali falado (Chevrot & Ghimenton, 2018, p. 513).

A produção científica em AD2 resultante de migrações debruçou-se principalmente na componente fonológica, tendo feito uso de metodologias diversas que, em geral, recolheram dados de produção, focando-se naquilo que era diferente entre as variedades, isto é, as diferentes realizações das variáveis linguísticas, e a conclusão de que houve aquisição ou não advém do cálculo da percentagem de uso da variante do D2 frente à do D1 (Siegel, 2010, p. 22). Observa-se que, no geral, os resultados da AD2 são muito inconsistentes: por um lado, há participantes que não utilizam nenhuma das variantes do D2; por outro, há evidências de aquisição (tanto entre adultos, quanto entre crianças). Neste último caso, não é incomum que se desenvolva o seguinte cenário: os falantes são identificados como falantes não nativos do D2 por falantes deste dialeto, mas também podem soar estrangeiros aos ouvidos dos falantes do seu D1 (Siegel, 2010, p. 55).

De facto, Kupisch et al. (2023) observam exatamente isto entre falantes nativos do português brasileiro (PB) que imigraram para Portugal em idade adulta: os falantes nativos do PB reconhecem que esses imigrantes já não soam completamente como os brasileiros não imigrados, ao mesmo tempo que os falantes nativos do português europeu (PE) não consideram que soam como os portugueses. Neste sentido, o D1 e o D2 surgem como extremidades dentro de um contínuo dialetal e a fala dos falantes bidialetais jaz nalgum ponto desse contínuo (cf. Chambers, 1992, p. 695). Importantemente, isto significa que há indivíduos que adquirem (algumas) propriedades do D2 (ainda que possam diferir muito entre si, com diferentes taxas de uso de variantes por indivíduo), mas outros parecem não o conseguir (Siegel, 2010, p. 51). Dada esta alta variabilidade nos desfechos observados, sugere-se que certas variáveis (como a exposição ao D2 e as atitudes individuais) poderão estar por detrás da promoção ou da inibição da AD2, um ponto que Kupisch et al. (2023) não abordaram.

O presente estudo tem como objetivo complementar o estudo de Kupisch et al. (2023), nomeadamente pela identificação das possíveis variáveis que condicionarão a influência do PE no sotaque dos nativos do PB imigrados em Portugal. Este artigo está organizado da seguinte maneira: na Secção 2 traz-se a fundamentação do estudo, abordando-se alguns fatores individuais que parecem condicionar a AD2 e a aquisição de sotaque não nativo, bem como as principais diferenças fonológicas entre as duas variedades do português aqui consideradas; na Secção 3 descreve-se o estudo experimental, apresentando as questões de investigação e hipóteses, a metodologia e os resultados; na Secção 4 discutem-se esses resultados; por fim, a Secção 5 traz as considerações finais.

2. Fundamentação

2.1. Fatores individuais na aquisição de segundo dialeto

Tradicionalmente, os estudos em aquisição da fonologia duma L2 consideraram como importantes algumas variáveis individuais neste processo, como a idade de aquisição e o grau de exposição ao *input* da L2 (Piske et al., 2001). Similarmente, os estudos em AD2 no contexto de migrações também se debruçaram sobre tais fatores. Por exemplo, Siegel (2010, p. 84) comenta que os indivíduos com uso consistente das variantes do

D2 (acima de 90%) ou que chegam a obter competência quase nativa são aqueles que foram expostos ao D2 quando jovens, especialmente se tinham 13 anos ou menos. Já Payne (1980, p. 175), no seu estudo com propriedades fonológicas de dialetos do estado da Filadélfia nos Estados Unidos, concluiu que o sucesso na aquisição dos traços é mais provável quando a idade de primeiro contacto se dê até aos oito anos de idade.[1]

Seguindo as ideias aventadas pela Hipótese do Período Crítico (Lennenberg, 1967), de que a aquisição normal de certas propriedades duma língua após um período ótimo para tal não ocorre, Long (2007, p. 67) considera a existência de períodos sensíveis em que há maior facilidade na aquisição da fonologia da L2 (e, por extensão, do D2). Estes períodos durariam até aos sete anos de idade para as regras fonológicas complexas e até aos 13 anos de idade para regras fonológicas simples e suprassegmentais. Havendo a aquisição dentro desses períodos, chegar à proficiência quase nativa não seria garantida, mas tampouco seria algo inesperado; após, contudo, não seria esperado que acontecesse, embora não fosse impossível.

Por outro lado, enquanto a idade de chegada à zona de acolhimento parece ter um papel fundamental, o tempo de residência não. Payne (1976), por exemplo, observou, após dividir os participantes de acordo com o seu tempo de residência (de quatro a sete anos e de oito a 16 anos), um efeito do tempo de residência, mas que estava correlacionado com uma idade de aquisição precoce (indivíduos que foram expostos entre os cinco e os nove anos de idade). Apesar das fracas evidências sobre a relevância do tempo de residência no sucesso da AD2, Foreman (2003, p. 229) relata que todos os falantes no seu estudo com falantes do inglês americano imigrados na Austrália que utilizaram variantes do D2 tinham cinco anos de residência ou mais. Mais além, a autora refere que o uso do D2 na residência dos imigrantes também se mostrou relevante[2] (p. 230). Em conjunto, estes dados reforçam a importância da exposição ao *input*, sugerindo que uma quantidade mínima de exposição a ele será necessária (embora a simples exposição ao *input* não seja o bastante para a aquisição ocorrer, principalmente no caso de estruturas pouco salientes e/ou comunicativamente redundantes; cf. Long, 2007, p. 112).

Uma parte substancial da AD2 passa pelas atitudes individuais dos falantes em contextos bidialetais. Lecky (1961, pp. 51–52) afirma que os indivíduos não mudam facilmente a sua identidade e os conceitos que têm sobre si mesmos, visto que construí-los é um trabalho árduo. Destarte, o modo de falar é uma das partes intrínsecas à autenticidade dos indivíduos e alterá-lo sem nenhuma razão comunicativa (como ocorre, por exemplo, na aquisição de L2) seria visto como falso ou não autêntico (Siegel, 2010, p. 149). Todavia, se um falante de um D1 se identificar com os falantes (ou com a região) do seu D1, tal facto acaba por dificultar a AD2 (Stanford, 2008); em compensação, pode-se postular que uma maior identificação com a comunidade ou com a região do D2 estimulará a AD2. Em Foreman (2003, pp. 234–245), dos 12 participantes que adquiriram traços do inglês australiano, oito afirmaram ter, ainda que parcialmente, uma identidade australiana, enquanto outros dois não se identificavam nem com a nação de residência, nem com a nação de origem; já os que mantiveram os traços do D1 identificavam-se mais com o seu país de origem.

Não obstante, estes dados têm de ser vistos com cuidado, pois pode ter havido efeitos de idade de chegada: os que se identificavam mais com o país de acolhimento chegaram mais jovens, pelo que tiveram mais hipóteses de adquirir os traços do D2. Ainda assim, a autora refere três participantes que chegaram durante a infância à Austrália (com sete, oito e dez anos). Enquanto o primeiro atingiu uma competência quase nativa no D2, os outros dois adquiriram somente alguns traços. A análise consoante à identificação com a Austrália mostrou que o primeiro se identificava fortemente com aquele país, mas os outros dois não.

Note-se que as variedades regionais ou sociais podem ter um estatuto sociocultural muito mais proeminente do que as línguas. Noutras palavras, os dialetos estão mais ligados à identidade dos falantes do que

[1] Note-se que indivíduos que tenham chegado à área do novo dialeto quando jovens, muito provavelmente terão tido uma parcela considerável do seu ensino formal no D2, o que pode, também, influenciar o processo de aquisição (Siegel, 2010, p. 86).

[2] Em contrapartida, a exposição ao D2 no ambiente de trabalho não apresentou significância estatística. A autora acredita que isto tenha ocorrido pelo facto de os falantes trabalharem em empresas australianas, pelo que esta variável não deverá apresentar variação o suficiente para ser detetada pelos métodos estatísticos empregados (Foreman, 2003, p. 209).

as línguas, e falar um ou outro dialeto é uma característica que os distingue como pertencentes a um certo grupo[3] (Rickford & Rickford, 2000, p. 22); por conseguinte, um falante de um D1, ao falar o D2, pode ter reações negativas dos falantes nativos deste dialeto (e até mesmo dos falantes do D1) – um tipo de reação raro na aquisição duma L2 (Siegel, 2010, pp. 146–147). Por exemplo, em Jacobs-Huey (1997, p. 356), três homens (dois nativos do inglês afro-americano e um estudante universitário branco de classe média interessado em e identificado com a cultura afro-americana) foram gravados a falar o dialeto afro-americano e as gravações foram posteriormente avaliadas por falantes de diferentes origens: afro-americanos, anglo-americanos, latino-americanos e ásio-americanos. O estudante foi considerado um falante nativo do inglês afro-americano baseado na sua fala por 92% dos participantes. Todavia, ao saberem que se tratava de um homem branco de classe média, muitos dos avaliadores demonstraram reações negativas (Jacobs-Huey, 1997, p. 363).

É importante ressaltar que esse sentimento de pertença e de propriedade não se restringe a variedades de menor prestígio (como no caso do inglês afro-americano). Por exemplo, há reações negativas de falantes do inglês britânico D1 para com falantes do inglês canadiano D1 que estavam a alterar a sua pronúncia em direção àquela variedade. Para Giles e Smith (1979, p. 62), tais reações advêm da sensação que os falantes do D1 têm da perda da sua distinção cultural aquando da aquisição pelos canadianos da característica mais fundamental da sua identidade (o sotaque britânico). Neste sentido, se falar um dialeto servir para sinalizar a pertença a um grupo social em particular, somente aqueles que são membros desse grupo terão o direito de o falar (Siegel, 2010, p. 153).

Por outro lado, Ekberg e Östman (2020) relatam o caso duma imigrante bósnia numa região da Finlândia que possui o seu próprio dialeto do sueco (que difere do sueco finlandês padrão). Ela diz ser criticada na escola por não utilizar o dialeto local e preferir o sueco finlandês padrão; porém, quando o utiliza, é criticada por esses mesmos colegas por causa da sua pronúncia. Falantes do inglês norte-americano na Austrália têm experiências similares. Foreman (2003, p. 239) aponta que um informante no seu estudo afirmou que nativos do D1 se riram das suas tentativas de utilizar expressões típicas do inglês australiano. Ademais, outros informantes dizem temer ser vistos como "falsos" por terem adquirido o dialeto australiano. Como conclui a autora, a mudança em direção ao D2 não é vista como autêntica e seria um exemplo de falta de lealdade, como se a mudança no sotaque indicasse uma tentativa de pertencer a um lugar ao qual não se pertence ou de ser alguém (um australiano) que não se é (Foreman, 2003, p. 241).

Os estudos sobre a AD2 envolvendo variedades nacionais do português são recentes e demonstram que é possível a aquisição do PE por falantes nativos do PB, sendo condicionada, por exemplo, pela idade de chegada (se se deu durante a infância ou não) e da propriedade gramatical em questão, seja ela morfossintática (e.g., Castro et al., 2016, 2017, 2020; Pereira, em preparação) ou fonológica (e.g., Kupisch et al., 2023). Por o presente estudo investigar esta última componente gramatical, apresentam-se a seguir certos aspetos contrastivos entre a fonologia destas duas variedades.

2.2. **Diferenças fonológicas entre o português brasileiro e o português europeu**

Para os ouvidos daqueles que não conheçam as peculiaridades da fonologia do PE e do PB, num primeiro contacto com estas variedades, é provável que elas soem como duas línguas sem relação entre elas (cf. Mateus & d'Andrade, 2000, p. 129). Dadas as suas diferenças a nível fonológico, abordam-se aqui algumas das principais propriedades segmentais e suprassegmentais que as possam distinguir.[4]

[3] Um revisor anónimo questiona se os traços regionais mais salientes tendem a ser mais conservados por determinados grupos sociais. Espera-se este cenário desde que o indivíduo se identifique com o restante do grupo, como visto nos parágrafos anteriores. Por outro lado, espera-se o "abandono" desses traços ao não haver identificação ou por haver a identificação com outro grupo.

[4] Não se pretende fazer uma descrição exaustiva de todas as diferenças entre as duas variedades devido à restrição de espaço e também porque as duas variedades não são homogéneas. Isto é, nem sempre as mesmas propriedades se observam em todo o território onde cada variedade é falada. Neste sentido, muito do que é apresentado nesta Subsecção se baseia na pronúncia do português falado em Lisboa (para o PE) e em São Paulo e no Rio de Janeiro (para o PB). Veja-se Vigário (2022) e Rodrigues (2024) para variação no PE e Battisti (2021) para variação no PB.

Segundo Vigário (2022, p. 872), o arcabouço fonológico das duas variedades é muito similar. Relativamente às consoantes, as diferenças principais encontram-se na existência de processos específicos a cada variedade. Por exemplo, em posição de ataque silábico as consoantes plosivas dentais /t/ e /d/ são palatizadas no PB quando ocorrem antes de /i/ ou /j/ (ou das suas versões nasais), sendo realizadas como [tʃ] e [dʒ], respetivamente:

(1a) tio – [ˈtiu] (PE) / [ˈtʃiu] (PB)

(1b) dia – [ˈdiɐ] (PE) / [ˈdʒiɐ] (PB)

Em ambas as variedades a realização da consoante rótica em posição de ataque silábico é variável, podendo emergir como uma vibrante alveolar [r] (menos frequente no PB) ou uvular [ʁ], como as fricativas [ʁ] ou [χ] ou ainda aspiradas [h] (esta última realização restrita ao PB; cf. Rennicke, 2015, p. 40):

(2) rio – [ˈriu], [ˈʁiu], [ˈʁiu], [ˈχiu] (PE, PB), [ˈhiu] (PB)

No PE, em particular, as consoantes plosivas sonoras /b/, /d/ e /g/ podem ser realizadas como fricativas [β], [ð] e [ɣ], respetivamente, desde que se encontrem entre duas vogais ou precedidas por uma consoante líquida ou por uma semivogal oral (Rodrigues, 2024, p. 216):

(3) uma bodega – [ˈumɐ βu ˈðɛɣɐ] (PE) / [ˈumɐ buˈdɐgɐ] (PE, PB)

Outras diferenças entre as duas variedades surgem em segmentos que se encontram em coda silábica. Por exemplo, no PE, o fonema /l/ é realizado como [ɬ] nessa posição, ao passo que, no PB, o mesmo fonema é velarizado como [w]:

(4) final – [fi ˈnaɬ] (PE) / [fi ˈnaw] (PB)

Semelhantemente, sibilantes em posição de coda silábica são tendencialmente realizadas como fricativas pós-alveolares no PE, bem como na variedade do PB falada no Rio de Janeiro. Na variedade de São Paulo, a sua realização é tendencialmente alveolar (Callou, 2009, p. 134):

(5) dois – [dojʃ] (PE, PB – Rio de Janeiro) / [dojs] (PB – São Paulo)

A consoante rótica em coda silábica, estável no PE como [ɾ], apresenta-se variável no PB. Por exemplo, na variedade de São Paulo, a pronúncia mais comum é [ɾ] (embora certos falantes, principalmente oriundos do interior do Estado, possam usar uma pronúncia retroflexa [ɻ]); na variedade do Rio de Janeiro, a pronúncia é recuada, geralmente realizada como [χ] ou [h] (Callou, 2009, p. 138; Rennicke, 2015, pp. 46-47):

(6) mar – [maɾ] (PE, PB – São Paulo) / [maχ] (PB – Rio de Janeiro)

As diferenças entre as duas variedades serão mais percetíveis nos processos que envolvem os segmentos vocálicos, embora o arcabouço seja, novamente, semelhante.[5] Em geral, no que tange às sílabas tónicas, poucas

[5] Salienta-se que ambas as variedades possuem vogais nasais que têm um comportamento praticamente idêntico (vd. Mateus & d'Andrade, 2000). Por este motivo, a descrição dos sistemas vocálicos focar-se-á apenas nas vogais orais.

diferenças se veem.[6] Por exemplo, no PE, o fonema /e/ quando realizado antes de um som palatal é centralizado para [ɐ]:

(7) telha – [ˈtɐʎɐ] (PE) / [ˈteʎɐ] (PB)

As diferenças mais visíveis entre o PB e o PE estão na realização das vogais átonas. Em ambas as variedades, existe uma tendência para a elevação das vogais átonas pós-tónicas finais; contudo, o PE, além de elevar a vogal /e/, também a centraliza. Assim, ao passo que o PE apresenta as vogais [ɐ], [ɨ], [i][7] e [u], o PB apresenta somente [ɐ], [i] e [u]:

(8a) nome – [ˈnomɨ] (PE) / [ˈnomi] (PB)

(8b) júri – [ˈʒuri] (PE/PB)

(8c) casa – [ˈkazɐ] (PE, PB)

(8d) livro – [ˈlivru] (PE, PB)

O PE não apresenta distinção entre vogais átonas pré e pós-tónicas, aplicando os mesmos processos supracitados a todas as vogais átonas.[8] No PB, as vogais átonas em posição pré-tónica, em geral, mantêm-se, exceto pela perda de contraste entre as vogais médias:

(9a) sofá – [suˈfa] (PE) / [soˈfa] (PB)

(9b) menino – [mɨˈninu] (PE) / [meˈninu] (PB[9])

No PE, além da consistente elevação das vogais, observa-se a sua redução. De facto, em determinadas situações, é possível que a vogal seja apagada de todo, resultando, a nível fonético, na adjacência de sons outrora não "permitida". No PB, este fenómeno raramente ocorre:

(10) envelhecido – [ẽv(ɨ)ʎ(ɨ)ˈsidu] (PE) / [ẽveʎeˈsidu] (PB)

O PB, por outro lado, tende à epêntese de /i/ em encontros consonantais que não envolvam uma consoante líquida[10]:

(11) pneu – [pnew] (PE[11]) / [p(i)new] (PB)

---

[6] Note-se que determinadas palavras têm vogais médias de alturas distintas em cada variedade, o que fica claro pelo tipo de acento gráfico utilizado (quando necessário):
a. bebé – [bɛˈbɛ] (PE) / bebê – [beˈbe] (PB)
b. sinónimo – [siˈnɔnimu] (PE) / sinônimo – [siˈnonimu] (PB)

[7] Como bem aponta um revisor anónimo, a produção de [i] em posição final não é produtiva no PE, apesar de se verificar a sua realização em empréstimos ou em palavras cultas, tal como no exemplo (8b).

[8] No entanto, exceções há no PE (e.g., retórica – [ʁɛˈtɔrikɐ]). Para Vigário (2016, pp. 844–845), a presença de vogais átonas não reduzidas no PE advém de especificações a nível lexical (e não a nível morfológico), ou de condicionamentos prosódicos ou morfológicos.

[9] Em certas variedades do PB pode haver a elevação das vogais médias neste contexto quando a vogal da sílaba tónica for alta (i.e., harmonia vocálica; cf. Barbosa et al., 2019). Logo, a pronúncia [miˈninu] é também possível.

[10] Este processo parece ser tão recorrente a ponto de os falantes do PB percecionarem a existência de epêntese até mesmo quando nenhuma vogal está inserida entre consoantes, similarmente ao que ocorre com falantes de japonês (vd. Dupoux et al., 2011).

[11] Casos de epêntese no PE são raros, mas não estão ausentes de todo. Todavia, nesta variedade a vogal inserida é [ɨ] (Vigário, 2022, p. 856).

Tais processos contribuem para a perceção de ritmo diferenciado entre as duas variedades, pois a complexidade silábica parece estar associada a diferentes ritmos (Frota & Vigário, 2001). A literatura na área considera tradicionalmente a existência de um padrão rítmico baseado na regularidade temporal entre as sílabas (ritmo silábico) e outro baseado na regularidade temporal entre os acentos tónicos (ritmo acentual), além de um terceiro padrão baseado na regularidade temporal entre as moras[12] (ritmo moraico). Neste sentido, línguas com um ritmo acentual aparentam possuir uma maior variedade de tipos de sílabas, além de apresentarem processos fonológicos que envolvem a redução das suas vogais, enquanto as línguas com ritmo silábico não só possuem uma menor variedade de sílabas, como também apresentam processos fonológicos que simplificam a sua estrutura, como, por exemplo, a epêntese (Frota & Vigário, 2011).

Concernente às variedades do português aqui descritas, Frota e Vigário (2001) defendem que ambas apresentam um ritmo misto, mas com distribuição diferente. O PE, por um lado, assemelha-se às línguas germânicas no ritmo acentual quando se considera a variabilidade nos intervalos consonantais (ΔC), mas não se distancia das outras línguas românicas no seu ritmo silábico quando se considera a proporção de intervalos vocálicos[13] (%V). Já o PB possui uma variabilidade nos ΔC que o equipara a outras línguas românicas, mas a sua proporção de %V assemelha-se à de línguas de ritmo moraico, com o japonês e o basco.[14]

Assim, é provável que a adaptação à pronúncia do PE por falantes nativos do PB não seja uma tarefa simples, dadas as diferenças existentes entre as duas variedades, algumas delas subtis. Ao mesmo tempo, por conta dessas mesmas diferenças, a identificação dos falantes de cada variedade deverá ocorrer com relativa facilidade, mesmo que os falantes brasileiros apresentem um certo grau de influência do PE na sua fala (cf. Kupisch et al., 2023). Sendo este o ponto central deste estudo, volta-se a atenção a seguir à perceção de sotaque não nativo.

2.3. **Aquisição e perceção de sotaque não nativo**

Um sotaque não nativo é descrito como a pronúncia utilizada por aprendentes duma L2 que difere (muitas vezes sistematicamente) da pronúncia dos falantes nativos dessa L2 (Southwood & Flege, 1999). A literatura nesta área baseia-se largamente em tarefas de avaliação de sotaques. Nelas, falantes nativos duma determinada língua avaliam o quão nativo soam os falantes que a têm como L2. Os resultados são, em geral, variados: ao passo que alguns participantes são avaliados como não nativos, outros são avaliados como nativos.

Dada tal variabilidade, postula-se que alguns fatores individuais são responsáveis por modular a capacidade de um falante de L2 soar como um nativo. Piske et al. (2001) comentam que a literatura na área teve como foco fatores[15] como a idade de aquisição da L2, tempo de residência, género, instrução formal, motivação, aptidão em aprender línguas e a quantidade de uso da L1 e da L2; embora a idade de aquisição tenda a apresentar-se como o fator mais importante, todos os outros também parecem influenciar o sotaque dos participantes até certo ponto, embora nenhum deles aparente ser determinante. De facto, os autores puderam confirmar que tanto a idade de aquisição da L2 como a manutenção no uso da L1 tiveram um papel na avaliação de falantes de italiano imigrados na região anglófona do Canadá. Neste sentido, uma menor idade de aquisição e um menor uso da L1 estiveram associados a uma menor perceção de sotaque não nativo. Importantemente, até mesmo alguns falantes que chegaram ao Canadá em idade adulta soavam mais como nativos, especialmente

[12] A maioria das línguas românicas faz parte do primeiro grupo, as línguas germânicas fazem parte do segundo e o japonês surge como um representante do terceiro grupo (Abercrombie, 1967).
[13] As línguas de ritmo acentual tendem a apresentar sílabas que contêm mais consoantes, além de intervalos consonantais com duração variada. Isto resulta num maior rácio entre consoantes e vogais que se traduz numa %V mais baixa e num ΔC mais alto. As línguas de ritmo silábico seguem a tendência oposta (cf. Ramus et al., 1999).
[14] Adianta-se que o desenho experimental neste estudo não consistirá numa análise minuciosa da pronúncia dos falantes, daí a explicação resumida sobre as diferenças suprassegmentais entre o PE e o PB apresentada. Referem-se ao leitor interessado os textos originais citados, bem como os de Frota e de Moraes (2014) e de Frota et al. (2015) para dados contrastivos entre as variedades relativamente, entre outros, aos seus contornos entoacionais.
[15] Note-se que estes fatores são, basicamente, os mesmos que modulam a AD2 (cf. Subsecção 2.1).

se o seu uso de italiano fosse infrequente. Por outro lado, o género e o tempo de residência não tiveram um efeito significativo na perceção do seu sotaque.

Mais além da avaliação de sotaques não nativos, a investigação nesta área também se interessou pela perceção do sotaque de falantes não nativos sob a perspetiva dos preconceitos que os falantes nativos têm para com aqueles por possuírem uma pronúncia não nativa. Sendo o sotaque uma parte importante da identidade dos indivíduos, o qual também expressa um certo grau de informação social, como, por exemplo, a pertença a um determinado grupo numa sociedade (Edwards, 1999), as perceções dos falantes nativos relativamente a sotaques não nativos são, geralmente, negativas. Por exemplo, declarações anónimas proferidas por indivíduos que possuem um sotaque não nativo são, em geral, avaliadas pelos nativos como menos credíveis (Lev-Ari & Keysar, 2010). Também Lorenzoni et al. (2024) demonstram que os falantes não nativos são avaliados negativamente em relação a uma série de características sociais (e.g., honestidade, credibilidade, inteligência, competência, etc.). Como consequência, evidências demonstram o efeito desses preconceitos na obtenção de emprego por falantes não nativos: alguns empregadores são intolerantes a certos sotaques (Eisenstein, 1983) e quanto menos nativo um falante soar, menores são as hipóteses de ser percecionado como um professor e mais baixas são as suas hipóteses de ser contratado para essa posição (Boyd, 2003).

Salienta-se que as avaliações nem sempre são negativas. Por exemplo, falantes do inglês americano percecionam os falantes do inglês canadiano positivamente no que tange à amigabilidade e à competência. Em contrapartida, percecionam italianos falantes de inglês L2 positivamente em relação à amigabilidade, mas não em relação à competência; já asiáticos falantes de inglês L2 são percecionados negativamente em relação à amigabilidade, mas positivamente em relação à competência (Fiske, 2018; Lee & Fiske, 2006). No geral, tais avaliações advêm, ainda que parcialmente, do quanto um falante nativo consegue identificar um sotaque não nativo e, a partir daí, categorizá-lo e estereotipá-lo pela sua pertença a um determinado grupo (Dragojevic & Giles, 2016; Ryan, 1983).

No contexto de contacto entre o PE e o PB, Kupisch et al. (2023) preocuparam-se em avaliar o sotaque global de imigrantes nativos do PB em Portugal à procura de influência do PE na sua fala. Foram extraídos trechos de cerca de dez segundos das amostras de fala do estudo de Castro et al. (2020) em que os participantes tiveram de narrar uma história a um entrevistador nativo do PB (modo PB) e a um entrevistador nativo do PE (modo PE) em diferentes momentos. Os áudios provinham de 15 imigrantes precoces que chegaram a Portugal entre os 2 e os 8 anos de idade (média de 5;8) e de 15 imigrantes tardios que chegaram a Portugal entre os 19 e os 43 anos de idade (média de 25;10). Os falantes do primeiro grupo possuíam entre 14 e 45 anos de residência (média de 23;7) e indicaram usar o PB em 20,59% dos casos e o PE em 79,41%; já os do segundo grupo residiam em Portugal havia entre seis e 30 anos (média de 10;10) e utilizavam o PB em 58% dos casos e o PE em 42%. Além dos referidos participantes, trechos de fala de cinco falantes nativos do PB residentes no Brasil e de cinco falantes nativos do PE residentes em Portugal, os quais serviram como grupos de controlo, foram incluídos.

As amostras de áudio foram avaliadas por 25 falantes nativos do PB e 25 falantes nativos do PE originários de diferentes partes de cada país e que nunca tinham residido noutro país de língua oficial portuguesa além do seu país de origem. A tarefa consistia em escutarem as amostras e decidirem se o falante era ou brasileiro ou português e o quão certos estavam da sua avaliação. Importantemente, os avaliadores foram instruídos a ignorar traços morfossintáticos e lexicais nos áudios (embora os autores tenham tido o cuidado de não selecionarem trechos que incluíssem pistas de tal natureza), focando-se apenas na pronúncia.

Os autores transformaram as avaliações numa escala de 1 a 6, na qual o valor 1 correspondia à certeza de que o falante era brasileiro e o valor 6 à certeza de que era português. Os resultados indicaram que os avaliadores brasileiros e portugueses não tiveram problemas em identificar os controlos – foram facilmente identificados como brasileiros ou portugueses com alto grau de certeza. Os imigrantes precoces também foram percecionados como falantes do PE com um alto grau de certeza, mas somente nas amostras obtidas no modo PE; nas amostras no modo PB ambos os grupos de avaliadores os julgaram como falantes do PB, mas sem muita certeza. Já os imigrantes tardios foram percecionados como falantes do PB, mas os avaliadores diferiram no grau de certeza:

os avaliadores brasileiros tiveram menos certeza do que os avaliadores portugueses de que os trechos ouvidos pertenciam a um falante nativo do PB[16] (Tabela 1).

Tabela 1

Resultados da Tarefa de Avaliação de Sotaques de Kupisch et al. (2023)

	Modo PB			**Modo PE**		
	Controlo PB	Imigrantes precoces	Imigrantes tardios	Controlo PE	Imigrantes precoces	Imigrantes tardios
Avaliadores PB	1,01	2,77	2,41	5,93	5,21	2,27
Avaliadores PE	1,07	2,75	1,61	6,00	5,17	1,80

Os autores argumentam que as diferenças observadas entre o grupo que chegou durante a infância e o grupo que chegou em idade adulta se deve a restrições maturacionais, às quais se soma um maior *input* do PE precocemente. Mais além, apesar de não terem recolhido dados sobre este ponto, consideram a possibilidade de que imigrantes tardios possam, consciente ou inconscientemente, durante o seu longo período de residência em Portugal, ter suprimido certos traços mais salientes do seu sotaque original, ao mesmo tempo que adquiriram certos traços do PE.[17]

Apesar dos dados inovadores obtidos na sua pesquisa, os autores não se preocuparam em cruzar os resultados com outras variáveis além da idade de chegada a Portugal. Logo, segue pendente a identificação de fatores, principalmente daqueles associados à exposição ao PE e às atitudes individuais, que podem condicionar essa influência do PE na fala dos imigrantes nativos do PB, uma lacuna que este estudo pretende colmatar. Assim, descreve-se, a seguir, o estudo experimental.

3. Estudo experimental

3.1. Questões de investigação e hipóteses

Com o objetivo geral de verificar que variáveis individuais modulam a influência do PE[18] na fala de falantes nativos do PB que emigraram para Portugal em idade adulta, questiona-se o seguinte:

Questão 1: Até que ponto podem os nativos do PB imigrados em Portugal ser percecionados como falantes do PE por falantes nativos da variedade europeia?

Questão 2: O grau de sotaque dos imigrantes detetado pelos nativos do PE é moldado por fatores individuais dos imigrantes (idade de chegada, tempo de residência, exposição ao PE em casa, exposição ao PE no trabalho, identificação com Portugal e vontade de ser percecionado como um português)?

[16] Apesar de Kupisch et al. (2023) não discorrerem sobre isto, uma explicação para tal diferença poderá residir naquilo que Siegel (2010, p. 60) defende: ao interagir com um falante bidialetal, o foco de um falante nativo do mesmo D1 recai sobre os traços adquiridos do D2; já o foco de um falante nativo do D2 recai sobre os traços não adquiridos. Neste sentido, os avaliadores nativos do PB, por se focarem nos traços supostamente adquiridos do PE, podem ser confundidos por eles, diminuindo a sua certeza relativamente ao sotaque dos imigrantes tardios. Por outro lado, os avaliadores nativos do PE focam-se nos traços do PB, os quais provavelmente aparecem mais frequentemente, aumentando o seu grau de certeza. Ao mesmo tempo, isto também poderá explicar o porquê de não os julgarem a par com os controlos falantes do PB: por mais que os traços do PE não sejam o seu foco, eles ainda estão presentes e deverão ser de alguma forma percecionados e as suas avaliações refletem tal facto.

[17] De facto, Oliveira (2022, pp. 33–39) descreve, a partir de comentários em redes sociais e em notícias jornalísticas, preconceitos pelos quais brasileiros imigrados em Portugal passaram devido à variedade do português que falam.

[18] Seguindo um modelo conexionista da produção da linguagem (vd. Blanco-Elorrieta & Caramazza, 2021), assume-se neste trabalho que as evidências de influência, ainda que não emerjam consistentemente nas produções dos indivíduos, correspondem à aquisição dessas propriedades.

Hipotetiza-se o seguinte:

Hipótese 1: Similarmente ao que Kupisch et al. (2023) observaram, espera-se que esses falantes sejam percecionados como falantes do PB, mas com um grau de certeza mais baixo em comparação aos não imigrados, sugerindo uma possível influência fonológica do PE na sua fala.

Hipótese 2: Seguindo a literatura na área de AD2, espera-se um efeito dos fatores individuais, mas estes não serão transversais a todos os participantes, dada a variabilidade dos contextos de aquisição e do perfil dos imigrantes.

3.2. Metodologia

3.2.1. Participantes

Para o atingimento dos objetivos, realizou-se uma Tarefa de Perceção de Sotaques inspirada na tarefa aplicada por Kupisch et al. (2023). Um total de 63 falantes nativos do PE voluntariou-se, mas 15 foram excluídos por terem indicado no questionário sociolinguístico que viveram noutro país de língua oficial portuguesa ou que eram filhos de falantes de variedades do português que não a europeia. Dos 47 participantes (21 mulheres, 1 não-binário) cujos dados foram analisados, o questionário sociolinguístico apontou que todos eram escolarizados ao nível do ensino secundário, encontrando-se a pluralidade na faixa etária entre os 18 e os 29 anos (46,81%). Provinham de diversas regiões do país, sendo a maioria oriunda da região de Lisboa e Vale do Tejo (53,19%). Ademais, relatavam estar, em média, expostos ao PB e a variedades africanas do português a uma taxa de pouco mais de um terço durante uma semana típica (Tabela 2).

Tabela 2

Perfil Sociolinguístico dos Avaliadores

N	**47**	
Género	Feminino = 21 Masculino = 25 Não-binário = 1	
Faixa Etária	18-29 = 22 30-39 = 5 40-49 = 8 Mais de 50 = 13	
Escolaridade	Secundário = 7 Superior = 24 Pós-Graduação = 16	
Origem	Açores = 1 Alentejo = 1 Algarve = 7 Centro = 4	Lisboa e Vale do Tejo = 25 Madeira = 1 Norte = 8
Média de frequência de exposição a outras variedades do português numa semana típica (intervalo)	41% (0 – 97)	

3.2.2. Instrumento e procedimento

A tarefa foi administrada online por meio da plataforma Gorilla Software Builder[19] (Anwyl-Irvine et al., 2019) durante o mês de setembro de 2024. Os avaliadores inicialmente completaram o questionário

[19] www.gorilla.sc

sociolinguístico e em seguida realizaram a tarefa. Inicialmente, viam no ecrã uma cruz por 1000ms. Assim que desaparecesse, um ícone de som surgia e um áudio começava a tocar automaticamente. Os participantes foram instruídos a ouvi-lo e avaliar se o falante era brasileiro ou português, bem como o grau de certeza que tinham desse julgamento, por meio duma escala que surgia por debaixo do ícone de áudio quando o áudio terminava. Podiam também, se quisessem, carregar outra vez no ícone de áudio para ouvi-lo por (somente) mais uma vez.[20] Após terem tomado uma decisão, o processo repetia-se. A Figura 1 esquematiza a visão que os participantes tinham durante a realização da tarefa:

Figura 1

Ecrã de Realização da Tarefa



No total, tiveram de avaliar 50 trechos de áudios que provinham do banco de dados de Pereira (em preparação). Neles, ouvem-se os falantes a narrar um excerto a que assistiram de um filme mudo estrelado por Charles Chaplin, no qual, ao perambular por um zona degradada, encontra um bebé abandonado num beco e procura uma maneira de resolver este problema, encontrando diferentes personagens pelo caminho.[21] Estes trechos continham cerca de 11 segundos cada com pouca ou nenhuma pausa ou hesitação por parte dos falantes. Ademais, priorizaram-se trechos neutros, isto é, que não incluíssem pistas lexicais ou morfossintáticas que pudessem identificar a variedade nativa do falante. Trinta trechos de áudios correspondiam a falantes do PB a residir em Portugal (EXP) (12), 8 áudios correspondiam a falantes do PB a residir no Brasil (CPB) (13) e 12 áudios correspondiam a falantes do PE a residir em Portugal (CPE) (14). Estes dois últimos grupos serviram de grupos de controlo. O perfil destes falantes está descrito na Tabela 3.

(12) ... esse beco é um beco onde há muito lixo e, portanto, é a ideia de pouco valor que ali está e ele... ele vai em direção à saída do beco...

(13) ... quando ela se deu conta que não era a criança dela, brigou com ele, enfim, e fez com que ele retirasse a criança do carrinho...

(14) ... o vídeo é um excerto de um filme do Charles Chaplin em que a personagem do Chaplin anda num cenário meio destruído...

[20] Adianta-se aqui que nenhum dos avaliadores fez uso deste artifício.
[21] A cena da qual se extraiu o excerto pode ser vista na plataforma YouTube: https://www.youtube.com/watch?v=eO1HvF2G2Sw

Tabela 3

Perfil dos Grupos de Controlo (CPB e CPE) e do Grupo Experimental (EXP)

	CPB	EXP	CPE
n	8	30	12
Género	Feminino = 4 Masculino = 4	Feminino = 21 Masculino = 9	Feminino = 6 Masculino = 6
Média de idade (intervalo)	34 (22 – 44)	41;7 (24 – 63)	29;2 (19 – 42)
Escolaridade	Secundário = 3 Superior = 3 Pós-Graduação = 2	Secundário = 2 Superior = 8 Pós-Graduação = 20	Secundário = 4 Superior = 6 Pós-Graduação = 2
Origem[22]	Nordeste = 4 Sul = 4	Centro-Oeste = 3 Nordeste = 4 Sudeste = 17 Sul = 6	Alentejo = 1 Algarve = 4 Centro = 1 Lisboa e Vale do Tejo = 4 Norte = 2

Os falantes do grupo experimental chegaram a Portugal a partir dos 18 anos de idade (M = 31;6) e residiam em diversas regiões de Portugal continental, sendo a maioria na região de Lisboa e Vale do Tejo (60%), há, no mínimo, seis anos (M = 10;6). Numa semana típica, relataram estar mais expostos ao PE no ambiente de trabalho (70%) do que em casa (43,9%). Relativamente à sua identificação com Portugal, relatam identificar-se com este país numa média de 6,27 numa escala de 0 a 10, em que 0 correspondia a "nenhuma identificação" e 10 a "identificação total". Sobre a sua motivação autodeclarada em falar o PE, a média foi de 4,03 numa escala de 0 a 10, em que 0 correspondia a "nenhuma motivação" e 10 a "motivação total". A sua vontade de serem percecionados como falantes do PE em interações com desconhecidos portugueses atingiu uma média de 2,93 numa escala de 0 a 10, em que 0 correspondia a "nenhuma vontade" e 10 a "vontade total". Por fim, por meio duma escala em que 0 significava "nenhuma diferença" e 10 "completamente diferentes", consideram a pronúncia das duas variedades diferente (M = 7,47; DP = 1,99). Além disso, quando questionados sobre o aspeto que mais lhes causou dificuldade na compreensão dos falantes do PE ao chegarem a Portugal, 23 (76,7%) dos participantes relataram ser a pronúncia o principal fator. Similarmente, metade dos falantes refere ser a sua pronúncia o aspeto com que os falantes do PE mais tinham dificuldade aquando da sua chegada a Portugal. A Tabela 4 reúne estes dados:

[22] Um revisor anónimo questiona o facto de o grupo CPB não conter falantes do Sudeste, a região mais representativa entre os imigrantes. Sobre isto, recorda-se que os áudios provinham do banco de dados de Pereira (em preparação), o qual não continha amostras de falantes do Sudeste. No entanto, como se verá mais adiante, os resultados assemelham-se aos de Kupisch et al. (2023) cujos participantes do grupo de controlo PB provinham de diversas regiões do Brasil. Neste sentido, não parece que esta lacuna poderá ter sido relevante para os resultados deste estudo.

Tabela 4

Perfil dos Imigrantes

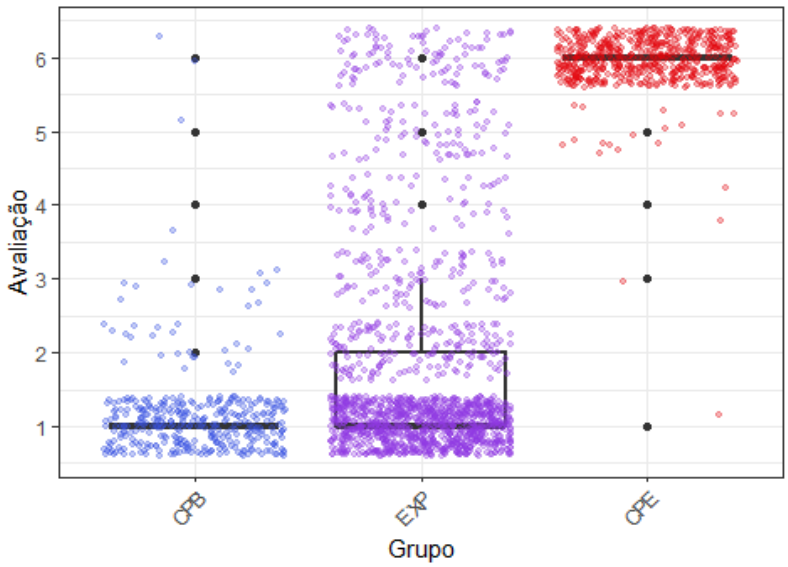
Residência	Alentejo = 1 Algarve = 1 Centro = 5	Lisboa e Vale do Tejo = 18 Norte = 5
Idade de chegada (média e intervalo)	31;6 (18 – 57)	
Tempo de residência (média e intervalo)	10;6 (6 – 34)	
Exposição ao português em casa (média e intervalo)	PB = 56,1% (0 – 100)	PE = 43,9% (0 – 100)
Exposição ao português no trabalho (média e intervalo)	PB = 30% (16,67 – 100)	PE = 70% (0 – 83,33)
Identificação com Portugal (média e intervalo)	6,27 (0 – 10)	
Motivação autodeclarada em falar PE (média e intervalo)	4,03 (0 – 10)	
Vontade de ser percecionado como falante do PE (média e intervalo)	2,93 (0 – 10)	
Perceção de distância fonológica (média e intervalo)	7,47 (5 – 10)	
Dificuldade mais frequente na compreensão do PE aquando da sua chegada	Estrutura da frase = 2 Pronúncia = 23 Vocabulário = 2 Sem problemas = 3	
Dificuldade mais frequente em serem compreendidos aquando da sua chegada	Estrutura da frase = 3 Pronúncia = 15 Vocabulário = 5 Sem problemas = 7	
Frequência com que portugueses já lhes disseram que têm um sotaque aportuguesado	Nunca = 5 Raramente = 13 Às vezes = 4 Frequentemente = 4 Muito frequentemente = 4	

3.3. **Resultados**

Para a análise (descritiva e estatística) dos resultados, utilizou-se o programa RStudio versão 2024.4.2.764 (Posit Team, 2024). Inicialmente, procedeu-se a uma conversão das respostas para uma escala de Likert de seis pontos codificada da seguinte maneira: 1 = brasileiro: tenho a certeza; 2 = brasileiro: tenho quase a certeza; 3 = brasileiro: não tenho a certeza; 4 = português: não tenho a certeza; 5 = português: tenho quase a certeza; 6 = português: tenho a certeza. A média aritmética das avaliações apontou que o grupo EXP obteve uma média de 1,88 (DP = 1,55), o grupo CPB obteve uma média de 1,17 (DP = 0,6) e o grupo CPE uma média de 5,95 (DP = 0,32). A distribuição das avaliações pode ser vista na Figura 2, em que cada ponto representa uma avaliação:

Figura 2

Gráfico de Caixas com as Avaliações da Tarefa de Avaliação de Sotaques por Grupo



Vê-se que as avaliações do grupo CPB se concentraram no valor 1, enquanto as avaliações do grupo CPE se concentraram no valor 6, ainda que, em ambos os grupos, tenha havido avaliadores que atribuíram outros valores aos áudios, mais notadamente no grupo CPB.[23] As avaliações do grupo EXP também se concentraram no valor 1; contudo, houve uma maior dispersão por toda a escala.

Devido à variável dependente (i.e., a avaliação) estar numa escala do tipo Likert, a análise estatística fez uso do pacote *ordinal* (Christensen, 2023) para a realização de modelos de regressão com dados ordinais. A avaliação da performance dos modelos foi feita pelo cálculo dos valores de R^2 (condicional e marginal) obtidos por meio do pacote *performance* (Lüdecke et al., 2021). Utilizando o grupo como efeito fixo e o avaliador e o falante como efeitos aleatórios, a análise intergrupos demonstrou que o grupo EXP é estatisticamente diferente do grupo CPB (β = -1,948, z = -2,479, p = 0,013) e do grupo CPE (β = 9,782, z = 12,699, p < 0,001) (Tabela 5[24]):

Tabela 5

Modelo de regressão ordinal entre grupos

	β	**erro padrão**	**valor *z***	**valor *p***
CPB	-1,948	0,786	-2,479	0,013*
CPE	9,782	0,770	-12,699	< 0,001*

Nota. R^2 Condicional = 0,882; R^2 Marginal = 0,695

[23] É possível notar áudios do grupo CPB avaliados acima do valor 3, bem como áudios do grupo CPE avaliados abaixo do valor 4. Embora isto possa sugerir que determinados falantes dos grupos de controlo pudessem ser percecionados como falantes da outra variedade, salienta-se que, assim que um participante selecionasse a sua resposta, a tarefa avançava. Neste sentido, se o participante o fizesse involuntariamente, já não era possível alterar a resposta. Em função do número residual de casos, é mais provável que essas avaliações especificamente tenham sido fruto desse tipo de descuido.

[24] Os valores na tabela referem-se à comparação de cada grupo de controlo com o grupo experimental. Note-se que, diferentemente de outros modelos, as regressões ordinais não apresentam valores para o intercepto.

Para a análise das variáveis individuais do grupo experimental, computou-se um modelo[25] (Tabela 6) contendo o avaliador e o falante como efeitos aleatórios e como efeitos fixos as seguintes variáveis: idade de chegada (IDEC), tempo de residência (TDER), exposição ao PE em casa (EXPC), exposição ao PE no trabalho (EXPT), identificação com Portugal (IDPT) e a vontade de ser percecionado como falante do PE por portugueses (QRPT). Verificou-se que somente as variáveis IDEC e QRPT apresentaram significância estatística (*ps* ≤0,039), sugerindo, respetivamente, que quanto mais cedo o imigrante tiver chegado a Portugal e quanto maior for a vontade ser percecionado como falante do PE por portugueses, maior é a probabilidade de ser avaliado como português na tarefa de perceção de sotaques (vd. Figura 1).

Tabela 6

Modelo de Regressão Ordinal com as Variáveis Individuais

	β	erro padrão	valor *z*	valor *p*
IDEC	-0,062	0,030	-2,067	0,039*
TDER	0,046	0,042	1,094	0,274
EXPC	0,011	0,008	1,407	0,159
EXPT	0,011	0,014	0,841	0,400
IDPT	0,126	0,095	1,330	0,184
QRPT	0,406	0,068	5,982	< 0,001*

Nota. R^2 condicional = 0,689; R^2 marginal = 0,361

Note-se que os dados dos modelos estatísticos consideram os resultados a nível do grupo. Para uma melhor visualização da variabilidade individual, optou-se por realizar uma análise detalhada das avaliações que os imigrantes receberam. Primeiramente, calculou-se a média aritmética das avaliações de cada participante e constatou-se que, dos 30 áudios incluídos na amostra, apenas nove possuíam uma média aritmética maior do que a do grupo; destes, quatro (A, B, C e D) eram percecionados como brasileiros, mas com um nível de certeza mais baixo, um (E) esteve basicamente no ponto central (o que se pode interpretar como falta de certeza de que seja brasileiro ou português), e quatro (F, G, H, e I) foram percecionados como portugueses, mas com diferentes graus de certeza (Tabela 7):

Tabela 7

Média do Grupo Experimental e dos Participantes com as Maiores Médias

	Grupo	A	B	C	D	E	F	G	H	I
Média	1,88	1,94	2,02	2,21	2,34	3,47	3,72	3,79	4,72	5,47

Para referência, calcularam-se as médias das variáveis para dois grupos, um (Alto) contendo os nove falantes supramencionados e outro (Baixo) contendo os demais falantes (Tabela 8):

Tabela 8

Comparação entre as Médias dos Grupos Alto e Baixo por Variável Individual

Grupo	Média	IDEC	TDER	EXPC	EXPT	IDPT	QRPT
Alto	3,30	27;6	15;11	70,37%	81,48%	6,33	5,67
Baixo	1,28	33;3	8;3	32,54%	65,08%	6,24	1,76

Vê-se, portanto, que não só a média das avaliações para o grupo Alto (3,30) é superior à do grupo Baixo (1,28), mas também a pontuação em todas as outras variáveis se mostra superior naquele grupo, à exceção da média de idade de chegada. Por fim, organizaram-se os dados de modo que demonstrassem como cada falante

[25] Outro modelo elaborado confirmou o facto de que os falantes que relatavam ouvir que tinham um sotaque "aportuguesado" mais frequentemente tiveram, tendencialmente, avaliações mais altas ($\beta = 0{,}056$, $z = 7{,}017$, $p < 0{,}001$).

respondeu às questões sobre as variáveis individuais na Figura 3, destacando-se em cinzento as respostas dos nove falantes do grupo Alto:

Figura 3

Distribuição das Respostas dos Participantes por Variável Individual

IDEC	TDER	EXPC	EXPT	IDPT	QRPT
18	6	0	16,67	0	0
20	6	0	33,33	0	0
20	6	0	33,33	1	0
22	6	0	50	1	0
23	6	0	50	4	0
23	6	0	50	5	0
23	6	0	50	5	0
23	6	0	50	5	0
23	6	0	66,67	5	0
25	6	0	66,67	5	0
27	6	33,33	66,67	6	0
29	7	33,33	66,67	6	0
30	7	33,33	66,67	6	0
30	7	50	66,67	7	0
30	7	50	66,67	7	1
30	8	50	83,33	7	2
31	8	50	83,33	7	2
32	9	50	83,33	7	2
32	9	50	83,33	7	3
33	9	50	83,33	7	4
33	9	66,67	83,33	8	5
33	10	66,67	83,33	8	5
35	13	66,67	83,33	8	5
39	15	83,33	83,33	8	5
42	16	83,33	83,33	8	5
42	16	100	83,33	10	9
44	17	100	83,33	10	10
47	22	100	100	10	10
49	26	100	100	10	10
57	34	100	100	10	10

Nota. Destaque cinzento nas respostas dos participantes do grupo Alto

Apresenta-se a discussão dos dados a seguir.

4. Discussão

Este estudo teve como objetivo complementar os dados obtidos por Kupisch et al. (2023) na sua Tarefa de Avaliação de Sotaques com imigrantes nativos do PB em Portugal ao analisar que variáveis individuais poderiam modular a influência do PE no sotaque de imigrantes adultos. O estudo original de Kupisch et al.

(2023) considerou apenas a idade de chegada dos participantes e demonstrou que, ao passo que os imigrantes precoces adaptavam a sua fala em direção à variedade do modo em que estavam a ser testados, o mesmo não se pôde dizer relativamente aos imigrantes tardios (cf. Tabela 3). Embora os avaliadores os tenham percecionado como brasileiros, o nível de certeza que apresentavam sobre tal facto era mais baixo do que o grau de certeza que tinham de que os falantes do grupo de controlo eram brasileiros, uma diferença estatisticamente significativa. Mais além, isto ocorreu independentemente do modo, ou seja, tanto no modo PB, quanto no modo PE, não foram percecionados a par com os falantes dos grupos de controlo, indicando que se encontram em pontos intermédios no contínuo dialetal PB-PE (cf. Chambers, 1992, p. 695), embora esse ponto se encontre mais próximo do PB do que do PE, confirmando a existência de influência do PE na sua fala, a qual se estende até mesmo a situações em que não precisariam de usar o PE (e.g., no modo PB).

Dada a Questão 1 deste estudo (*Até que ponto podem os nativos do PB imigrados em Portugal ser percecionados como falantes do PE por falantes nativos da variedade europeia?*), o desenho experimental considerou apenas a fala de imigrantes tardios ao serem entrevistados por um falante nativo do PE (i.e., modo PE no estudo supracitado) e avaliações de nativos do PE relativamente ao sotaque desses imigrantes. Viu-se que os avaliadores facilmente identificaram os controlos brasileiros e portugueses; já os imigrantes foram, a nível de grupo, identificados como brasileiros, mas com um grau de certeza significativamente mais baixo comparativamente aos brasileiros não imigrados, havendo um subgrupo de nove participantes cujas médias foram superiores à média do grupo de imigrantes. Estes resultados apresentam-se em linha com o que a Hipótese 1 previa e confirmam os resultados de Kupisch et al. (2023), ao observar-se que, em geral, os imigrantes tardios apresentam um certo grau de influência do PE na sua fala que os diferem dos nativos do PB residentes no Brasil (Tabela 9). Notavelmente, os resultados assemelham-se mesmo com as amostras alargadas de áudios e de avaliadores utilizadas neste estudo.

Tabela 9

Comparação entre os Resultados deste Estudo e de Kupisch et al. (2023)

	Este estudo	**Kupisch et al. (2023)**
Grupo de controlo PB	1,17	1,07
Imigrantes tardios	1,88	1,80
Grupo de controlo PE	5,95	6,00

A segunda análise realizada teve como intuito relacionar certas variáveis individuais com os resultados da Tarefa de Avaliação de Sotaque para responder à Questão 2 (*O grau de sotaque dos imigrantes detetado pelos nativos do PE é moldado por fatores individuais dos imigrantes (idade de chegada, tempo de residência, exposição ao PE em casa, exposição ao PE no trabalho, identificação com Portugal e vontade de ser percecionado como um português)?*). Os resultados demonstraram que, dentre as variáveis analisadas, somente a idade de chegada e a vontade de querer ser percecionado como português por portugueses foram estatisticamente significantes (cf. Tabela 6), o que possibilita uma resposta afirmativa parcial à Questão 2.

A idade de chegada costuma ser um fator importante quando se consideram imigrantes durante o período sensível para a aquisição da fonologia (cf. Long, 2007), pelo que se compreenderia a ausência de significância estatística com o grupo experimental deste estudo, pois todos chegaram a Portugal após os 18 anos de idade. Porém, os resultados indicam uma tendência duma maior probabilidade de ser percecionado como português quanto mais cedo o imigrante tiver chegado a Portugal. De facto, como visto na Figura 1, há diversas avaliações a rondar os valores 5 e 6 (os quais equivalem, respetivamente, a "português: tenho quase a certeza" e "português: tenho a certeza"). Embora isto se aplique, provavelmente, ao subgrupo com maiores avaliações (Tabelas 7 e 8), é possível que os mecanismos para a aquisição de propriedades fonológicas de outro sistema linguístico possam estar ativos mesmo após o período sensível, especialmente se outros fatores individuais estiverem presentes, como a vontade de ser percecionado como português.

Neste sentido, tal resultado sugere que, para estes imigrantes, uma parte importante de "ser português" abarca a variedade do português falada em Portugal, o que contrasta com a identificação com Portugal, que não apresentou diferença estatisticamente significativa. Relembra-se que estes imigrantes consideram alta a distância a nível fonológico entre as duas variedades, sendo também esta a componente em que mais encontraram dificuldade aquando da sua chegada a Portugal (cf. Tabela 4). É provável que, globalmente, deverá ser a fonologia a componente gramatical que mais associam a um português, sugerindo que será principalmente por esta via que a vontade de ser percecionado como um se manifestará. Note-se, porém, que ser percecionado como português não se resume apenas à variedade do português em Portugal: há também um nível extralinguístico, que inclui fatores culturais e comportamentais, bem como a inserção dos falantes na comunidade portuguesa, e poderá moldar a perceção de que alguém é português (vd. Sobral, 2012, para uma discussão). Com efeito, um indivíduo surdo-mudo poderá ser percecionado como português por meio desses outros fatores; logo, o nível linguístico, apesar de relevante, não será exclusivo. Assim, sugere-se que investigações futuras que tencionem recolher dados sobre estas variáveis sejam mais explícitas sobre o tipo de informação que pretendem obter dos falantes.

Voltam-se as atenções agora para as variáveis que não apresentaram significância estatística. Sobre o tempo de residência, a literatura em AD2 já sugere que um aumento no tempo de residência não leva a um aumento expressivo na produção de variantes do D2, pelo que este resultado não é de todo surpreendente. Não obstante, salienta-se que sete dos nove falantes com a média mais alta (grupo Alto, referido na subsecção anterior) residiam em Portugal havia pelo menos 9 anos aquando da realização do estudo. Por outro lado, visto que metade da amostra residia em Portugal havia seis ou sete anos, poderá ter ocorrido um efeito da falta de variabilidade nos dados, pelo que se sugere que este ponto seja visto com mais atenção futuramente.

Ressalta-se, também, a inexistência de significância estatística na exposição ao PE, seja em casa, seja no trabalho. Foreman (2003, p. 239) relata uma relação entre a exposição ao D2 em casa, mas não no trabalho, o que a autora considera advir do facto de o D2 ser a variedade principal utilizada nesse meio na zona de acolhimento, pelo que pouca variação se deteta neste âmbito a ponto de estabelecer algum tipo de associação (vd. Nota 2). De facto, neste estudo, os falantes relatam estar expostos ao PE no trabalho cerca de 70% do seu tempo; mais além, 22 dos 30 participantes reportam estar expostos a esta variedade pelo menos dois terços do seu tempo (vd. Figura 3), o que parece insinuar que, novamente, a baixa variabilidade estará por detrás da sua falta de significância.

Ademais, poder-se-ia considerar que os falantes mais expostos ao PE no trabalho seriam os que mais se adaptariam a esta variedade, apresentando, a longo prazo, uma maior influência do PE na sua fala, dado o nível mais formal do ambiente de trabalho, o qual poderá criar pressões sociolinguísticas no sentido de reforçarem o uso duma norma (vd. Holmes & Stubbe, 2003); estas pressões não estariam presentes no ambiente domiciliar. Sem embargo, é preciso clarificar que, mais além da simples exposição ao PE, se faz também necessário determinar que tipo de trabalho esses falantes exercem. Ou seja, mesmo que dois indivíduos estejam expostos ao PE com a mesma frequência no ambiente de trabalho, isto não significa que a possível pressão sociolinguística que sofrerão será a mesma. Finalmente, mais além da frequência de exposição ao PE, também se deve questionar se os imigrantes passam a mesma quantidade de horas nesses ambientes (casa e trabalho), algo em que investigações futuras se podem debruçar.

Os desfechos reportados, embora curiosos, não são inesperados. A Hipótese 2 já previa que as atitudes individuais e a exposição ao *input* seriam relevantes para a AD2, mas não seriam determinantes. A distribuição dos dados na Figura 3 mostra que as respostas dos participantes do grupo Alto se encontram, em geral, na base da tabela, exceto a idade de chegada, o que corrobora, em grande parte, os resultados do modelo de regressão ordinal (cf. Tabela 6). Ao mesmo tempo, verifica-se uma falta de linearidade nos resultados. Tome-se, por exemplo, a vontade de ser percecionado como português: embora quatro falantes do grupo Alto indiquem uma vontade total (ou quase total) de ser percecionado como português, outros três têm nenhuma (ou quase nenhuma) vontade, enquanto outros dois se situam num ponto intermédio. Por outro lado, um participante do grupo Baixo indica que uma vontade total de ser percecionado como português, ao passo que alguns outros se

encontram num ponto intermédio. Uma tendência similar observa-se com as outras variáveis, o que demonstra como as experiências dos imigrantes são tão particulares que cada variável poderá ter um peso distinto para cada um deles.

Por sinal, até mesmo quando o perfil dos imigrantes se assemelha é possível encontrar variabilidade. Trudgill (1986) relata o caso peculiar de irmãos gémeos canadianos imigrados na Austrália. Os seus perfis eram basicamente idênticos: tinham a mesma idade de chegada, o mesmo tempo de residência aquando da participação na pesquisa, frequentavam a mesma escola, tinham o mesmo grau de exposição ao *input*, tanto do inglês canadiano, quanto do inglês australiano, entre outros fatores. Ainda assim, cada um adquiriu propriedades fonológicas distintas do inglês australiano e, ao serem analisadas aquelas propriedades que ambos adquiriram, constatou-se que não foram adquiridas ao mesmo ritmo. Embora este caso seja aparentemente extremo, é de se concluir que, tendo os imigrantes neste estudo perfis sociolinguísticos particulares, a variabilidade apresentada não é surpreendente. Como argumenta Pereira (em preparação), "é possível que uma parte expressiva da variabilidade encontrada em estudos em AD2 provenha da variabilidade encontrada entre os perfis dos participantes."

Por fim, salienta-se que o facto de um falante não apresentar indícios de influência do PE não significa que elas não existam (cf. Nycz, 2015). Isto é, neste estudo, os avaliadores tomaram a sua decisão baseados em trechos de fala de cerca de 11 segundos extraídos de um contexto específico (a narração de um excerto de um filme). Ademais, recorde-se de que foi feito um esforço para evitar trechos que incluíssem pistas lexicais e morfossintáticas das variedades nos áudios analisados. Neste sentido, poderá ser o caso de a influência fonológica tender a surgir associada a propriedades de outras componentes gramaticais, especialmente se se considera um modelo conexionista da produção da linguagem (cf. Blanco-Elorrieta & Caramazza, 2021; veja-se também Lønes et al., 2023).

Ademais, mesmo que no trecho utilizado para a pesquisa nenhum indício possa ter estado presente, isto não significa que os falantes não os apresentem noutras situações fora do contexto experimental desta pesquisa. Com efeito, Siegel (2010, p. 5) comenta que os falantes em contextos de imersão noutro dialeto poderão acomodar (cf. Giles et al., 1973) a sua fala (por meio da sua monitorização) de modo que fique mais próxima (expressando aprovação ou solidariedade) ou mais distante da do interlocutor (expressando desaprovação ou distância), o que requer processos mais conscientes de controlo da produção linguística. Não obstante, por ora, não se pode determinar precisamente com estes dados que mecanismo(s) psicolinguístico(s) específico(s) poderá(ão) estar por detrás da presença (e da ausência) de indícios de influência do PE no sotaque da população estudada.

5. Considerações finais

Em suma, este estudo atingiu o objetivo proposto: determinar se e quais variáveis associadas à exposição ao PE e às atitudes individuais dos falantes nativos do PB imigrados em Portugal condicionavam a influência fonológica do PE na sua fala. Relativamente à avaliação global do sotaque dos participantes, ao replicar a Tarefa de Avaliação de Sotaques de Kupisch et al. (2023) com esta mesma população, obtiveram-se resultados extremamente semelhantes, reforçando o facto de que adquirir nem que sejam alguns traços fonológicos de outro dialeto é possível.

Importantemente, salienta-se que a tarefa, tal como foi constituída, avaliou somente a perceção global do sotaque dos participantes. Kupisch et al. (2023) relatam que alguns avaliadores no seu estudo apontaram informalmente certos traços segmentais e suprassegmentais que influenciaram a sua decisão. Por exemplo, mencionaram o uso elevado de [ʃ] em posição de coda silábica (tendencialmente realizada como [s] no PB), as vogais mais curtas e um discurso mais rápido dos imigrantes precoces no modo PB, e o falar mais "rítmico" dos imigrantes tardios no modo PE, o que parece refletir o ritmo silábico associado ao PB, com menos redução vocálica e epêntese (cf. Subsecção 2.2). Neste sentido, poderá ser também interessante avaliar que traços fonológicos específicos têm maior probabilidade de serem adquiridos (e aqueles que mais resistem à mudança)

e se são os mesmos que servem como pista para os avaliadores decidirem se o falante é brasileiro ou português. Ademais, visto que tanto o PB quanto o PE apresentam variação diatópica na componente fonológica (cf. Battisti, 2021; Vigário, 2022), será igualmente interessante determinar a importância das propriedades específicas do dialeto falado nas regiões de origem dos imigrantes e do dialeto falado nas suas zonas de residência em Portugal para este processo.

Sobre os fatores individuais, identificaram-se a idade de chegada e a vontade de ser percecionado como português como fatores relevantes na modulação dessa influência, corroborando a literatura na área, que demonstra a importância de tais variáveis para a AD2 e para a aquisição de sotaque não nativo (cf. Subsecções 2.1 e 2.3). Contudo, ainda não está claro como a vontade ser percecionado português age neste processo; isto é, não se consegue determinar que "incentivo" poderão ter estes falantes ao serem percecionados como portugueses. Dada a breve revisão da literatura na Subsecção 2.3, percebe-se que a perceção de sotaques não nativos por falantes nativos influenciará a experiência que os imigrantes têm na zona de acolhimento. No contexto de contacto entre o PB e o PE em Portugal, indaga-se se esta não poderá ser uma estratégia para minimizar algum tipo de impacto social (ou obter algum tipo de benefício social), dados os relatos de preconceito contra brasileiros decorrentes da sua identificação como tal por meio da sua fala (vd. Oliveira, 2022, pp. 33–39). Por outro lado, não se descartam também possíveis efeitos duma restruturação das suas gramáticas devido à exposição ao PE, sem haver, à partida, motivações sociais (cf. Nycz, 2015; Siegel, 2010, p. 5). Logo, sugere-se que investigações futuras intersecionem diferentes áreas, como a Sociolinguística e a Psicolinguística, ao considerarem mais detalhadamente as atitudes dos imigrantes e as suas experiências na zona de acolhimento, sem ignorar os contextos em que as interações ocorrem e os efeitos associados, por exemplo, à recência do uso do D2 e a exposição a ele, tendo sempre em conta a utilização de metodologias variadas que possam trazer diferentes perspetivas sobre o processo de AD2.

Agradecimentos

Agradece-se à Professora Doutora Tanja Kupisch (Universidade de Lund) pela assistência no estabelecimento do protocolo para este estudo, à Professora Doutora Ana Madeira e à Professora Doutora Alexandra Fiéis (Universidade NOVA de Lisboa) pelo auxílio na interpretação dos resultados, aos revisores anónimos pelos comentários e sugestões, ao público do 40.º Encontro Nacional da Associação Portuguesa de Linguística e à equipa editorial da Revista da Associação Portuguesa de Linguística. Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e Tecnologia, I.P., no âmbito do projeto UID/03213 – Centro de Linguística da Universidade NOVA de Lisboa (CLUNL) e da bolsa de doutoramento 2021.05667.BD atribuída ao primeiro autor deste artigo.

Referências

Abercrombie, D. (1967). E*lements of general phonetics*. Aldine.

Anwyl-Irvine, A., Massonnié, J., Kirkham, N., & Evershed, J. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 358–407. https://doi.org/10.3758/s13428-019-01237-x

Barbosa, P., Papa, P. B., Silva, B. A., & Mourão, N. (2019). Harmonia vocálica e coarticulação vogal a vogal em duas variedades do português brasileiro. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, *35*(2). https://doi.org/10.1590/1678-460x2019350202

Battisti, E. (2021, 23 fevereiro). Phonological variation and change in Brazilian Portuguese. *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.488

Blanco-Elorrieta, E., & Caramazza, A. (2021). A common selection mechanism at each linguistic level in bilingual and monolingual language production. *Cognition*, *213*, 104625. https://doi.org/10.1016/j.cognition.2021.104625

Boyd, S. (2003). Foreign-born teachers in a multilingual classroom in Sweden: The role of attitudes to foreign accent. *International Journal of Bilingual Education and Bilingualism*, *6*(3), 283–295. https://doi.org/10.1080/13670050308667786

Callou, D. (2009). Um perfil da fala carioca. In S. S. C. Ribeiro, S. B. B. Costa & S. A. M. Cardoso (Eds.), *Dos sons às palavras: Nas trilas da língua portuguesa* (pp. 129–152). EDUFBA.

Castro, T., Rothman, J., & Westergaard, M. (2016). Comparing anaphora resolution in early and late Brazilian Portuguese-European Portuguese bidialectal bilinguals: Theoretical and experimental issues. *Revista Española de Lingüística Aplicada*, *29*, 429–461. https://doi.org/10.1075/resla.29.2.03cas

Castro, T., Rothman, J., & Westergaard, M. (2017). On the directionality of cross-linguistic effects in bidialectal bilingualism. *Frontiers in Psychology*, *8*, 1–14. https://doi.org/10.3389/fpsyg.2017.01382

Castro, T., Rothman, J., & Westergaard, M. (2020). Syntact contrasts in early and late Brazilian Portuguese-European Portuguese bidialectal bilinguals: Data from production. In K. V. Molsing, C. B. Perna & A. M. Ibaños (Eds.), *Linguistic approaches to Portuguese as an additional language* (pp. 35–59). J. Benjamins. https://doi.org/10.1075/ihll.24.02cas

Chambers, J. K. (1992). Dialect acquisition. *Language*, *68*, 673–705. https://doi.org/10.2307/416850

Chevrot, J., & Ghimenton, A. (2018). Bilingualism and Bidialectalism. In A. de Houwer & L. Ortega (Eds.), *The Cambridge handbook of bilingualism* (pp. 510–523). Cambridge University Press. https://doi.org/10.1017/9781316831922.026

Christensen, R. (2023). *ordinal - Regression Models for Ordinal Data* (R package version 2023.12-4.1). Obtido em https://CRAN.R-project.org/package=ordinal

Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*, *42*(3), 396–420. https://doi.org/10.1111/hcre.12079

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*, 199–210. https://doi.org/10.1016/j.jml.2010.12.004

Edwards, J. (1999). Refining our understanding of language attitudes. Journal of Language and Social Psychology, 18, 101–110. https://doi.org/10.1177/0261927X99018001007

Eisenstein, M. (1983). Native reactions to non-native speech: A review of empirical research. *Studies in Second Language Acquisition*, *5*, 160–176.

Ekberg, L., & Östman, J. (2020). Identity construction and dialect acquisition among immigrants in rural areas - The case of Swedish-language Finland. *Journal of Multilingual and Multicultural Development*, 1–14. https://doi.org/10.1080/01434632.2020.1722681

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Foreman, A. (2003). *Pretending to be someone you're not: A study of second dialect acquisition in Australia* [Tese de Doutoramento, Universidade de Monash]. https://doi.org/10.4225/03/59c9f2bfcdc9b

Frota, S., & de Moraes, J. A. (2016). Intonation in European and Brazilian Portuguese. In W. L. Wetzels, J. Costa & S. Menuzzi (Eds.), *The handbook of Portuguese linguistics* (pp. 141–166). John Wiley. https://doi.org/10.1002/9781118791844.ch9

Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions. The European/Brazilian Portuguese case. *Probus*, *13*, 247–273.

Frota, S., Cruz, M., Svartman, F., Collischonn, G., Fonseca, A., Serra, C., Oliveira, P., & Vigário, M. (2015). Intonational variation in Portuguese: European and Brazilian varieties. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 235–283). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199685332.003.0007

Giles, H., & Smith, P. M. (1979). Accomodation theory: Optimal levels of convergence. In H. Giles & R. St Clair (Eds.), *Language and social psychology* (pp. 45–65). Blackwell.

Giles, H., Taylor, D., & Bourhis, R. (1973). Toward a theory of interpersonal accomodation through speech: some Canadian data. *Language in Society*, *2*, 177–192.
Holmes, J., & Stubbe, M. (2003). *Power and politeness in the workplace: A sociolinguistic analysis of talk at work*. Pearson.
Jacobs-Huey, L. (1997). Is there an authentic African American speech community? Carla revisited. *University of Pennsylvania Working Papers in Linguistics*, *4*, 331–370.
Kupisch, T., Castro, T., Krämer, M., & Westergaard, M. (2023). Phonological influence in bilectal speakers of Brazilian and European Portuguese. *International Journal of Bilingualism*, 1–18. https://doi.org/doi.org/10.1177/13670069231156341
Lecky, P. (1961). *Self-consistency, A theory of personality*. Shoe String Press.
Lee, T. L., & Fiske, S. T. (2006). Not an ougroup, not yet an ingroup: Immigrants in the Stereotype Content Model. *International Journal of Intercultural Relations*, *30*, 751–768. https://doi.org/10.1016/j.ijintrel.2006.06.005
Lennenberg, E. (1967). *Biological foundations of language*. John Wiley.
Lev-Ari, S., & Keysar, B. (2010). Why don't we believe in non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/doi.org/10.1016/j.jesp.2010.05.025
Lønes, E. H., Kamide, Y., & Melinger, A. (2023). *Dialect-specific grammatical features: Evidence from abstract priming across the lexicon* [Manuscrito não publicado]. Universidade de Dundee.
Long, M. H. (2007). *Problems in SLA*. Lawrence Erlbaum Associates Publishers.
Lorenzoni, A., Faccio, R., & Navarrete, E. (2024). Does foreign-accented speech affect credibility? Evidence from the illusory-truth paradigm. *Journal of Cognition*, *7*(1), 26. https://doi.org/doi.org/10.5334/joc.353
Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139
Mateus, M. H., & d'Andrade, E. (2000). *The phonology of Portuguese*. Oxford University Press.
Nycz, J. (2015). Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass*, *9*(11), 469–482. https://doi.org/10.1111/lnc3.12163
Oliveira, B. D. (2022). *Quando se fala "brasileiro": o preconceito linguístico sofrido por brasileiros que vivem em Portugal* [Trabalho de Conclusão de Curso, Universidade Federal da Paraíba]. Repositório Institucional da Universidade Federal da Paraíba. https://repositorio.ufpb.br/jspui/handle/123456789/23177
Payne, A. C. (1976). *The acquisition of the phonological system of a second dialect* [Tese de Doutoramento, Universidade da Pensilvânia]. ProQuest Dissertations and Theses. https://www.proquest.com/openview/6fe5af23f08899283da912501196be0b/1?pq origsite=gscholar&cbl=18750&diss=y
Payne, A. C. (1980). Factors controlling the acquisition of the Phildelphia dialect by out-of-state children. In W. Labov (Ed.), *Locating language in time and space* (pp. 143–178). Academic Press.
Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215. https://doi.org/10.006/jpho.2001.0134
Pereira, R. (2025). *Aquisição de segundo dialeto e erosão linguística: Os pronomes objeto nas gramáticas dos imigrantes adultos nativos do português brasileiro em Portugal* [manuscrito em preparação]. Universidade NOVA de Lisboa.
Posit Team (2024). *RStudio: Integraded Development Environment for R*. Posit Software, PBC.
Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in speech. *Cognition*, *73*, 265–292. https://doi.org/10.1016/S0010-0277(00)00101-3
Rennicke, I. E. (2015). *Variation and change in the rhotics of Brazilian Portuguese* [Tese de Doutoramento, Universidade Federal de Minas Gerais]. Repositório Institucional da UFMG. http://hdl.handle.net/1843/MGSS-A8ZH8U

Rickford, J. R., & Rickford, R. J. (2000). *Spoken soul: The story of Black English*. John Wiley.
Rodrigues, C. (2024). Sound variation in European Portuguese. In A. Zampaulo (Ed.), *The Routledge handbook of Portuguese phonology* (pp. 201–221). Routledge. https://doi.org/10.4324/9781003266648-14
Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition*, *5*, 148–159.
Siegel, J. (2010). Second Dialect Acquisition. Cambridge University Press. https://doi.org/10.1017/CBO9780511777820
Sobral, J. M. (2012). *Portugal, portugueses: Uma identidade nacional*. Fundação Francisco Manuel dos Santos.
Southwood, M. H., & Flege, J. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics and Phonetics*, *13*(5), 335–349. https://doi.org/10.1080/026992099299013
Stanford, J. N. (2008). A sociotonetic analysis of Sui dialect contact. *Language Variation and Change*, *20*, 1–42. https://doi.org/10.1017/S0954394508000161
Trudgill, P. (1986). *Dialects in contact*. Blackwell.
Vigário, M. (2016). Segmental phenomena and their interactions. Evidence for prosodic organization and the architecture of grammar. In S. Fischer & C. Gabriel (Eds.), *Manual of grammatical interfaces in Romance* (pp. 41–73). De Gruyter. https://doi.org/10.1515/9783110311860-005
Vigário, M. (2022). Portuguese. In C. Gabriel, R. Gess & T. Meisenburg (Eds.), *Manual of Romance phonetics and phonology. Part II - Phonetics and phonology of Romance languages* (pp. 839–881). De Gruyter. https://doi.org/10.1515/9783110550283-027

ELBT e desenvolvimento da competência escrita em manuais didáticos de PL2

TBLT and the development of writing competence in PL2 textbooks

Juliano Sippel[1]

[1] Centro de Linguística da Universidade NOVA de Lisboa

Abstract

In this study, we investigated the use of Task-Based Language Teaching (TBLT) in teaching materials developed for teaching text production in Portuguese L2 (PL2). We previously analysed PL2 textbooks and interviewed teachers, highlighting limitations in existing resources and the need for a procedural approach to writing. Based on ELBT, we developed activities involving planning, producing, and revising texts. The resources were used in a writing workshop with PL2 level B1 university students and the results were promising, including better use of lexicon and textual progression and suitability for the genre and target audience. Data collection will continue with new groups to broaden the conclusions about the effectiveness of the materials developed.

Keywords: Task-Based Language Teaching (TBLT), writing competence, textbook, Portuguese L2 (PL2).

Resumo

Neste estudo investigamos o uso do Ensino de Língua Baseado em Tarefas (ELBT) em materiais didáticos desenvolvidos para o ensino da produção textual em português L2 (PL2). Analisamos previamente manuais de PL2 e entrevistamos docentes, destacando limitações nos recursos existentes e a necessidade de abordagem processual de escrita. Com base no ELBT, foram desenvolvidas atividades envolvendo planificação, produção e revisão textual. Os recursos foram utilizados em uma oficina de escrita com estudantes universitários de PL2 nível B1 e os resultados foram promissores, incluindo melhor uso de léxico, progressão textual e adequação ao gênero e ao público-alvo. A coleta de dados continuará com novos grupos para ampliar as conclusões sobre a eficácia dos materiais desenvolvidos.

Palavras-chave: Ensino de Línguas Baseado em Tarefas (ELBT), competência escrita, manual didático, Português L2 (PL2).

1. Introdução

Por considerar que o ensino da competência escrita em L2 exige instruções explícitas (Cumming, 2001; Leki et al., 2008; Rinnert & Kobayashi, 2009) e que os manuais didáticos desempenham um papel essencial no desenvolvimento das habilidades comunicativas (Tomlinson & Masuhara, 2018), testamos a aplicação do Ensino de Língua Baseado em Tarefas (ELBT) (Byrnes & Manchón, 2014; Castro, 2017, 2020; Ellis, 2003, 2009; Long, 2015; Willis & Willis, 2007) em materiais voltados para o ensino da produção textual em português L2 (PL2).

Partindo da ideia de que a aprendizagem de L2 está focada na resolução de problemas relacionados ao mundo real (Skehan, 1996) e de que as tarefas de escrita são atos de construção de sentido voltados para a

aquisição das habilidades necessárias ao engajamento comunicativo nas esferas pessoais e profissionais (Byrnes & Manchón, 2014), optamos pelo ELBT, porque permite reflexões durante a produção de textos e estimula o uso real do PL2 (Pinto, 2020). Além disso, por abranger aspectos da língua, do contexto social e dos processos mentais dos estudantes (Van den Branden *et al*, 2009), acreditamos que essa abordagem pode ser integrada ao modelo de escrita como processo (Hyland, 2009), o qual tem sido considerado adequado (Graham, 2018) para uma reestruturação do ensino, que inclui não apenas as formas textuais, mas também a atenção aos processos de composição dos estudantes.

O que apresentamos neste artigo são os resultados parciais de uma investigação mais ampla, desenvolvida em diferentes fases. Na etapa inicial, que foi realizada anteriormente (Sippel, 2023), analisamos manuais didáticos de PL2 e constatamos que fazem uso da versão fraca da abordagem comunicativa (Richards & Rodgers, 2001) e adotam um modelo de ensino de escrita centrado no texto (*text-oriented research and teaching*) (Hyland, 2009).

Cientes de que manuais e recursos didáticos são utilizados de diferentes formas (devido a fatores como planificação curricular dos cursos de L2, formação dos docentes, necessidades específicas dos estudantes e preferências pessoais), prosseguimos com a coleta de dados, que envolveu entrevistas semiestruturadas. Durante esse processo, entrevistamos cinco docentes de PL2 da NOVA-FCSH e dois da Universidade Federal Tecnológica do Paraná (Brasil), com o objetivo de compreender como utilizam os manuais didáticos em suas aulas e quais lacunas identificam no ensino da competência escrita.

Das conversas realizadas, destacamos as observações recorrentes e as que consideramos mais relevantes, a saber: (i) necessidade de fornecimento de instrução para a produção escrita; (ii) relevância do trabalho com a planificação da produção textual em sala de aula; (iii) importância do incentivo e do fornecimento de instruções para realização de revisões textuais; (iv) oferta de *inputs* estruturados e adequados aos níveis e aos diferentes perfis socioculturais presentes na sala de aula; e (v) ensino de padrões textuais (formas típicas dos gêneros, léxico especializado e adequação pragmática aos diferentes propósitos) dos gêneros trabalhados.

Neste estudo, apresentaremos alguns efeitos do uso dos materiais elaborados com os pressupostos do ELBT (com base na análise dos manuais e das entrevistas realizadas), aplicados em uma oficina de escrita com estudantes universitários de PL2 no nível B1. O artigo está estruturado da seguinte forma: na seção 2, oferecemos uma visão breve da literatura sobre o desenvolvimento da competência escrita em L2 e sobre o ELBT; a seção 3 apresenta a caracterização do estudo, incluindo a questão de investigação e a metodologia adotada; os resultados são descritos e discutidos na seção 4; na seção 5, se apresentam as conclusões e os próximos passos da investigação.

2. Enquadramento teórico

2.1 Desenvolvimento da competência escrita em L2

Comparada às demais competências comunicativas, a escrita é uma habilidade linguística mais complexa. Em termos gramaticais e em contraste à oralidade,[1] um texto bem redigido exige correção ortográfica, uso de orações completas e ordem de palavras mais rígida, além de alguma densidade lexical com utilização constante de sinônimos (Cassany, 1999, 2017). O processo de escrita envolve a seleção e organização de expressões linguísticas para representar conhecimentos de forma coerente e coesa, atuando em dois níveis: macroestrutural (organização geral) e microestrutural (combinação de expressões) (Barbeiro & Pereira, 2007). O escritor deve ativar conhecimentos prévios, enquadrar o texto em um gênero, redigir e revisar, integrando os processos de planificação, textualização e revisão (Grabe & Kaplan, 1996) de forma interativa.

[1] Mosterín (2002) traça um paralelo interessante entre *fala* e *escrita* e diz que "El habla se produce en el tiempo y desaparece tan pronto como se ha producido, no viaja en el tiempo. El habla reflejada en el espacio, fijada y hecha visible en el espacio, es la escritura" (p. 22).

A tradição do ensino de escrita, tanto em L1 quanto em L2, pode ser compreendida a partir de três modelos principais, conforme apontam Cumming (2001) e Hyland (2003, 2009): centrados no texto, no escritor e no leitor. No modelo orientado ao texto (*text-oriented*), o foco está na estrutura linguística e na precisão gramatical, partindo do princípio de que quanto maior o domínio léxico-gramatical, melhor a produção escrita. Já o modelo centrado no escritor (*writer-oriented*) entende a escrita como um processo dinâmico e não linear, que envolve planificação, textualização e revisão contínua. Modelos de Hayes e Flower (1980, 1981), Hayes (1996) e Scardamalia e Bereiter (1987) são referências importantes nessa abordagem e investigações sobre processos mentais mostram que alunos mais proficientes dedicam mais tempo à planificação e à revisão, demonstrando maior controle textual. Por fim, o modelo orientado ao leitor (*reader-oriented*) trata a escrita como uma prática interativa e social, que leva em conta aspectos políticos e ideológicos, enfatizando o ensino de gêneros textuais para inserir os alunos em comunidades discursivas e ajudá-los a negociar significados e relações de poder.

Graham (2018) destaca que a escrita tem sido analisada pelas perspectivas comportamentais, cognitivas e sociais. Atualmente, predomina uma visão que combina bases cognitivas - envolvendo mecanismos complexos de escrita - e sociais, através da pedagogia do gênero, que concebe a escrita como uma prática interativa. Por ser um conhecimento explícito,[2] o domínio da competência escrita requer instrução explícita, e seu ensino (também explícito) tem sido apontado (Bangert-Drowns et al., 2004; Graham, 2018; Graham & Herbert, 2011; Leki et al., 2008; Rinnert & Kobayashi, 2009) como critério de êxito na aprendizagem.

Ainda que Tomlinson (2010a, 2010b, 2011, 2016) e Pinnard (2016) não foquem especificamente na escrita em L2, sugerem práticas úteis para o ensino dessa competência. Recomendam que os alunos leiam textos envolventes, explorem formas linguísticas, identifiquem padrões, façam generalizações e produzam novos textos, inspirados nos modelos fornecidos.

No âmbito do ensino da escrita em PL2, Lopes e Pinto (2022) mostram que a densidade de ideias na escrita melhora com instruções adequadas; Pinto (2020) destaca que uma abordagem baseada em tarefas ajuda no desenvolvimento da escrita por estimular usos reais da língua; Parracho (2011) revela que intervenções pedagógicas focadas no processo de escrita beneficiam crianças não nativas de português. Destacamos, ainda, a investigação de Barbosa e Bizarro (2015), que defendem uma *pedagogia da escrita em L2*, ressaltando a necessidade do desenvolvimento de manuais que expliquem processos de escrita e suas relações culturais.

2.2 **ELBT e desenvolvimento da competência escrita em L2**

O ELBT se apresenta como uma extensão da abordagem comunicativa,[3] mas com tarefas como unidade central do ensino da L2 (Castro, 2020; Long, 1991, 1998; Richards, 2006), mantendo a comunicação como objetivo central e concentrando-se em aspectos linguísticos à medida que surgem durante a interação.

As tarefas são projetos de trabalho que requerem processamento pragmático para assegurar comunicação eficaz. Elas demandam atenção ao significado e ao uso autêntico da língua, ativando processos cognitivos como seleção, ordenação e avaliação de informações. Com objetivos e resultados definidos, as tarefas integram as quatro competências comunicativas da L2 na resolução de problemas (Ellis, 2003; Prabhu, 1987).

Segundo Nunan (2004), alguns princípios do ELBT incluem: (i) fornecimento de modelos de apoio aos estudantes; (ii) construção de tarefas progressivas (com uma história pedagógica); (iii) estímulo à aprendizagem

[2] Conhecimento explícito e implícito diferenciam-se sobretudo em torno da questão da consciência: não temos consciência sobre o que aprendemos de forma implícita, ao passo que a aprendizagem explícita exige atenção consciente, memorização e testagem de hipóteses (Rebuschat, 2015).

[3] A abordagem comunicativa apresenta duas versões principais: a *fraca* (ou moderada), que propõe o ensino isolado dos componentes das competências comunicativas, criando oportunidades para o uso da língua em atividades com fins pedagógicos – como no modelo *Presentation, Practice, Production* (Howatt, 1984); e a *forte*, que defende a aquisição da L2 exclusivamente por meio da comunicação, com foco na interação e na descoberta de regras linguísticas durante o uso real da língua – conforme o modelo *Focus on Form* (Long, 1991, 1998). Na versão fraca, parte-se da ideia de *aprender para usar*; já na forte, de *usar para aprender* (Richards & Rodgers, 2001).

ativa; (iv) integração de formas gramaticais a funções comunicativas; (v) uso de tarefas reprodutivas para exercitar formas, sentidos e funções; e (vi) inclusão de *feedback* e reflexão.

O ELBT não exclui o trabalho com formas linguísticas, já que focar apenas no significado não garante a produção adequada de estruturas gramaticais, uma vez que a análise formal ajuda na reestruturação da interlíngua, destacando estruturas necessárias às tarefas (Skehan, 1996). Como consequência, essa abordagem deve ser vista como complementar às centradas na forma, criando contextos que estimulem a aprendizagem natural da L2 em vez de sistematizar a língua gradualmente.

Mais recentemente, Ellis (2009) afirma que uma tarefa deve atender a quatro critérios[4]: (i) foco no significado, promovendo processamento semântico e pragmático; (ii) presença de uma lacuna que exija expressar opiniões ou transmitir informações; (iii) uso de recursos linguísticos e não linguísticos próprios dos estudantes; e (iv) um resultado definido além do uso da língua.

Com esses critérios é possível diferenciar *tarefa de exercício gramatical contextualizado* de *tarefa com* foco – outra distinção feita por Ellis (2003, 2009). Segundo o autor, as tarefas podem ser *com* ou *sem* foco, sendo estas as projetadas para oferecer ao estudante possibilidades de usar a L2 de forma livre e, aquelas, as que demandam o uso de alguma estrutura gramatical específica, que deve ser oculta (sem que se diga aos estudantes que recurso específico necessitam). Dessa distinção, surge outra, igualmente importante, que é a oposição do ensino de línguas *baseado* em tarefas e o ensino de língua *com o apoio* de tarefas: o primeiro prioriza tarefas sem foco, enquanto o segundo segue o modelo *Presentation, Practice, Production* (PPP) com exercícios gramaticais finais (Castro, 2017).

Porque a aprendizagem de L2 promovida pelo ELBT está ligada à solução de problemas, adotamos essa abordagem para desenvolver materiais voltados ao ensino da competência escrita. As tarefas textuais expandem eventos comunicativos, exigindo resolução de problemas em contextos reais. Ademais, o ELBT pode ser uma alternativa promissora associada ao modelo de ensino de escrita como processo (Bygate et al., 2011; Manchon, 2014, Robinson, 2011; Van den Branden *et al*, 2009). Os modelos cognitivos de escrita (Bereiter & Scardamalia 1987; Flower & Hayes 1980, 1981; Hayes 1996) concebem a natureza da resolução de problemas relacionada à capacidade do escritor de atender a objetivos de forma simultânea, gerindo a alternância entre os componentes do processo de escrita (planificar, escrever e revisar). Tais modelos mostram também que a representação da tarefa, a definição de objetivos e a resolução do problema emergem da interação entre o escritor e a tarefa de escrever, reforçando a conexão com o ELBT.

No contexto de ensino de gêneros textuais,[5] uma abordagem articulada ao ELBT favorece a prática situada da linguagem em contextos comunicativos significativos. As tarefas de escrita baseadas em gêneros oferecem oportunidades autênticas para o uso de recursos léxico-gramaticais recorrentes, encenando situações reais que motivam os alunos a construir sentido de forma contextualizada. Além disso, possibilitam o monitoramento e a reflexão sobre o uso de elementos linguísticos típicos de diferentes gêneros, promovendo maior consciência sobre a relação entre forma (os recursos linguísticos disponíveis) e função (os contextos de usos discursivo da língua) (Byrnes, 2014).

Byrnes e Manchon (2014) listam mais alguns fatores que atestam o alinhamento do desenvolvimento da competência escrita ao ELBT, dentre os quais destacamos a natureza idiossincrática da atividade de resolução de problema que caracteriza a produção de texto - que implica um processamento linguístico mais expansivo e

[4] Skehan (1998) propôs anteriormente cinco critérios: (i) o foco principal é o sentido; (ii) os estudantes não devem reproduzir discursos de terceiros; (iii) as tarefas devem ter relação com o mundo real; (iv) a realização e a conclusão da tarefa devem ser aspectos prioritários; e (v) a tarefa deve ser avaliada com base nos resultados obtidos. Consideramos a proposta da Ellis (2003, 2009) mais interessante pela exigência da definição prévia de um resultado esperado.

[5] Dentre as diversas teorizações sobre gênero textual, destaca-se seu papel central na mediação entre linguagem e cultura, pois permitem, por meio da produção textual, o acesso, a interpretação e a apropriação de conteúdos culturais das línguas. Gêneros podem ser definidos como formas retóricas dinâmicas e situadas culturalmente, que articulam forma e conteúdo, constroem estruturas sociais e delimitam comunidades discursivas (Berkenkotter & Huckin, 1995). Para Miller (2012), são artefatos culturais por meio dos quais se realizam ações sociais e comunicativas. No contexto acadêmico, configuram-se como eventos comunicativos com propósitos compartilhados, reconhecidos pelas comunidades discursivas às quais pertencem (Swales, 1990, 1998).

profundo do que é possível na comunicação oral. Os autores sugerem a importância de se investigar as relações entre conhecimento linguístico e comportamento estratégico acionado pelos estudantes, consoantes às metas que estabelecem.

Atendendo a essa sugestão, Macaro (2014) mostra que as tarefas de escrita exigem que os escritores considerem o conhecimento de L2 que possuem, o conhecimento do gênero da escrita e a disponibilidade de fontes de apoio externas como dicionários, glossários etc., e que o nível de comportamento estratégico adotado pode afetar o resultado da produção, pois o grau de ativação de conhecimento linguístico resulta em diferentes comportamentos estratégicos. Dito de outra forma, o nível de ativação linguística (léxico, morfossintático e pragmático) afeta o desempenho e deve ser considerado no desenvolvimento de tarefas.

Embora investigações experimentais nem sempre sejam viáveis em sala de aula ou em manuais didáticos, estudos qualitativos podem fornecer subsídios relevantes para o desenvolvimento de materiais - esse é o objetivo central deste trabalho.

3. Caracterização do estudo

Nosso estudo se fundamenta na investigação qualitativa (Minayo, 2012), o que nos possibilitou compreender os processos relacionados à problemática investigada, com foco na análise das perspectivas e experiências dos sujeitos envolvidos. Adotamos a estratégia do estudo de caso, valendo-nos de observações, entrevistas e análise de produções realizadas, com o objetivo de identificar e compreender os fatores que influenciam as práticas pedagógicas e seus efeitos no aprendizado dos estudantes.

Tendo em consideração a análise de manuais didáticos de PL2 e a entrevista com docentes de PL2, previamente realizadas, formulamos a questão de investigação a que procuraremos responder neste estudo: *O ELBT pode ser eficaz para o desenvolvimento da competência escrita em PL2?*

Para responder à questão, construímos materiais, utilizando como abordagem pedagógica o ELBT, e os testamos em uma oficina de escrita com quatro estudantes universitários (doravante identificados com as letras A a D) que recentemente haviam terminado o curso de PL2 nível B1, na NOVA-FCSH. A oficina foi ministrada em cinco encontros presenciais, com duas horas de duração, totalizando dez horas letivas. Os gêneros selecionados para o trabalho com tarefas foram: sinopse de filmes; narrativa de histórias em imagem; e reclamações formais.

Organizamos o trabalho com as tarefas, para cada gênero, em três fases (Ellis, 2009; Willis & Willis, 1996), conforme descrevemos na Tabela 1.

Tabela 1

Desenvolvimento de Atividades para Produção Textual

Pré-tarefa	Planificação da produção textual com tarefas de levantamento de léxico especializado, leitura crítica de *inputs* e reconhecimento de estruturas recorrentes no gênero.
Tarefa	Tarefas de produção escrita (individual e/ou em pares).
Pós-tarefa	Tarefas com foco na forma para promoção da consciencialização linguística e revisão da produção textual realizada.

As etapas de desenvolvimento do gênero *sinopse de filmes* foram as seguintes:

(i) Pré-tarefa: solicitamos a produção de uma sinopse sem o fornecimento de instruções, na sequência, realizamos a leitura de diferentes sinopses de filmes com o objetivo de identificar os elementos macro do gênero (propósito, público-alvo e suporte) e os elementos micro (aspectos visuais, informações técnicas das obras, vocabulário e estruturas recorrentes). Promovemos a ativação do léxico por meio

de tarefas como a adequação de adjetivos a diferentes gêneros cinematográficos e fornecemos instruções detalhadas para a escrita de sinopses.

(ii) Tarefa: assistimos a pequenos filmes de um minuto[6] e produzimos sinopses dessas obras, tendo como foco o treinamento do resumo de enredos e a identificação das informações técnicas essenciais para o gênero.

(iii) Pós-tarefa: promovemos atividades *Focus on Form* (FonF), como o reconhecimento e a classificação de conjunções coordenativas presentes em uma sinopse. Revisamos o texto inicialmente escrito com base nos critérios descritos na Tabela 2.

Tabela 2

Grelha para Revisão de Sinopse

	Pontos de atenção	Sim/Não	Comentários
Adequação	A sinopse cumpre as funções comunicativas básicas		
	Identifica os personagens principais		
	Considera tempo e espaço		
	Apresenta um breve resumo do filme		
	Faz menção ao conflito do enredo		
	A linguagem é adequada ao público-alvo		
	Utiliza adjetivos que valorizam a história		
	Os tempos verbais utilizados são adequados		
Coerência	O texto apresenta uma estrutura clara e ordenada, sem repetição de informações		
	Identifica-se no texto introdução, desenvolvimento e conclusão		
	As orações estão inter-relacionadas com pontuação, pronomes, conjunções		
	Os elementos de coesão são facilmente recuperados e guiam a leitura		
Correção	As palavras estão escritas de forma correta, sem erros de ortografia		
	O vocabulário é variado, sem repetição de termos		

As fases de desenvolvimento do gênero *narrativa de histórias em imagem* foram as seguintes:

i. Pré-tarefa: solicitamos a escrita da transcrição em texto de uma imagem, sem o fornecimento de instruções, fizemos a leitura (em conjunto e individualmente) de diferentes narrativas em imagens[7] com o objetivo de identificar os elementos macro (propósito, público-alvo e suporte) e micro (estrutura gráfica, personagens, cenários, mensagens e metáforas), e fornecemos instruções detalhadas para a produção das transcrições.

[6] *One Minute Short Film*, ou microfilmes, são produções cinematográficas de até 60 segundos. Assistimos a filmes vencedores de diferentes competições como *The Gotta Minute Film Festival*, Canadá (2023), *Vesuvius International Film Fest*, Itália (2023), e *Film Festival*, Paquistão (2023).

[7] Utilizamos muitas das narrativas visuais contidas no *Bologna Children's Book Fair "museum*", uma espécie de museu virtual que exibe livros e ilustrações expostos na *BolognaFiere*. As imagens podem ser consultadas em https://galleries.bolognachildrensbookfair.com.

ii. Tarefa: reescrevemos coletivamente um texto mal produzido, tendo em conta os critérios apresentados e discutidos na etapa anterior.

iii. Pós-tarefa: por termos optado por exercitar a escrita desse gênero com foco sobretudo no sentido, não fizemos tarefas FonF, somente revisamos os primeiros textos produzidos com base nos critérios da Tabela 3.[8]

Tabela 3

Grelha para Revisão de Transcrição de Narrativa em Imagem

	Pontos de atenção	Sim/Não	Comentários
Adequação	O texto cumpre as funções comunicativas básicas		
	Descreve os personagens e suas expressões		
	Informa o(s) cenário(s) onde se passa a história		
	Identifica a presença de humor ou crítica social		
	Faz uma correta interpretação de símbolos e metáforas presentes na história		
	O texto não possui marcas de oralidade		
	Os tempos verbais utilizados são adequados		

O gênero *reclamação formal* foi desenvolvido de acordo com as seguintes etapas:

i. Pré-tarefa: solicitamos a escrita de uma reclamação, sem fornecer instruções, realizamos a leitura, em pares e individualmente, de diferentes exemplos de reclamações formais[9], com o objetivo de identificar os elementos macro (propósito, público-alvo e suporte) e micro (título, identificação do problema, local e data, tentativas de resolução, pedido de verificação, prazo de resposta etc.), ativamos léxico, explorando estruturas típicas usadas em reclamações relacionadas a serviços, produtos ou experiências e fornecemos instruções detalhadas para guiar a produção desse gênero.

ii. Tarefa: apresentamos contextos de situações problemáticas e solicitamos a escrita de uma reclamação formal, orientando os alunos a considerarem os critérios discutidos na etapa anterior.

iii. Pós-tarefa: desenvolvemos atividades FonF, como a identificação de marcas de formalidade e expressões de cortesia características de uma reclamação formal. Revisamos as primeiras versões produzidas pelos alunos, com base nos critérios da Tabela 4.

[8] Reproduzimos apenas a coluna *Adequação* porque as colunas *Coerência* e *Correção* possuem os mesmos critérios da Tabela 3, pelo que não as iremos incluir nas próximas grelhas de revisão.
[9] Usamos textos autênticos e atuais submetidos no *Portal da Queixa*, disponível em https://portaldaqueixa.com.

Tabela 4

Grelha para Revisão de Reclamação

	Pontos de atenção	Sim/Não	Comentários
Adequação	O texto cumpre as funções comunicativas básicas		
	Contém um título claro e conciso, que indica o motivo da reclamação		
	Faz uma descrição detalhada do problema, incluindo data, local, como o problema/situação que gerou a reclamação foi descoberto		
	Informa a(s) tentativa(s) prévia(s) de resolução tomada(s)		
	Solicita resolução e prazo de resposta		
	Contém uma despedida formal e cordial		
	Mantém-se ao longo do texto o registo formal (saudação, corpo do texto e despedida)		
	O texto não possui marcas de oralidade		
	Os pronomes de tratamento são adequados (*você/senhor*)		

Conforme descrevemos nas etapas anteriores, antes do início das atividades propostas, solicitamos aos estudantes que produzissem pequenos textos para cada um dos gêneros trabalhados na oficina, sem que recebessem orientações prévias. Esse procedimento teve como objetivo avaliar seus conhecimentos gramaticais e suas habilidades de produção textual de forma espontânea.

Paralelamente, aplicamos um questionário, por meio do *Google Forms*, com o intuito de investigar a experiência prévia dos participantes com a escrita, tanto em sua L1 quanto em PL2, além de identificar os comportamentos que costumam adotar durante o processo de escrita. O instrumento continha afirmações relacionadas às atitudes diante da escrita e à importância atribuída à produção textual em PL2, as quais os alunos deveriam avaliar segundo seu grau de concordância.

Adicionalmente, para compreender e analisar as estratégias adotadas pelos participantes durante a produção textual, realizamos entrevistas semiestruturadas ao final da oficina. As perguntas abrangeram aspectos da planificação (como a formulação de ideias em L1 ou em português e o uso de esquemas prévios ou listas de palavras), da revisão (se os participantes revisam o texto durante ou após a escrita, e se retomam as ideias inicialmente planejadas) e do conhecimento linguístico (como foi a identificação de lacunas lexicais e/ou gramaticais ao longo do processo de escrita).

4. Resultados e discussão

4.1. Inquérito com estudantes

As informações a seguir referem-se aos dados fornecidos pelos participantes por meio do inquérito a que responderam antes de iniciarem a oficina.

No que diz respeito à escrita em PL2, 100% dos estudantes afirmaram ter recebido instrução em cursos de língua. Contudo, apenas 50% declararam ter realizado atividades de produção textual nesses cursos, e 25% disseram ter recebido instrução na universidade. Nenhum dos participantes utilizou manuais de ensino com propostas específicas para a produção escrita.

Quanto à escrita em L1, todos relataram ter recebido instrução na escola e/ou na universidade, elaborado trabalhos escolares e acadêmicos, e afirmaram conhecer as normas gramaticais e ortográficas. Além disso, 75% utilizaram manuais didáticos que ensinavam a escrever.

Sobre os comportamentos adotados ao escrever em PL2, 100% revisam o texto após concluí-lo; 75% fazem revisões durante o processo de escrita; 50% utilizam ferramentas eletrônicas de correção e tradução, e escrevem na L1 para depois traduzir; e um participante cria listas temáticas de vocabulário e esquemas de planificação antes de escrever.

Ao serem questionados sobre opiniões relacionadas à escrita em português, 75% concordam que escrever nessa língua é importante para manter contato com colegas e fazer amizades, que escreveriam melhor com mais prática, e que ler textos em português contribui para a melhoria da escrita; 50% acreditam que escrever em português os ajudaria a progredir na vida profissional; 25% concordam que um bom vocabulário na língua portuguesa é suficiente para escrever bem e que os cursos que frequentaram deram à escrita a mesma importância atribuída à produção oral; nenhum dos participantes concorda que o domínio da gramática por si só é suficiente para escrever bem ou que a escrita se tornou irrelevante devido às ferramentas de tradução eletrônica.

Ao comparar as respostas referentes à escrita em L1 e PL2, nota-se que, embora todos os estudantes tenham recebido instrução em ambas as línguas, apenas metade relatou ter realizado atividades de produção textual nos cursos de PL2. Ademais, nenhum estudante declarou ter utilizado manuais didáticos voltados para a escrita em português. Esses resultados corroboram os dados de nossa análise de manuais, realizada na primeira etapa do estudo, que evidenciaram a pouca (ou menor) ênfase dada à competência escrita nos cursos de L2, como já apontado na literatura (Larios et al., 2007; Silva, 1990, 1993).

A importância atribuída à escrita pelos estudantes fica evidente quando se observa que nenhum deles acredita que ferramentas eletrônicas de revisão e tradução possam substituir seu trabalho com o texto, embora metade as utilize regularmente. Sobre o ensino dessa competência, destaca-se que a maioria concorda que a prática da escrita e a leitura de textos em português são mais eficazes para melhorar a produção textual do que apenas o domínio de vocabulário e gramática, que não são vistos como os fatores mais importantes.

4.2. **Análise de produções escritas**

Devido à opção por um enfoque qualitativo com estratégia de estudo de caso e às limitações de espaço, neste artigo reproduziremos e analisaremos uma amostra representativa das produções de cada participante da oficina de escrita, considerando tanto as versões iniciais quanto as revisadas. Os textos nomeados como *Versão 1* correspondem às produções realizadas pelos estudantes sem a oferta de instruções prévias e antes do início das atividades propostas na oficina, enquanto as *Versão 2* representam as produções finalizadas, elaboradas após a conclusão das etapas descritas na Tabela 1.

Embora tenhamos considerado a correção gramatical como um critério de revisão, este aspecto não será o foco de nossas análises; buscaremos avaliar, entre as diferentes versões: (i) a adequação ao gênero e ao público-alvo; (ii) o uso de léxico especializado e de estruturas típicas dos gêneros; e (iii) a progressão lógica.[10]

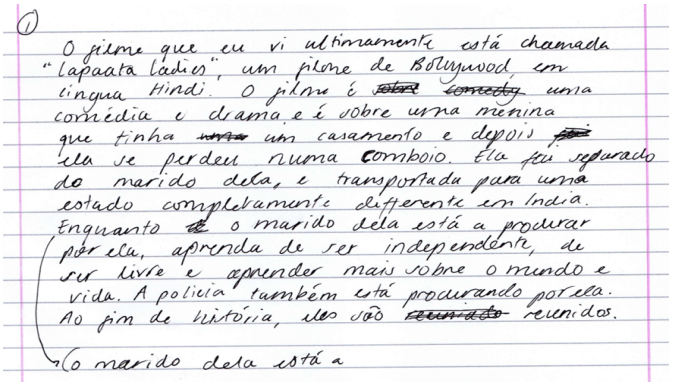
Além da análise das produções textuais dos participantes, apresentaremos comentários adicionais sobre o comportamento estratégico que adotaram durante as produções, com base nas informações recolhidas nas entrevistas realizadas ao final da oficina.

Iniciaremos nossa análise com o primeiro gênero textual trabalhado na oficina. A Figura 1 reproduz o texto produzido por A, solicitado com o seguinte comando: “Escreva uma sinopse de um filme”, sem o oferecimento de demais instruções.

[10] Por *progressão lógica*, referimo-nos ao encadeamento de segmentos textuais por meio de articuladores (ou marcadores discursivos), considerados em dois níveis: (i) organização global, que trata das relações de *coerência* entre sequências textuais; e (ii) organização intermediária, voltada à *coesão* entre parágrafos e sentenças (Koch, 2014).

Figura 1

Versão 1 da Sinopse de A



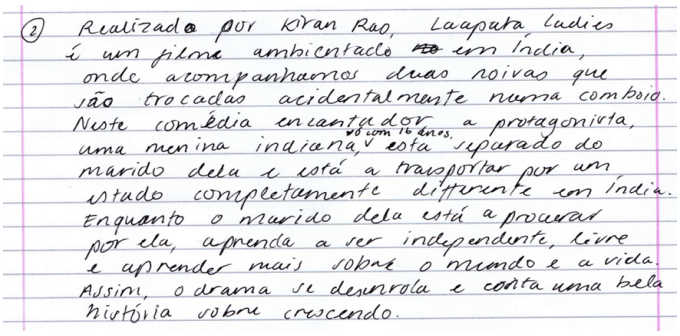
A sinopse apresenta a premissa principal do filme, o enredo central e o desfecho, cumprindo adequadamente a função de informar sobre o conteúdo da obra. A protagonista (a menina) e o seu marido são identificados como personagens centrais, e a presença da polícia como agente de apoio na busca ajuda a compor o quadro narrativo. Há referências temporais (antes e depois do casamento) e espaciais (um comboio, diferentes estados da Índia) que contribuem para a contextualização da história. No entanto, ao iniciar com a expressão pessoal "O filme que eu vi ultimamente está chamada" e ao revelar o desfecho com "Ao fim da história, eles são reunidos", o texto descaracteriza o gênero sinopse e compromete a adequação às suas convenções. Isso porque, nesse tipo de texto, não se espera a exposição do final da trama nem a inclusão de marcas explícitas de autoria (como o uso do pronome pessoal "eu").

Embora se reconheça uma estrutura básica - com introdução (ainda que breve), desenvolvimento e conclusão -, o desenvolvimento da narrativa ocorre de forma apressada, aproximando-se rapidamente do desfecho. Além disso, algumas frases mostram-se demasiado curtas ou desconectadas, como "A polícia também está procurando por ela", que poderia ser articulada com a informação anterior sobre a busca realizada pelo marido. A leitura seria mais fluida com o uso de pronomes e conectores que evitassem repetições (como "ela" e "marido dela") e garantissem maior coesão entre as partes do texto para melhorar sua progressão lógica.

Apresentamos na sequência a segunda versão do texto, após termos realizado as tarefas e a revisão, seguindo os parâmetros anteriormente descritos.

Figura 2

Versão 2 da Sinopse de A


② Realizado por Kiran Rao, Laapata Ladies é um filme ambientado em India, onde acompanhamos duas noivas que são trocadas acidentalmente numa comboio. Neste comédia encantador a protagonista, uma menina indiana, de com 16 anos, está separado do marido dela e está a transportar por um estado completamente different em India. Enquanto o marido dela está a procurar por ela, aprenda a ser independente, livre e aprender mais sobré o mundo e a vida. Assim, o drama se desenrola e conta uma bela histórria sobre crescendo.

A segunda versão apresenta maior fluidez e clareza em comparação à anterior. A organização do enredo está mais eficiente, evitando repetições desnecessárias e facilitando a compreensão do leitor. Destaca-se também pela introdução do título do filme, do nome do realizador e da ambientação logo no início, utilizando estruturas típicas do gênero sinopse (como a expressão "ambientado em"), que contribuem para situar o leitor de forma objetiva.

O enredo é exposto de maneira linear e bem estruturada, destacando eventos centrais como a troca acidental das noivas, a busca empreendida pelo marido e o processo de amadurecimento da protagonista. O uso de conectivos, como "onde" após a introdução do cenário, contribui para a coesão textual e a fluidez da narrativa.

A conclusão é adequada e envolvente, sintetizando a essência do filme com a frase "Assim, o drama se desenrola e conta uma bela história sobre crescendo", que contribui para manter o interesse do leitor.

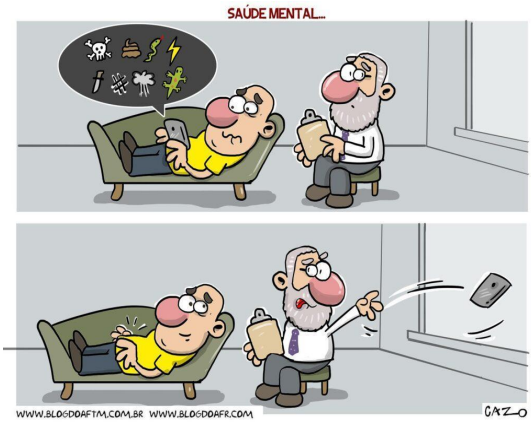
De modo geral, o texto se destaca por incorporar vocabulário mais específico (termos como "ambientado" e "protagonista") e por utilizar adjetivos que valorizam a obra (como "comédia encantadora"). Além disso, adota uma estrutura narrativa coerente com o gênero da sinopse, ao enfatizar a ambientação, os temas centrais e os elementos essenciais da trama, sem antecipar *spoilers*.

Em relação aos comportamentos adotados durante as produções, A inicialmente concebe ideias em sua L1, anota fragmentos dessas ideias na mesma língua e, conforme constrói as sentenças, realiza a transposição para o português. No início do processo, não dedica muita atenção ao vocabulário, preferindo planificar de forma mais ampla o que deseja expressar. Posteriormente, revisita seu esquema inicial, complementando-o com as palavras necessárias. Esse padrão tornou-se mais evidente na produção da Versão 2, o que destaca a relevância do trabalho de pré-tarefa, com ênfase na planificação da produção textual.

O segundo gênero textual que exercitamos foi a transcrição textual de narrativas em imagens. Iniciamos o trabalho da mesma forma que fizemos com as sinopses - solicitando a transcrição da história contada na Figura 3, sem o fornecimento de instruções.

Figura 3

Saúde Mental

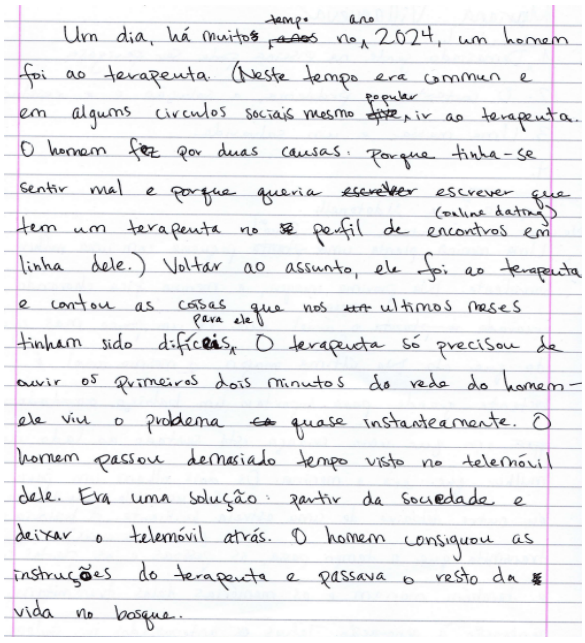


Nota. Fonte: Cazo (s.d.).

Reproduzimos na Figura 4 o texto produzido por B e, na sequência, faremos nossos comentários.

Figura 4

Versão 1 da Transcrição de Imagem de B



Um dia, há muitos ~~anos~~ tempos no ano 2024, um homem foi ao terapeuta. (Neste tempo era commun e em alguns circulos sociais mesmo popular ir ao terapeuta. O homem ~~foi~~ por duas causas. Porque tinha-se sentir mal e porque queria ~~escrever~~ escrever que tem um terapeuta no ~~se~~ perfil de encontros em linha (online dating) dele.) Voltar ao assunto, ele ~~foi~~ ao terapeuta e contou as coisas que nos ~~ultimos~~ ultimos meses para ele tinham sido difíceis. O terapeuta só precisou de ouvir os primeiros dois minutos do rede do homem - ele viu o problema ~~em~~ quase instantaneamente. O homem passou demasiado tempo visto no telemóvil dele. Era uma solução: partir da sociedade e deixar o telemóvil atrás. O homem consiguou as instruções do terapeuta e passava o resto da vida no bosque.

A Versão 1 da transcrição de *Saúde Mental* é uma tentativa de adaptar a mensagem do texto original com uma abordagem narrativa, para contextualizar o tema dentro de um cenário fictício (ano 2024) e criar uma conexão com o público contemporâneo (que usa aplicações de encontros *online*), o que o torna pouco objetivo na descrição dos elementos visuais e da mensagem central.

Embora mencione aspectos como terapia e dependência de *smartphones*, o texto introduz informações que não estão presentes na história original (como a referência a "perfis de encontros em linha" e a decisão de "passava o resto da vida no bosque"). Essas adições fazem que o texto produzido por B não cumpra o critério de adequação ao gênero.

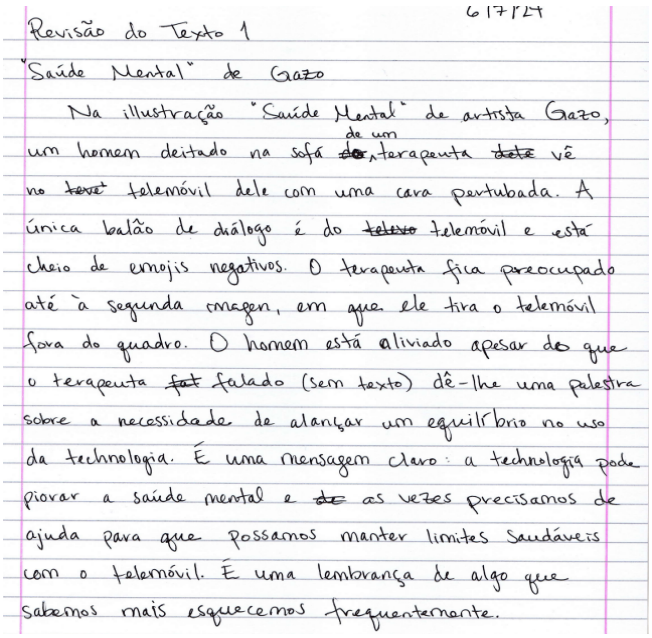
O critério de progressão lógica também é prejudicado, porque o texto demonstra falta de clareza, com interrupções desnecessárias e trechos repetitivos (como "voltar ao assunto, ele foi ao terapeuta", depois de um longo período entre parênteses). O uso de parênteses e termos genéricos (como "coisas") prejudica a fluidez e a precisão da narrativa. O desfecho é abrupto e desconectado da mensagem original, falhando em transmitir a crítica irônica de forma consistente.

A produção escrita poderia se beneficiar de uma linguagem mais técnica e estruturada, com foco na descrição objetiva dos elementos visuais e suas implicações. A omissão de detalhes visuais importantes e a inclusão de informações externas comprometem a fidelidade à obra original e dificultam a compreensão da mensagem crítica do texto original, o que também compromete o critério de uso de estruturas típicas do gênero.

A Figura 5 a seguir reproduz a segunda versão do texto, revisado depois de terem sido fornecidas as introduções sobre o gênero em causa.

Figura 5

Versão 2 da Transcrição de Imagem de B



6/7/24

Revisão do Texto 1

"Saúde Mental" de Gazo

Na ilustração "Saúde Mental" de artista Gazo, um homem deitado na sofá ~~do~~ de um terapeuta ~~dele~~ vê no ~~text~~ telemóvil dele com uma cara perturbada. A única balão de diálogo é do ~~telem~~ telemóvil e está cheio de emojis negativos. O terapeuta fica preocupado até à segunda imagem, em que ele tira o telemóvil fora do quadro. O homem está aliviado apesar do que o terapeuta ~~fat~~ falado (sem texto) dê-lhe uma palestra sobre a necessidade de alcançar um equilíbrio no uso da technologia. É uma mensagem clara: a technologia pode piorar a saúde mental e ~~às~~ as vezes precisamos de ajuda para que possamos manter limites saudáveis com o telemóvil. É uma lembrança de algo que sabemos mais esquecemos frequentemente.

A segunda versão mostra avanços importantes em relação à primeira, destacando-se por contextualizar a produção com o título do texto e o nome do autor, além de descrever os elementos visuais de forma mais precisa.

Agora, é possível compreender o que acontece na narrativa mesmo sem acesso ao texto-base, uma característica essencial das transcrições de narrativas visuais. Essa precisão na descrição dos elementos (como o balão de *emojis* negativos e a ação do terapeuta) demonstra maior fidelidade à mensagem original, sem a introdução de elementos externos ou fictícios, que não estão presentes nas ilustrações. A mensagem central é identificada e explicitada com precisão: "a technologia pode piorar a saúde mental, e as vezes precisamos de ajuda para manter limites saudáveis", refletindo o objetivo da narrativa.

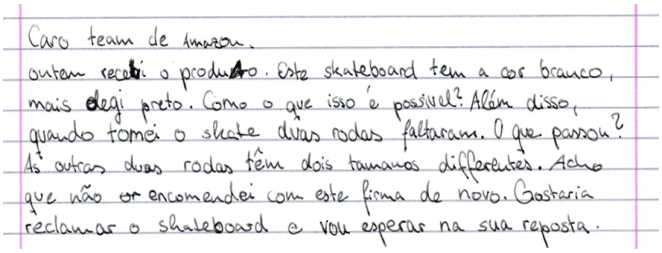
Além disso, o texto utiliza estruturas típicas de transcrição de narrativas em imagem, como a organização clara dos eventos e a explicitação da mensagem central: descreve o cenário e os personagens, apresenta o conflito central (o uso do telemóvel e seus impactos) e explica como o desfecho reflete uma crítica ao uso excessivo da tecnologia. Essas melhorias garantem que a mensagem seja compreendida de forma mais coerente, respeitando a ordem de transição das imagens e atendendo melhor ao critério de progressão lógica.

Relativamente aos comportamentos adotados por B durante a produção dos seus textos, observa-se que organiza suas ideias de forma ativa em português. Inicialmente, escreve palavras e tópicos de maneira simples, progressivamente desenvolvendo ideias mais complexas. Durante esse processo, reestrutura a ordem dos tópicos conforme necessário. Além disso, relê constantemente o que escreveu, identificando palavras faltantes ou ajustes que são necessários para completar suas ideias. B descreve esse processo como um ciclo contínuo de tentativas e revisões. Comparativamente à A, B dedica mais tempo à planificação de suas produções escritas, gerenciando de forma mais interativa os três componentes do processo. Esse comportamento reflete uma maior maturidade em sua competência escrita, o que resulta em textos mais coesos e com um número maior de palavras.

O terceiro gênero textual praticado na oficina correspondeu às reclamações formais. Seguindo a mesma orientação metodológica, pedimos aos participantes que escrevessem uma reclamação sobre qualquer tema, sem fornecer instruções. Reproduzimos nas Figuras 6 e 7 os textos escritos por C e por D.

Figura 6

Versão 1 da Reclamação de C


Caro team de Amazon.
ontem recebi o produto. Este skateboard tem a cor branco, mais eu pedi preto. Como o que isso é possível? Além disso, quando tomei o skate duas rodas faltaram. O que passou? As outras duas rodas têm dois tamanhos differentes. Acho que não or encomendei com este firma de novo. Gostaria reclamar o skateboard e vou esperar na sua reposta.

O texto escrito por C apresenta falhas na organização típica de uma reclamação formal. A introdução menciona o recebimento de um produto com problemas, mas não fornece informações essenciais, como o número do pedido ou o contexto do contato, nem a data e hora de compra. O desenvolvimento lista os defeitos do produto (cor errada, rodas faltando ou com tamanhos diferentes), mas a sequência das ideias é pouco clara (como as informações sobre as rodas, que são dadas em momentos diferentes). A conclusão expressa a insatisfação do autor e solicita uma resposta, mas com o uso de uma linguagem informal, sem o tom esperado para o gênero, comprometendo o critério de adequação.

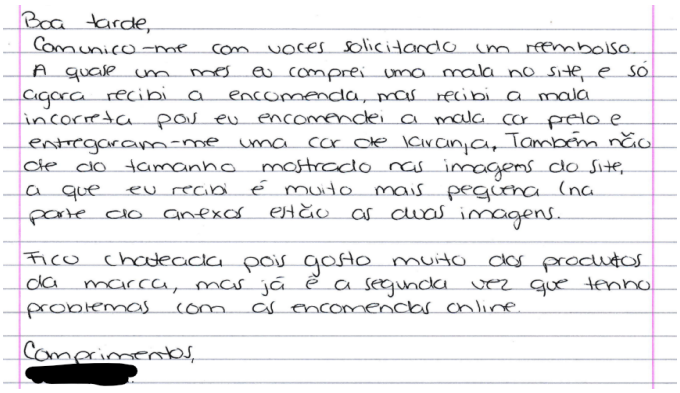
Quanto ao uso de expressões e estruturas, a reclamação contém frases de descontentamento e um pedido de resposta, mas novamente carece de formalidade e precisão. Não há a utilização de expressões típicas de reclamações formais, como "venho por este meio reclamar" ou "aguardo uma resposta breve". Além disso, falta polidez, com trechos que soam mais como desabafos do que como uma comunicação formal para uma empresa.

O texto reflete o esforço do estudante para seguir o que é esperado na produção desse gênero, mas sem um tom profissional; seria necessário reestruturá-lo em uma introdução, desenvolvimento e conclusão claros e com uma linguagem com marcas de cordialidade e de polidez, evitando frases como "O que passou?" ou "não encomendei com este firma de novo".

Analisamos a seguir novamente uma primeira versão de uma reclamação escrita por outro estudante.

Figura 7

Versão 1 da Reclamação de D


Boa tarde,
Comunico-me com voces solicitando um reembolso. A qual um mes eu comprei uma mala no site, e só agora recibi a encomenda, mas recibi a mala incorreta pois eu encomendei a mala cor preto e entregaram-me uma cor de laranja, Também não ete do tamanho mostrado nas imagens do site, a que eu recibi é muito mais pequena (na parte do anexos etão as duas imagens.

Fico chateada pois gosto muito dos produtos da marca, mas já é a segunda vez que tenho problemas com as encomendas online.

Comprimentos,
[redacted]

A primeira versão do texto escrito por D apresenta uma tentativa inicial de adequação ao gênero reclamação formal, mas ainda revela traços de informalidade e uma estrutura pouco desenvolvida. D opta por uma abertura direta ("Boa tarde"), sem identificação clara do destinatário, o que limita a formalidade e a clareza do público-alvo. O corpo do texto apresenta o motivo da reclamação - erro na cor e no tamanho da mala recebida -, mas de forma pouco organizada e com estrutura bastante linear, sem recorrer aos elementos estruturais típicos do gênero (como indicação de data da compra, número do pedido ou solicitação explícita de solução).

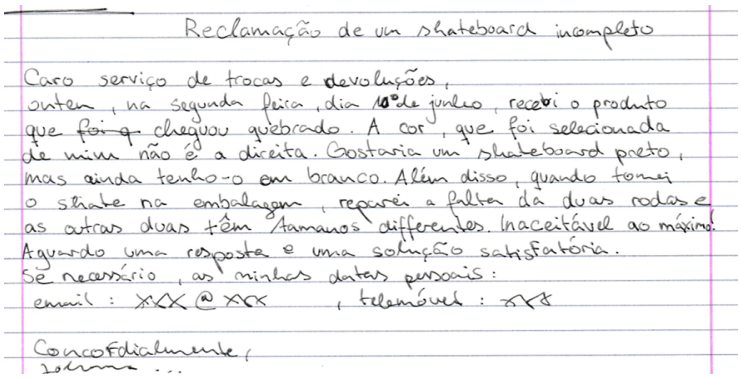
O uso de léxico especializado e de estruturas típicas ainda é incipiente. Termos como "reembolso" e "encomenda" apontam para algum conhecimento do vocabulário pertinente, mas faltam expressões formais como "gostaria de solicitar", "aguardo providências" ou "solicito a troca/reembolso", comuns nesse tipo de reclamação. A informalidade se acentua com o uso de expressões como "fico chateada", que, apesar de reforçar o descontentamento, não contribuem para a objetividade exigida nesse gênero.

No que diz respeito à progressão lógica, há uma apresentação clara do problema (recebimento de produto errado e de tamanho inadequado), mas a organização das ideias ainda precisa de refinamento. As frases são longas e pouco articuladas, e a solicitação principal (o que exatamente D espera que seja feito) fica implícita. A ausência de uma conclusão clara com expectativa de resposta ou prazo também prejudica o cumprimento completo da função comunicativa do gênero.

As Figuras 8 e 9 reproduzem as reclamações revisadas, de acordo com o trabalho que viemos desenvolvendo.

Figura 8

Versão 2 da Reclamação de C



Reclamação de um shateboard incompleto

Caro serviço de trocas e devoluções,
ontem, na segunda feira, dia 10 de junho, recebi o produto que ~~foi~~ chegou quebrado. A cor, que foi selecionada de mim não é a direita. Gostaria um shateboard preto, mas ainda tenho-o em branco. Além disso, quando tomei o skate na embalagem, reparei a falter de duas rodas e as outras duas têm tamanos differentes. Inaceitável ao máximo! Aguardo uma resposta e uma solução satisfatória.
Se necessário, as minhas datas pessoais:
email: xxx@xxx, telemóvel: xxx

Concordialmente,

A segunda versão da reclamação escrita por C demonstra melhor progressão lógica e organização, com introdução, desenvolvimento e conclusão bem definidos. A introdução identifica o destinatário e contextualiza o problema e, no desenvolvimento, há a descrição dos defeitos do produto de forma objetiva, como a cor incorreta, o estado do produto (quebrado) e os problemas nas rodas. A conclusão inclui um pedido de solução, os dados de contato e o fechamento adequado com a assinatura formal, cumprindo os critérios de adequação ao gênero e ao público-alvo.

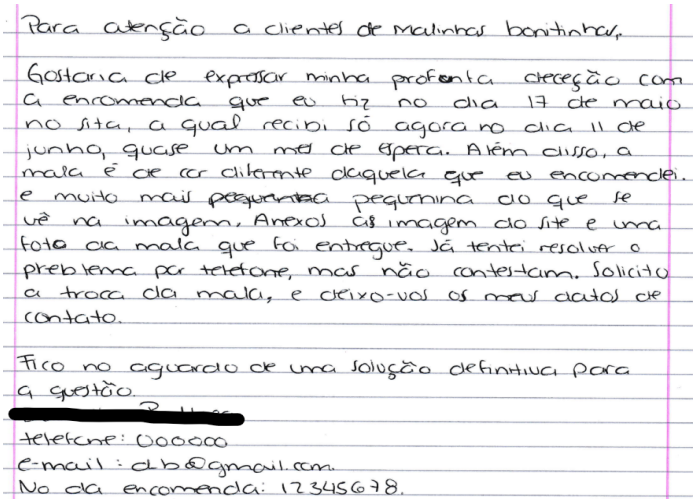
Quanto à utilização de expressões e estruturas típicas, o texto apresenta vocabulário apropriado e tom respeitoso, com frases como "Aguardo uma resposta e uma solução satisfatória.", evitando desabafos ou críticas excessivas.

Ainda em termos de adequação, o texto é funcional e segue as convenções de uma reclamação formal, mostrando um avanço significativo em relação à versão inicial, ainda que alguns itens presentes na grelha de revisão não tenham sido considerados como "Informa a(s) tentativa(s) prévia(s) de resolução tomada(s)" e "Solicita prazo de resposta".

Os comportamentos adotados por C na produção textual, especialmente durante a etapa de planificação, apresentam semelhanças aos de A. Inicialmente, C concebe as ideias gerais em sua L1. Em seguida, pensa nas estruturas gramaticais apropriadas ao gênero, registrando-as de forma isolada. Por fim, reflete sobre como intercalar as ideias anotadas no texto. Esse padrão indica uma habilidade ainda limitada para gerir os três componentes da produção escrita. Observa-se que C realiza poucas revisões em seus textos e não as integra de forma simultânea ao processo de escrita, o que vem a reforçar a necessidade da implementação de uma pedagogia que trabalhe de forma integrada e sistemática as três etapas da produção textual.

Figura 9

Versão 2 da Reclamação de D



Para atenção a cliente de Malinhas bonitinhas.

Gostaria de expressar minha profunda decepção com a encomenda que eu fiz no dia 17 de maio no site, a qual recibi só agora no dia 11 de junho, quase um mês de espera. Além disso, a mala é de cor diferente daquela que eu encomendei. e muito mais pequenina do que se vê na imagem. Anexo as imagem do site e uma foto da mala que foi entregue. Já tentei resolver o problema por telefone, mas não contestam. Solicito a troca da mala, e deixo-vos os meus dados de contato.

Fico no aguardo de uma solução definitiva para a questão.

[redacted]

telefone: 000000

e-mail: dlb@gmail.com.

No da encomenda: 12345678.

A nova versão da reclamação de D demonstra avanços significativos. Em termos de adequação ao gênero, o texto está bem estruturado: há uma saudação direcionada ao serviço de atendimento (“Para atenção ao cliente de Malinhas Bonitinhas”), uma apresentação clara do problema (demora na entrega e divergência no produto), uma tentativa de resolução já feita e um pedido objetivo de solução. O tom formal é mantido ao longo do texto, especialmente com o uso de verbos no condicional (“gostaria de expressar”), a ausência de marcas de informalidade e a inclusão de dados de contato no final da mensagem.

Quanto ao léxico e às estruturas típicas, há um progresso perceptível. Expressões como “profunda deceção”, “solicito a troca”, “fico no aguardo de uma solução definitiva” e o uso do termo “anexo” para indicar a inclusão de imagens estão bem alinhadas às convenções do gênero. A construção “quase um mês de espera” e o uso de termos como “cliente”, “encomenda”, “entregue”, “contataram” indicam familiaridade com o vocabulário próprio de situações de compra e atendimento ao consumidor.

A progressão lógica é clara e bem estabelecida. A reclamação segue uma sequência coerente: descrição do problema (atraso e entrega errada), comparação com o produto anunciado, referência à tentativa de resolução prévia, solicitação de solução e encerramento com expectativa de resposta. Essa organização permite ao leitor compreender o percurso da situação e qual é o pedido da cliente.

Relativamente ao comportamento estratégico adotado por D, observa-se que mistura sua L1 com o português para gerar as ideias iniciais, num esforço de organizar o conteúdo antes de focar na forma. Em um primeiro momento, anota tópicos de maneira livre, sem se preocupar com a correção linguística ou com a articulação completa das frases. Durante esse processo inicial, tende a descartar ideias que não consegue conectar de forma clara, priorizando uma estrutura mais linear e compreensível para si mesma.

No que diz respeito ao seu conhecimento linguístico, diz sentir dificuldade em reter vocabulário e evitar repetições ao longo do texto. Essa limitação interfere tanto na variedade lexical quanto na fluidez da escrita, levando-a a empregar estratégias alternativas para manter a progressão do texto, como o uso de sinônimos simples ou a simplificação de estruturas.

De forma geral, seu comportamento revela uma abordagem ainda em desenvolvimento quanto à planificação e à revisão textual. Apesar de apresentar um controle inicial sobre as ideias a serem incluídas, D

encontra desafios na construção de enunciados mais complexos e na diversificação do léxico, o que novamente sugere a importância de um trabalho mais sistemático com o enriquecimento vocabular e a articulação de ideias.

5. Considerações finais

Os resultados descritos neste artigo nos permitem responder de forma afirmativa à questão de pesquisa formulada*: O ELBT pode ser eficaz para o desenvolvimento da competência escrita em PL2?*

A análise dos dados apresentados revelou que o uso do ELBT tem um impacto positivo no aprimoramento da competência escrita. Após a fase de pós-tarefa, observamos uma melhoria significativa na amostra dos textos analisados, destacando-se o aumento do vocabulário especializado, o uso adequado de estruturas típicas dos diferentes gêneros textuais e maior adequação ao público-alvo, além de uma progressão lógica mais consistente.

Cabe ressaltar que pudemos associar o ELBT a um modelo de ensino de escrita que integra fundamentos cognitivos e sociais, uma vez que as tarefas realizadas não apenas abordaram o processo de produção dos estudantes, mas também o ensino sistemático dos gêneros textuais que trabalhamos. Durante a execução das atividades, identificamos comportamentos dos participantes que evidenciaram a relevância de uma abordagem pedagógica faseada no ensino da competência escrita, com atividades específicas para as etapas de planificação, textualização e revisão. Nesse sentido, nossas análises corroboram as observações da abordagem *writer-oriented*, que aponta que estudantes com maior proficiência dedicam mais tempo ao planejamento e à revisão de seus textos, demonstrando maior controle sobre a produção textual.

Embora a investigação descrita neste estudo tenha sido desenvolvida com metodologias qualitativas, as conclusões apresentadas são resultantes de um pequeno conjunto de dados, pelo que prosseguiremos com a recolha. Faremos novas oficinas com estudantes universitários de níveis B1 e também de B2 para termos mais materiais e dados de análise e ampliaremos nossas conclusões a respeito dos efeitos do ELBT na produção escrita em PL2.

Agradecimentos / Financiamento

Agradecemos aos professores de PL2 da NOVA-FCSH e da UTFPR-BR por participarem das entrevistas e compartilharem suas impressões sobre o uso de manuais didáticos e o ensino da competência escrita em suas aulas. Estendemos também nossa gratidão aos estudantes de PL2 que participaram nas oficinas, demonstrando interesse em aprimorar suas habilidades de escrita.

Parte deste trabalho é financiada por fundos nacionais através da FCT – Fundação para a Ciência e Tecnologia, I.P., no âmbito do projeto UID/03213 – Centro de Linguística da Universidade NOVA de Lisboa (CLUNL), e da bolsa de doutoramento 2023.05107.BD.

Referências

Bangert-Drowns, R., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research*, *74*(1), 29–58. https://doi.org/10.3102/00346543074001029

Barbeiro, L., & Pereira, L. (2007). *O ensino da escrita: A dimensão textual*. Ministério da Educação e Direção-geral de Inovação e de Desenvolvimento Curricular.

Berkenkotter, C., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication*. Hillsdale, Erlbaum.

Bygate, M., Norris, J, & Van den Branden, K. (2011). *Second language task complexity. Researching the Cognition Hypothesis of language learning and performance*. John Benjamins Publishing Company.

Byrnes, H. (2014). Linking task and writing for language development: Evidence from a genre-based curricular approach. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning. Insights from and for L2 writing* (pp. 237–263). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.7.10byr

Byrnes, H., & Manchón, R. M. (Eds.). (2014). *Task-based language learning. Insights from and for L2 writing*. John Benjamins Publishing Company. https://doi.org/10.1075/tblt.7
Cassany, D. (1999). *Construir la escritura*. Ediciones Paidós Ibérica.
Cassany, D. (2017). *Describir el escribir. Cómo se aprende a escribir*. Espasa Libros.
Castro, C. (2017). *Ensino de línguas baseado em tarefas. Da teoria à prática*. Lidel – edições técnicas, Lda.
Castro, C. (2020). Ensino de línguas baseado em tarefas: Princípios para a elaboração de tarefas. In P. Almeida & M. Vieira (Orgs.), *Por palavras e gestos: A arte da linguagem* (Vol. 2, pp. 51–62). Artemi.
Cazo. (s.d.). *Charge sobre saúde mental* [Imagem]. Blog do AFTM. Retirado a 18 maio 2025, de https://blogdoaftm.com.br/charge-saude-mental
Cumming, A. (2001). Learning to write in a second language: Two decades of research. *IJES*, *1*(2), 1–23.
Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*, 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x
Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387. https://doi.org/10.2307/356600
Grabe, W., & Kaplan, R. (1996). *Theory and practice of writing*. Longman.
Graham, S. (2018). Introduction to conceptualizing writing. *Educational Psychologist, 53*(4), 217–219. https://doi.org/10.1080/00461520.2018.1514303
Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81*(4), 710–744. https://psycnet.apa.org/doi/10.17763/haer.81.4.t2k0m13756113566
Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In C. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Routledge.
Hayes, J., & Flower, L. (1980). *Identifying the organization of writing processes. Cognitive processes in writing*. Lawrence Erlbaum Associates.
Howatt, A. (1984). *A history of language teaching*. Oxford University Press.
Hyland, K. (2003). *Second language writing*. Cambridge University Press.
Hyland, K. (2009). *Teaching and researching writing*. Pearson Education Limited.
Koch, I. V. (2014). *As tramas do texto*. Editora Contexto.
Larios, J. R., Manchón, R., & Murphy, L. (2007). Componentes básicos y evolutivos del proceso de formulación en la escritura de textos de lengua materna y lengua extranjera. *RESLA 20*, 159–183.
Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. Routledge Taylor & Francis Group.
Long, M. (1991). Focus on form: A design feature in language teaching. methodology. In K. de Boot, R. B. Ginsberg & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.2.07lon
Long, M. (1998). *Focus on form in task-based language teaching*. John Benjamins Publishing Company.
Long, M. (2015). *Second language acquisition and task-based language teaching*. Wiley Blackwell.
Lopes, A., & Graça Pinto, M. (2022). Assessing L2 Portuguese writing: Idea density and sentence complexity. *Signo*, *47*(88), 72–85.
Minayo, M. C. (2012). *Pesquisa social. Teoria, método e criatividade*. Editora Vozes.
Macaro, E. (2014). Reframing task performance: The relationship between tasks, strategic behaviour, and linguistic knowledge in writing. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning. Insights from and for L2 writing* (pp. 53–78). John Benjamins Publishing Company.
Manchón, R. M. (2014). The internal dimension of tasks: The interaction between task factors and learner factors in bringing about learning through writing. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning. Insights from and for L2 writing* (pp. 27–52). John Benjamins Publishing Company.

Miller, C. R. (2012). *Gênero textual, agência e tecnologia: Estudos*. Parábola Editorial.
Mosterín, J. (2002). *Teoría de la escritura*. Icaria Editorial.
Parracho, S. C. M. (2011). *Português L2: Ensino de escrita e input linguístico* [Dissertação de Mestrado, Escola Superior de Educação de Lisboa]. RCIPL – Repositório Institucional do Politécnico de Lisboa. http://hdl.handle.net/10400.21/2394
Pinnard, L. (2016). Looking outwards: Using learning materials to help learners harness out-of-class learning opportunities. *Innovation in language learning and teaching*, *10*(2), 133–143. https://doi.org/10.1080/17501229.2015.1090997
Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667336
Prabhu, N. (1987). *Second language pedagogy*. Oxford University Press.
Pinto, J. (2020). O ensino de línguas baseado em tarefas no ensino/aprendizagem da escrita em português língua segunda. Propostas didáticas. *Revista do GEL, 17*(2), 170–195. https://doi.org/10.21165/gel.v17i2.2425
Rebuschat, P. (2015). *Implicit and explicit learning of languages*. John Benjamins Publishing Company. https://doi.org/10.1075/sibil.48
Richards, J. (2006). *Communicative language teaching today*. Cambridge University Press.
Richards, J. C., & Rodgers, T. (2001). *Approaches and methods in language teaching.* Cambridge University Press. https://doi.org/10.1017/9781009024532
Rinnert, C., & Kobayashi, H. (2009). Situated writing practices in foreign language settings:
The role of previous experience and instruction. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching and research* (pp. 23–48). Multilingual Matters.
Robinson, P. (2011). *Second language task complexity. Researching the Cognition Hypothesis of language learning and performance*. John Benjamins Publishing Company. https://doi.org/10.1075/tblt.2
Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (pp. 142–175). Cambridge University Press. https://doi.org/10.1017/CBO9780511571374
Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing*: R*esearch insights for the classroom* (pp. 11–23). Cambridge University Press.
Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing. *TESOL Quarterly, 27*(4), 657–677. https://doi.org/10.2307/3587400
Sippel, J. (2023). Ensino de escrita e avaliação de manuais didáticos de Português L2. *Revista da Associação Portuguesa de Linguística*, *10*, 287–307. https://doi.org/10.26334/2183-9077/rapln10ano2023a16
Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, *17*(1), 38–62. https://doi.org/10.1093/applin/17.1.38
Swales, J. M. (1990). Genre analysis: English in academic and research settings. *Studies in Second Language Acquisition*, *15*(1), 125–126. Cambridge University Press. https://doi.org/10.1017/S0272263100011773
Swales, J. M. (1998). *Other floors, other voices: A textography of a small university building*. Lawrence Erlbaum.
Tomlinson, B., & Masuhara, H. (2018). *The complete guide to the theory and practice of materials development for language learning*. John Wiley & Sons.
Tomlinson, B. (2010a). Helping learners to fill the gaps in their learning. In F. Mishan & A. Chambers (Eds.), *Perspectives on language learning materials development* (pp. 87–108). Peter Lang.
Tomlinson, B. (2010b). Principles and procedures of material development. In N. Harwood (Ed.), *Materials in ELT: Theory and practice* (pp. 1–10). Cambridge University Press.
Tomlinson, B. (2011). *Materials development in language teaching*. Cambridge University Press. https://doi.org/10.1017/9781139042789

Tomlinson, B. (2016). The importance of materials development for language learning. In M. Azarnoosh, M. Zeraatpishe, A. Faravani & H. R. Kargozari (Eds.), *Issues in materials development* (pp. 1–9). Sense Publishers. https://doi.org/10.1007/978-94-6300-432-9_1

Van den Branden, K., Bygate, M., & Norris, J. (Eds.). (2009). *Task-based language teaching: a reader*. John Benjamins Publishing Company.

Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford University Press.

Aquisição de segunda língua e mudança linguística: O caso dos sujeitos nulos

L2 acquisition and linguistic change: The case of null subjects

Joana Teixeira[1], *Alexandra Fiéis*[1]

[1] Universidade NOVA de Lisboa, Faculdade de Ciências Sociais e Humanas | Centro de Linguística da Universidade Nova de Lisboa

Abstract

This study investigates parallels between L2 acquisition and language change, exploring the hypothesis that the transition from a null subject grammar to a non-null subject grammar follows common principles in both contexts. The study focuses on the acquisition of L2 English by native speakers of European Portuguese, comparing it with diachronic change data from Brazilian Portuguese. 64 Portuguese learners of L2 English (levels B1 to C2) and 12 native English speakers completed a speeded acceptability judgment task that crossed the variables *pronominal subject* (overt vs. null) and *referentiality* (2p vs. 3p [+human] vs. 3p [-human] vs. expletive). The results show that, unlike native English speakers, Portuguese learners of L2 English do not systematically reject null subjects. As proficiency increases, a gradual decrease in the acceptance of null subjects is observed, progressing from the [+referential] to the [-referential] end of the Referential Hierarchy. Our findings show that referentiality guide linguistic development in L2 acquisition (L1 EP – L2 English) and diachronic change (in BP).

Keywords: L2 acquisition, linguistic change, null subjects, European Portuguese.

Resumo

Este estudo investiga paralelismos entre a aquisição de L2 e a mudança linguística, explorando a hipótese de que a transição de uma gramática de sujeito nulo para uma gramática de sujeito obrigatório segue princípios comuns nos dois contextos. O estudo centra-se na aquisição de inglês L2 por falantes nativos de português europeu, comparando-a com dados da mudança diacrónica no português brasileiro. 64 aprendentes portugueses de inglês L2 (níveis B1 a C2) e 12 falantes nativos de inglês realizaram uma tarefa de juízos de aceitabilidade rápida que cruzou as variáveis *sujeito pronominal* (pleno vs. nulo) e *referencialidade* (2p vs. 3p [+humano] vs. 3p [-humano] vs. expletivo). Os resultados mostram que, ao contrário dos falantes nativos de inglês, os aprendentes portugueses de inglês L2 não rejeitam sistematicamente sujeitos nulos. À medida que a proficiência aumenta, observa-se uma redução progressiva da aceitação de sujeitos nulos, que progride do extremo [+referencial] para o extremo [-referencial] da Hierarquia de Referencialidade. Os nossos resultados mostram que a referencialidade guia o desenvolvimento linguístico tanto na aquisição de L2 (PE L1 – inglês L2) como na mudança diacrónica (no PB).

Palavras-chave: aquisição de L2, mudança linguística, sujeitos nulos, português europeu.

1. Introdução

A aquisição de segunda língua (L2) e a mudança linguística partilham traços fundamentais, como a instabilidade, a variação e a reestruturação das gramáticas mentais dos falantes. Estes fenómenos refletem processos de adaptação linguística, nos quais os sistemas gramaticais são reconfigurados em resposta a alterações no input, no contexto de uso ou nas capacidades cognitivas dos indivíduos. Vários autores têm sugerido que, apesar das diferenças entre os domínios da aquisição e da mudança[1], os mecanismos subjacentes à reestruturação gramatical podem ser, em muitos casos, análogos ou até partilhados (e.g., Donaldson, 2022).

Neste sentido, Meisel (2011) propõe uma abordagem integrada ao estudo do desenvolvimento linguístico, defendendo que a tradicional compartimentação entre áreas como a aquisição monolingue e bilingue de língua materna (L1), a aquisição de L2, a aquisição atípica e a mudança diacrónica pode limitar a compreensão dos processos envolvidos. Para o autor, uma visão mais abrangente e comparada entre estas áreas permitirá identificar padrões de reorganização gramatical semelhantes que orientam o desenvolvimento linguístico. Nas palavras do autor:

> […] the division of labour among research disciplines [monolingual and bilingual first language acquisition, second language acquisition, impaired language acquisition (SLI), the genesis of pidgins and creoles, and diachronic change] may represent an obstacle to achieving an adequate understanding of the driving forces and the mechanisms determining linguistic development. In fact, it is very likely that deeper insights into the mechanisms of development can only be attained by taking into consideration not merely the particularities of one type but also the commonalities across several or all of them. (Meisel, 2011, p. 122)

No sentido de testar esta proposta, o presente estudo explora a hipótese de que a transição de uma gramática de sujeito nulo para uma gramática de sujeito obrigatório pode constituir um processo semelhante tanto em contextos de mudança diacrónica como em aquisição de L2. Ao investigar comparativamente estes dois contextos, pretende-se contribuir para uma abordagem mais integrada do desenvolvimento linguístico e para o esclarecimento das condições sob as quais as gramáticas são passíveis de mudança estrutural.

O artigo encontra-se organizado da seguinte forma: na secção 2, apresenta-se um enquadramento teórico sobre pronominalização e referencialidade em dados diacrónicos e em aquisição de L2; na secção 3, apresenta-se a questão de investigação e a metodologia do estudo, com descrição dos participantes, da tarefa experimental, e da análise estatística dos resultados; na secção 4, descrevem-se os resultados; e, finalmente, na secção 5, discutem-se os resultados e apresentam-se as principais conclusões.

2. Pronominalização e referencialidade

A relação entre pronominalização e referencialidade tem sido amplamente discutida na literatura (e.g., Ariel, 1990, 2001; Cyrino et al., 2000; Fukumura & van Gompel, 2010), uma vez que o uso de pronomes é condicionado por fatores semânticos e pragmático-discursivos, principalmente no que diz respeito à acessibilidade e à saliência dos referentes no discurso.

Trabalhos como os de Ariel (1990, 2001) mostram uma correlação sistemática entre a forma da expressão anafórica e o grau de acessibilidade do antecedente no discurso: quanto mais acessível (isto é, quando já foi introduzido no discurso) ou saliente sintaticamente for um antecedente no contexto discursivo, mais reduzida tende a ser a forma morfossintática usada para o referir. A acessibilidade de um antecedente depende, entre outros fatores, da sua posição estrutural (posição de sujeito mais acessível do que posição não sujeito), do seu

[1] Apesar dos paralelos, é importante sublinhar algumas diferenças cruciais. Na mudança diacrónica, o processo é gradual e afeta toda a comunidade linguística, ocorrendo ao longo de gerações e interagindo com fenómenos sociolinguísticos e estilísticos. Já na aquisição de L2, a mudança afeta gramáticas individuais em contacto com uma língua-alvo estável, sendo influenciada por fatores como idade de aquisição, proficiência, contexto de exposição e interferência da L1.

estatuto discursivo (tópicos mais acessíveis do que não tópicos) e, segundo alguns autores (e.g., Fukumura & van Gompel, 2010; Morgado et al., 2018), dos seus traços de animacidade (+ humano mais acessível do que – humano).

A pronominalização pressupõe que o referente seja suficientemente identificável, dispensando formas mais explícitas. Os pronomes surgem, assim, associados a altos níveis de acessibilidade, o que significa que ocupam uma posição baixa na hierarquia de formas referenciais. Este princípio é central para a economia referencial e para a estruturação do discurso, e tem implicações tanto sincrónica quanto diacronicamente.

Estudos diacrónicos como os de Cyrino et al. (2000) mostram que alterações morfossintáticas, como o enfraquecimento da morfologia verbal, levaram a que o português do Brasil (PB) se tornasse uma língua de sujeito nulo parcial e afetaram diretamente os mecanismos de pronominalização, sendo estas mudanças influenciadas pela referencialidade. Por outro lado, trabalhos sobre L2 indicam que os aprendentes revelam dificuldades em eliminar a possibilidade de sujeito nulo na aquisição de línguas de sujeito obrigatório, havendo indícios de uma potencial influência da referencialidade neste processo.

A análise articulada de dados diacrónicos e de aquisição de L2 permitirá, assim, uma compreensão mais abrangente dos fatores que condicionam o comportamento pronominal nas línguas naturais, conforme será explorado nas subsecções seguintes.

2.1. **Dados diacrónicos**

A investigação em sintaxe diacrónica tem mostrado que a referencialidade desempenha um papel relevante nos processos de pronominalização em diferentes línguas. O PB constitui um caso paradigmático neste contexto. A investigação sobre o PB indica que esta língua tem evoluído gradualmente de um sistema de sujeito nulo consistente, semelhante ao português europeu (PE), para um sistema de sujeito nulo parcial (cf. Cyrino et al., 2000).

No PB contemporâneo, observa-se uma distinção clara entre sujeitos expletivos e sujeitos referenciais no que respeita à possibilidade de omissão. No caso dos sujeitos expletivos, a forma nula é obrigatória, como se mostra em (1), em que a inserção de um pronome sujeito (*ele*) resulta em agramaticalidade:

(1) (*Ele) parece que (*ele) vai chover

Por outro lado, as formas nulas referenciais têm vindo a desaparecer de forma gradual desde o século XIX. No PB atual, especialmente nas variedades urbanas e formais, é comum que sujeitos referenciais sejam realizados por pronomes plenos, mesmo em contextos em que, historicamente, seriam nulos. Este fenómeno reflete uma mudança neste sistema linguístico. Consideremos o exemplo (2):

(2) Você$_i$ me disse que você$_i$ está morando em Copacabana.

Neste caso, a repetição do pronome *você* em ambas as orações mostra a preferência atual pela realização explícita do sujeito, mesmo quando o referente se mantém o mesmo.

Segundo Cyrino et al. (2000), a mudança de um sistema de sujeito nulo consistente para um sistema de sujeito nulo parcial foi guiada pela Hierarquia de Referencialidade (HR) apresentada na Figura 1, em que a 1.ª e a 2.ª pessoas, com o traço inerente [+humano], ocupam as posições mais altas, e o expletivo a posição mais baixa.

Figura 1

Hierarquia de Referencialidade (Cyrino et al., 2000, p.59)

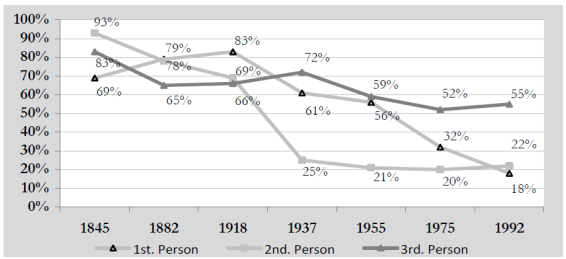
non-argument	proposition	[-human]		[+human]
			3rd p.	2nd p. 1st p.
-specif				+specif
[-ref] ◂				▸ [+ref]

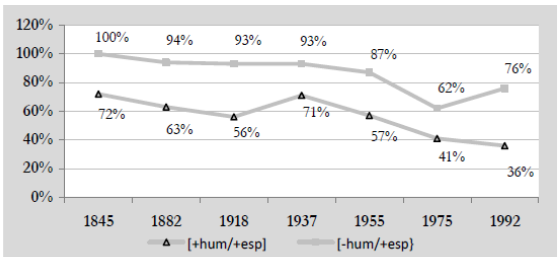
De acordo com a Hipótese do Mapeamento Implicacional, formulada por Cyrino et al. (2000), "the more referential, the greater the possibility of a non-null pronoun", e "a null variant at a specific point on the scale implies null variants to its left in the referential hierarchy" (Cyrino et al., 2000, p. 59). Assim, esta hipótese sugere que há uma hierarquia na pronominalização baseada na referencialidade, ou seja, se uma expressão com baixo grau de referencialidade pode ser pronominalizada, então expressões com maior grau de referencialidade também o poderão ser.

As propostas de Cyrino et al. (2000) relativas à trajetória de mudança da pronominalização no PB têm por base estudos diacrónicos. Como se observa nas Figuras 2 e 3 de Duarte e Varejão (2013), dados baseados em corpora revelam uma tendência de diminuição progressiva do uso de sujeitos nulos ao longo da história do PB (cf. Cyrino et al., 2000; Duarte, 1995; Duarte & Silva, 2016; Duarte et al., 2012). Esta diminuição é mais acentuada no caso de sujeitos nulos de 1.ª e 2.ª pessoas, que se tornaram residuais no PB contemporâneo. Já no caso dos sujeitos nulos de 3.ª pessoa, a redução é mais gradual: sujeitos nulos [+humanos] foram perdidos de forma mais rápida e em maior extensão do que sujeitos nulos [-humanos]. Importa notar que esta mudança não afetou os sujeitos nulos expletivos, que estão no extremo [– referencial] da HR e continuam a ser nulos em PB. Assim, diacronicamente, a eliminação dos sujeitos nulos progride ao longo da HR da direita para a esquerda.

Figuras 2 e 3

Sujeitos Nulos (vs Plenos) e HR na Diacronia do PB (Duarte & Varejão, 2013)



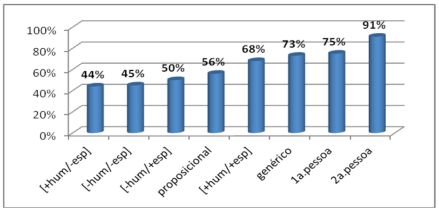


A redução significativa da produtividade dos sujeitos nulos em PB é tradicionalmente explicada pela perda de traços ricos de concordância verbal, que possibilitariam a identificação do sujeito com base na morfologia verbal (cf. Barbosa et al., 2005; Rizzi, 1982). A perda da riqueza de concordância verbal estaria na origem da mudança paramétrica observada, favorecendo a emergência de uma gramática mais próxima das gramáticas de línguas de sujeito obrigatório. A realização plena do sujeito torna-se, assim, uma estratégia para garantir a clareza e evitar ambiguidades.

A mudança ilustrada nas Figuras 2 e 3 é ainda mais evidente nos dados do PB contemporâneo apresentados na Figura 4, em que se observa que a realização dos sujeitos é fortemente condicionada pela HR.

Figura 4

Sujeitos Plenos (vs Nulos) e HR PB Contemporâneo (Kato & Duarte, 2014, p. 5)

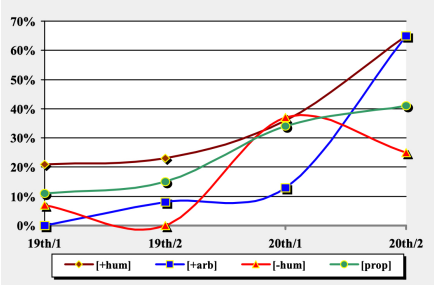


Como ilustrado na Figura 4, os pronomes sujeito de 1.ª e 2.ª pessoas, por serem geralmente humanos e discursivamente proeminentes, tendem a ser realizados. Por seu turno, os sujeitos de 3.ª pessoa, especialmente aqueles com referentes [-humanos] ou menor grau de referencialidade, são expressos com menor frequência. Este padrão é, pois, característico de uma gramática de sujeito nulo parcial.

De uma maneira geral, os dados diacrónicos e sincrónicos do PB são consistentes com a HR e a Hipótese do Mapeamento Implicacional. No entanto, pode-se contestar a inclusão de proposições na HR, uma vez que, na diacronia do PB, não se observam diferenças significativas entre sujeitos proposicionais e sujeitos [-humanos], como se observa na Figura 5. Na 2.ª metade do século XX, contrariamente ao que a HR prediria, ocorrem mais sujeitos realizados proposicionais do que [- humanos] (cf. Figura 5). Como a Figura 4 demonstra, na sincronia do PB, também não se identificam diferenças relevantes entre sujeitos proposicionais e [-humanos]. Assim, os sujeitos proposicionais não serão considerados na presente investigação.

Figura 5

Sujeitos Plenos (vs Nulos) e HR na Diacronia do PB (Cyrino et al., 2000, p. 63)



Considerados em conjunto, os dados diacrónicos e sincrónicos indicam uma tendência crescente para a realização plena do sujeito, sugerindo uma reconfiguração da sintaxe do PB no sentido de uma maior explicitação referencial. Esse quadro é compatível com a ideia de que o PB contemporâneo ocupa uma posição intermédia entre uma gramática de sujeito nulo consistente, como o PE, e uma gramática de sujeito obrigatório, como o inglês, configurando-se como um sistema híbrido em transição, cujos mecanismos de pronominalização estão ainda em fase de reestruturação.

2.2. **Dados de aquisição de L2**

No domínio da aquisição de L2, os sujeitos pronominais têm sido alvo de muita investigação nas últimas décadas, em particular em pares de línguas em que a L1 é uma língua de sujeito obrigatório e a L2 uma língua de sujeito nulo.

Os estudos mostram que, nestes pares de línguas, a aquisição da sintaxe do sujeito tende a ocorrer sem grandes dificuldades (e.g., Al-Kasey & Pérez-Leroux, 1998; Liceras, 1989; Pérez-Leroux & Glass, 1999; Rothman & Iverson, 2007), pois os aprendentes adquirem sujeitos nulos relativamente cedo (e.g., Rothman, 2009; White, 1985). Contudo, continuam a apresentar dificuldades no conhecimento das condições discursivas que regulam a distribuição de sujeitos pronominais plenos, mesmo em níveis altamente avançados (e.g., Belletti et al., 2007; Lozano, 2006; Sorace & Filiaci, 2006).

Já para o percurso inverso, ou seja, quando a L1 é uma língua de sujeito nulo e a L2 uma língua de sujeito obrigatório, os poucos estudos existentes indicam que os aprendentes têm dificuldades na aquisição da sintaxe do sujeito até níveis avançados, continuando a admitir sujeitos nulos em L2 como o inglês (e.g., Ballester, 2013; Judy, 2011; Judy & Rothman, 2010; Prentza & Tsimpli, 2013; Teixeira, 2019). Este padrão é observado quando a L1 do aprendente é uma língua de sujeito nulo, mas não quando é uma língua de sujeito obrigatório como o francês (Teixeira, 2019), o que indica que as dificuldades resultam de influência da L1.

De acordo com Teixeira (2019), as dificuldades sintáticas exibidas pelos falantes nativos de línguas de sujeito nulo podem ser ultrapassadas, já que, no nível quase-nativo, os aprendentes exibem comportamentos idênticos aos dos falantes nativos de inglês, rejeitando de forma consistente tanto sujeitos expletivos quanto referenciais nulos. Estes resultados reforçam a ideia de que, apesar dos atrasos, a referencialidade e a pronominalização plena são adquiríveis, e que, com o aumento da proficiência, a influência da L1 tende a ser progressivamente atenuada. Para além disso, este estudo sugere que, nas interlínguas dos aprendentes, a eliminação dos sujeitos nulos de 3.ª pessoa é influenciada pela referencialidade, seguindo a ordem: sujeito nulo referencial [+humano] > sujeito nulo referencial [-humano] e sujeito nulo expletivo.

Os resultados preliminares de Teixeira (2019) sugerem, assim, que a eliminação de formas nulas poderá ocorrer do extremo [+referencial] para o [-referencial] da HR, sendo plausível assumir-se que a aquisição de uma L2 como o inglês por falantes nativos de PE, línguas com comportamentos contrastantes no que diz respeito à realização de sujeitos e ao uso de pronomes, envolve uma mudança comparável à que se verificou na diacronia do PB.

A análise conjunta dos dados diacrónicos e dos estudos sobre aquisição de L2 permite observar que, na pronominalização, a referencialidade parece emergir como um princípio organizador importante, quer nas mudanças linguísticas observadas ao longo do tempo, quer nos desafios enfrentados por aprendentes em contextos de aquisição. Em todo o caso, os estudos existentes sobre sujeitos em inglês L2 têm investigado maioritariamente sujeitos de 3.ª pessoa com antecedentes DP, sem explorar todas as possibilidades na HR. Além disso, os estudos realizados têm-se concentrado predominantemente em um ou dois níveis de proficiência — sobretudo nos níveis intermédio-alto e avançado — o que impede uma compreensão abrangente e sistemática do percurso de aquisição. Deste modo, permanece por responder a seguinte questão: na aquisição de uma língua de sujeito obrigatório como L2 por falantes de línguas de sujeito nulo como L1, a reestruturação das gramáticas destes falantes é guiada pela HR, à semelhança do que se observa na diacronia do PB?

3. Questão de investigação e metodologia

Considerando o estado da arte apresentado na secção anterior, bem como as lacunas identificadas, nesta secção apresentamos as questões orientadoras do estudo e descrevemos a metodologia adotada para a sua exploração empírica, destacando-se a descrição dos participantes, a tarefa experimental utilizada e a análise estatística dos dados.

3.1. Questão de investigação

Este estudo analisa o fenómeno da transferência de sujeitos nulos por falantes nativos de PE na aquisição do inglês como L2, com o objetivo de determinar em que medida a reestruturação da interlíngua é orientada pela HR. A investigação parte do pressuposto de que a HR — originalmente proposta, como vimos na secção

anterior, no âmbito da mudança linguística diacrónica — pode igualmente constituir um princípio orientador no desenvolvimento linguístico em contextos de aquisição de L2. Para isso, a nossa questão de investigação centra-se na possibilidade de os falantes de PE L1 transferirem a propriedade de sujeitos nulos para o inglês L2, uma língua de sujeito obrigatório, e procura determinar se, nesse processo, a reestruturação da interlíngua é guiada pela HR. A hipótese correspondente prediz que, se a HR tiver um papel no desenvolvimento linguístico dos sujeitos pronominais, tal como tem sido descrito para a mudança linguística, então os aprendentes tenderão a eliminar gradualmente os sujeitos nulos da sua interlíngua, partindo do extremo [+ referencial] e avançando para o extremo [- referencial] da hierarquia.

Com o objetivo de responder a esta questão, este estudo investiga a realização e omissão de sujeitos pronominais de diferentes tipos em inglês L2 por falantes nativos de PE, recorrendo a uma tarefa de juízos de aceitabilidade rápida.

3.2. **Metodologia**

Nas secções seguintes, descrevem-se os participantes do estudo (3.2.1), a tarefa utilizada (3.2.2) e os procedimentos adotados na análise estatística (3.2.3).

3.2.1. **Participantes**

Participaram neste estudo 64 falantes nativos de PE, distribuídos pelos seguintes níveis de proficiência em inglês L2: B1 (n = 10), B2 (n = 18), C1 (n = 18) e C2 (n = 18). Todos os participantes são filhos de falantes monolingues de PE e têm esta língua como única L1. As suas idades variam entre os 18 e os 55 anos. À data do estudo, todos eram estudantes numa universidade portuguesa e nenhum tinha vivido num país de língua inglesa. Na maioria dos casos, a exposição ao inglês teve início no 1.º ciclo do ensino básico, sendo a média de idades de início de exposição igual ou superior a 5,9 anos em todos os subgrupos.

O nível de proficiência dos participantes foi determinado com base no último nível concluído no semestre imediatamente anterior à realização da tarefa experimental. Na universidade frequentada pelos participantes, os cursos de inglês estão organizados segundo o *Quadro Europeu Comum de Referência para as Línguas* (Conselho da Europa, 2001, 2020). Para assegurar que os estudantes sejam adequadamente agrupados conforme as suas competências linguísticas reais, esta instituição aplica testes de colocação em nível no momento da primeira matrícula. Dado que todos os participantes foram recrutados na mesma universidade, os níveis atribuídos oferecem uma base sólida para a análise realizada no presente estudo.

O estudo incluiu ainda um grupo de controlo constituído por 12 falantes nativos de inglês, com idades compreendidas entre os 19 e os 72 anos. Todos são filhos de falantes monolingues de inglês e têm esta língua como única L1. A Tabela 1 apresenta os dados relativos à idade e à idade de início de exposição ao inglês de cada grupo, recolhidos através de um questionário sociolinguístico preenchido antes da realização da tarefa experimental.

Tabela 1

Dados sobre os Participantes

Grupo	*N*	Idade		Idade de início de exposição ao inglês	
		M	**DesvPad**	**M**	**DesvPad**
B1	10	29,3	12,9	7,8	2,3
B2	18	22,2	8,1	7,7	2,4
C1	18	20,4	1,6	5,9	2,1
C2	18	26,8	9,3	7,4	2,2
Nativo	12	41,9	17,5	0	0

3.2.2. Tarefa

Todos os participantes realizaram uma tarefa de juízos de aceitabilidade rápida, desenvolvida no software PsychoPy (Peirce et al., 2019) e administrada online através da plataforma Pavlovia. Esta tarefa tinha um desenho 2x4, cruzando as seguintes variáveis: *tipo de sujeito pronominal* – pleno vs. nulo – e *referencialidade* – 2.ª pessoa do singular vs. 3.ª pessoa do singular [+ humano] vs. 3.ª pessoa do singular [- humano] vs. expletivo. Em todos os itens experimentais, o sujeito pronominal estava encaixado numa oração subordinada adverbial causal iniciada por *because* e era antecedido por um sintagma preposicional ou um sintagma adverbial. A frase tinha a ordem matriz-subordinada e tinha entre 14 e 16 palavras. Os verbos estavam no passado simples. Nos itens experimentais que testavam sujeitos referenciais, os sujeitos estavam encaixados num contexto de manutenção de tópico, retomando o antecedente na posição de sujeito na oração matriz. Em nenhum item, havia ambiguidade quanto ao antecedente potencial do sujeito pronominal. Foi usado este contexto discursivo por favorecer a omissão de sujeitos no PE. A tabela 2 apresenta itens de exemplo para cada condição.

Tabela 2

Itens de Exemplo

Referencialidade	**Tipo de sujeito**	
	Nulo	**Pleno**
2p	You couldn't see her flaws, because at the time [-] were in love with her.	You didn't get the job interview because in the application you submitted an incomplete resume.
3p [+humano]	The defendant was acquitted by the jury, because during the trial [-] was proven innocent.	The man made a series of regrettable decisions because that night he was drunk.
3p [-humano]	The tower is called 'Torre della Specola' because in the 18th century [-] contained an astronomical observatory.	The vaccine was recalled because last week it caused many cases of severe allergic reactions.
Expletivo	The streets of Chicago are covered in a white blanket because yesterday [-] snowed all day.	All the farmers in the region are upset because this week it rained every day.

A tarefa de juízos de aceitabilidade incluía 3 itens para cada uma das 8 condições experimentais e 30 distratores, tendo um total de 54 itens. Adicionalmente, foram incluídos 6 itens de treino, destinados a familiarizar os participantes com o ritmo da tarefa, e três itens de atenção, em que a frase apresentada palavra

por palavra instruía o participante a pressionar um determinado número de 1 a 5 para mostrar que estava atento. Todos os participantes incluídos no presente estudo responderam corretamente a todas as questões de atenção.

Em cada item da tarefa de juízos de aceitabilidade rápida, era inicialmente apresentado um ponto de fixação no centro do ecrã durante 1500 ms. Em seguida, a frase a ser avaliada surgia palavra por palavra, de forma não cumulativa, no centro do ecrã, a um ritmo fixo de 500 ms por palavra. Após a apresentação completa da frase, os participantes avaliavam a sua aceitabilidade numa escala de 1 a 5, sendo 1 correspondente a "totalmente inaceitável" e 5 a "totalmente aceitável". Caso não soubessem como avaliar a frase, deveriam pressionar a tecla "N". O ritmo de 500 ms por palavra foi estabelecido com base nos resultados de um teste-piloto que realizámos com um grupo de aprendentes de inglês L2 nos níveis considerados neste estudo, o qual indicou que esta janela temporal era suficiente para os participantes realizarem os processos normais de compreensão.

A tarefa de juízos de aceitabilidade rápida foi usada por dois motivos. Em primeiro lugar, Teixeira (2019) demonstrou que este tipo de tarefa é mais sensível às dificuldades dos aprendentes de inglês L2 com sujeitos do que uma tarefa sem pressão de tempo, como uma tarefa de produção induzida. Em segundo lugar, considerando que a omissão do sujeito em inglês é, em geral, objeto de correção explícita em contextos de ensino formal, contexto no qual os participantes deste estudo são expostos à língua, o recurso a uma tarefa com pressão de tempo é importante para limitar a influência de conhecimento linguístico explícito (i.e., conhecimento consciente de regras da língua) nas respostas dos participantes. Este tipo de conhecimento não está prontamente disponível para uso automático, envolvendo esforço de processamento e exigindo, portanto, tempo para ser usado (Hulstijn, 2005). Ao impor pressão de tempo, a tarefa de juízos de aceitabilidade rápida impede os participantes de refletirem sobre as frases e força-os a basearem as suas respostas essencialmente na sua intuição inicial (e.g., Bader & Häussler, 2010). Esta tarefa tem sido, por isso, validada por diversos estudos psicométricos como uma medida de conhecimento implícito (e.g., Bowles, 2011; Ellis, 2005; Ellis et al., 2009; Godfroid et al., 2015) ou, pelo menos, de conhecimento explícito automatizado (e.g., Kim & Nam, 2016; Suzuki & DeKeyser, 2017), i.e., conhecimento funcionalmente equivalente ao conhecimento implícito, mas que, ao contrário deste, não é completamente inconsciente.

3.2.3. **Análise estatística**

A análise estatística foi conduzida em R (versão 4.3.3), usando a função *lmer* do pacote *lme4* para aplicação dos modelos lineares de efeitos mistos, e o pacote *ggplot2* para a visualização dos dados. Optámos por realizar apenas análises estatísticas intragrupos, evitando comparações diretas com o grupo monolingue, com o objetivo de contornar a chamada *comparative fallacy* (Bley-Vroman, 1983; White, 2003) — ou seja, a falha em reconhecer as interlínguas como sistemas autónomos, decorrente de uma ênfase excessiva na comparação direta entre aprendentes de L2 e falantes monolingues. Seguindo White (2003), consideramos que uma abordagem mais adequada consiste em analisar se, dentro do mesmo grupo de L2, uma opção *a* (possível na língua-alvo) é tratada de forma significativamente diferente de uma opção *b* (impossível). Se tais distinções forem observadas, isto indica que a gramática da interlíngua representa a distinção relevante, ainda que o grau com que os aprendentes o manifestem possa divergir do observado nos falantes nativos. Além disso, as análises intragrupos permitem uma caracterização mais fina do desempenho dos aprendentes em cada nível, contribuindo para uma resposta mais precisa à questão de investigação.

Foram realizados três conjuntos de análises. No primeiro, para os dados de cada grupo, foi corrido um modelo linear de efeitos mistos, com *participantes* e *itens* especificados como efeitos aleatórios, e *tipo de sujeito* e *referencialidade* como efeitos fixos. Incluíram-se ainda declives aleatórios por participante para ambas as variáveis fixas, uma vez que são variáveis intrasujeito. No segundo conjunto de análises, para cada um dos quatro sujeitos testados (2.ª pessoa, 3.ª pessoa [+ humano], 3.ª pessoa [- humano], expletivo), foi analisado, em cada grupo, se os participantes distinguiam entre sujeitos plenos e nulos. Para tal, aplicou-se um modelo linear de efeitos mistos, incluindo *participantes* e *itens* como efeitos aleatórios, *tipo de sujeito* como efeito fixo, e

declives aleatórios por participante para esta variável. Por fim, o terceiro conjunto de análises examinou possíveis diferenças entre sujeitos nulos com base na referencialidade, comparando, dentro de cada grupo, os vários tipos de sujeitos nulos. As comparações relevantes para o presente estudo são: 2p vs. 3p [+ humano]; 3p [+ humano] vs. 3p [- humano]; 3p [- humano] vs. expletivo; 2p vs. expletivo. Nesta análise, foi aplicado um modelo linear de efeitos mistos, com *participantes* e *itens* como efeitos aleatórios, *referencialidade* como efeito fixo, e declives aleatórios por participante para esta variável fixa.

Em todas as análises, a variável *tipo de sujeito* foi codificada contrastivamente, com os valores -.5 para sujeito nulo e .5 para sujeito pleno. A variável *referencialidade*, com quatro níveis, foi codificada usando codificação contrastiva simples (UCLA Statistical Consulting Citation, 2011), que compara cada nível da variável categórica com um nível de referência. No primeiro conjunto de análises, o nível de referência foi a 2.ª pessoa.

Como a função *lmer* apenas gera valores de *t*, usámos este valor como medida de significância estatística. Um efeito fixo foi considerado estatisticamente significativo sempre que o valor absoluto de *t* é igual ou superior a 2.00.

4. Resultados

Na tarefa de juízos de aceitabilidade rápida, o grupo de falantes nativos de inglês apresenta um efeito principal de tipo de sujeito, aceitando sujeitos pronominais plenos e rejeitando a sua omissão em todas as condições experimentais, conforme ilustrado na Figura 6 e na Tabela 3. Não se observam efeitos de referencialidade nem nenhuma interação entre este fator e o tipo de sujeito. Deste modo, a referencialidade não influencia a aceitabilidade de sujeitos em inglês nas gramáticas nativas.

Figura 6

Aceitação de Sujeitos Nulos e Plenos por Falantes Nativos de Inglês

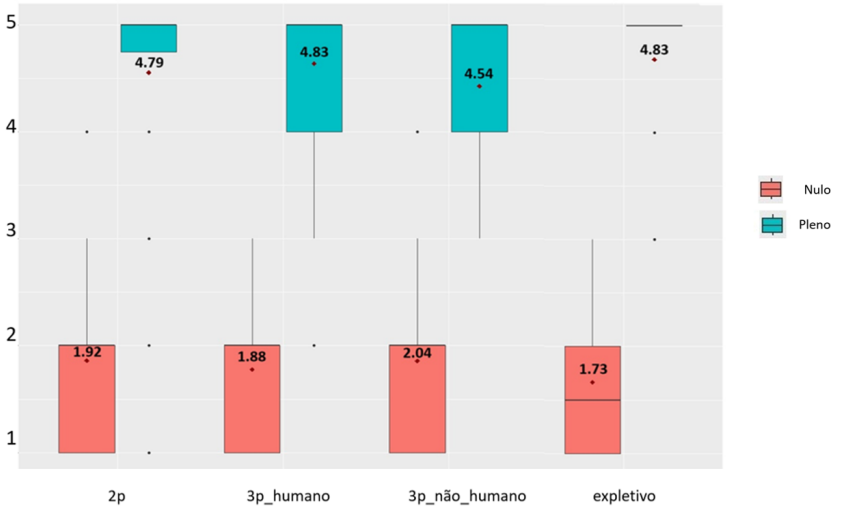


Tabela 3

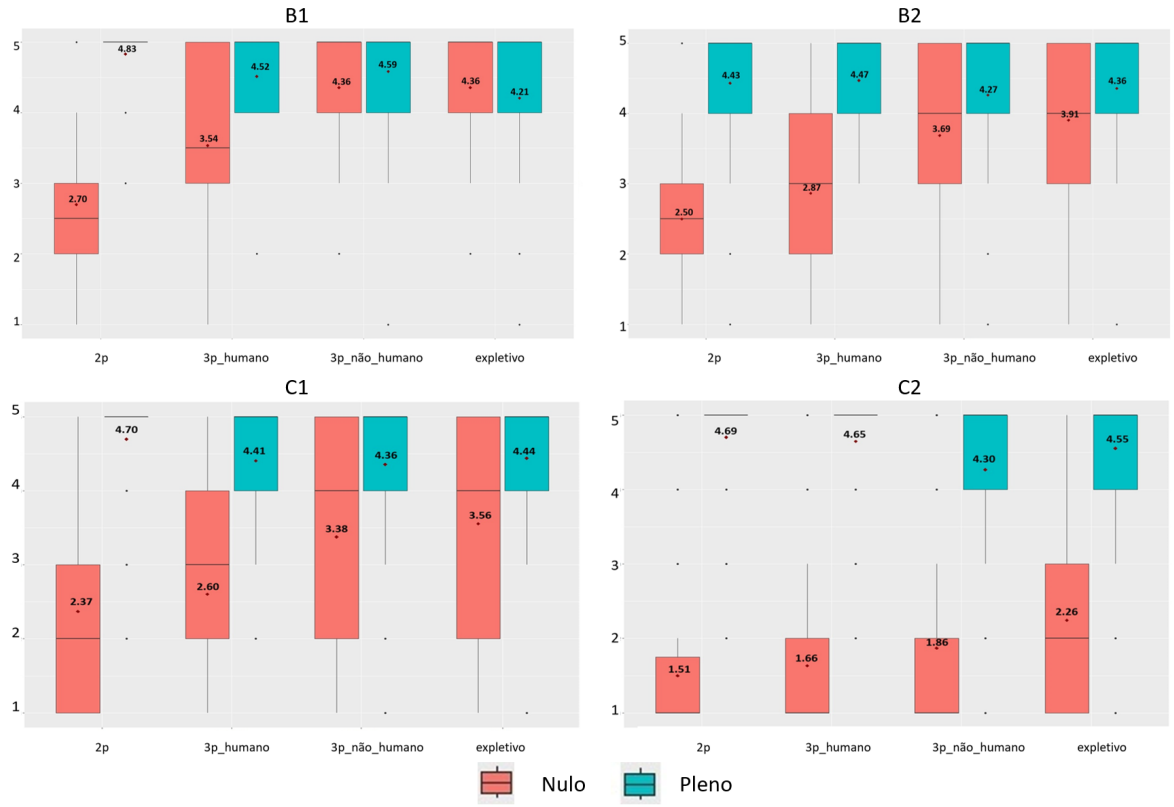
Resultados do Modelo Linear de Efeitos Mistos para o Grupo Nativo

	Estimativa	**Erro Padrão**	***t***
Tipo de sujeito	2.694	.328	8.214*
Referencialidade (2p vs. 3p_hum)	.0001	.149	.000
Referencialidade (2p vs. 3p_não_hum)	.0025	.164	.015
Referencialidade (2p vs. expl)	-.039	.161	-.244
Tipo de sujeito * Referencialidade (2p vs. 3p_hum)	.167	.258	.645
Tipo de sujeito * Referencialidade (2p vs. 3p_não_hum)	-.315	.315	-1.001
Tipo de sujeito * Referencialidade (2p vs. expl)	.311	.276	1.125

À semelhança dos falantes nativos, os aprendentes de inglês L2 exibem um efeito de tipo de sujeito em todos os níveis de proficiência. No entanto, diferentemente deles, revelam sensibilidade ao fator referencialidade na aceitação de sujeitos plenos e nulos em inglês. A influência deste fator varia em função do nível de proficiência, como ilustra a Figura 7.

Figura 7

Aceitação de Sujeitos Nulos e Plenos pelos Grupos de Aprendentes de Inglês L2



No nível B1, observa-se não só um efeito principal de tipo de sujeito, mas também uma interação significativa entre este fator e a referencialidade em todas as análises comparativas feitas (cf. Tabela 3). Verifica-se ainda um efeito principal de referencialidade, exceto na comparação entre sujeitos de 2.ª e 3.ª pessoas [+humanas]. As análises realizadas por condição revelam que, na presença de sujeitos humanos, os aprendentes de nível B1 preferem sujeitos pronominais plenos a nulos (2p: estimativa = 1.4250, EP = .3673, *t* = 3.879; 3p humano: estimativa = .9815, EP = .2772, *t* = 3.541), mas esta preferência não se verifica quando o sujeito é [-humano] ou expletivo. Nestes casos, não há diferenças significativas entre a aceitação de sujeitos nulos e plenos (3p não humano: estimativa = .2145, EP = .2822, *t* =.76; expletivo: estimativa = -.1200, EP = .5684; *t* = -.2119815). Por seu lado, a análise do subconjunto de respostas referentes aos sujeitos nulos indica que a sua aceitação no nível B1 varia de acordo com a referencialidade do sujeito: sujeitos nulos de 2.ª pessoa são menos aceites do que os de 3.ª pessoa [+humana] (estimativa = -.8283, EP = .3810, *t* = -2.174), que, por sua vez, são menos aceites do que os de 3.ª pessoa [-humana] (estimativa = -.8304, EP = .3031, *t* = -2.74). Não se observam diferenças entre sujeitos não humanos e expletivos (estimativa = -.09661, EP = .34966, *t* = -.276).

No nível B2, os aprendentes continuam a exibir efeitos principais de tipo de sujeito e de referencialidade, bem como uma interação significativa entre estes dois fatores (cf. Tabela 3). O fator referencialidade apenas deixa de ser estatisticamente relevante nas comparações entre sujeitos humanos de 2.ª e 3.ª pessoas. Nestas condições, ao contrário do observado no nível B1, os aprendentes de nível B2 não exibem diferenças na aceitação de sujeitos nulos, mantendo uma preferência consistente por sujeitos plenos. Diferentemente dos aprendentes de nível B1, os de B2 têm uma preferência por sujeitos plenos também na condição de 3.ª pessoa [-humana] (estimativa = .6445, EP = .3167, *t* = 2.035). No entanto, tal como no nível anterior, não se observa uma diferença significativa entre sujeitos plenos e nulos na condição de expletivo (estimativa = .4546, EP = .2873, *t* = 1.582). A análise restrita às respostas sobre sujeitos nulos mostra que, à semelhança dos aprendentes de nível B1, os de nível B2 exibem um efeito de animacidade na aceitação de sujeitos nulos referenciais de 3.ª pessoa, admitindo sujeitos nulos [-humanos] significativamente mais do que [+humanos] (estimativa = .7823, EP = .3383, *t* = 2.312). Também neste nível não se observa uma diferença significativa na aceitação de sujeitos nulos [-humanos] e expletivos (estimativa = -.2636, EP = .1971, *t* = -1.338).

Tabela 3

Resultados do Modelo Linear de Efeitos Mistos para os Grupos B1 e B2

	B1			B2		
	Estimativa	**Erro Padrão**	***t***	**Estimativa**	**Erro Padrão**	***t***
Tipo de sujeito	2.1333	.3259	6.546*	1.9413	.2464	7.877*
Referencialidade (2p vs. 3p_hum)	.2587	.1659	1.560	.1877	.1454	1.290
Referencialidade (2p vs. 3p_não_hum)	.7030	.2324	3.025*	.4935	.1784	2.766*
Referencialidade (2p vs. expl)	.4597	.1841	2.496*	.6776	.1405	4.823*
Tipo de sujeito * Referencialidade (2p vs. 3p_hum)	-1.1330	.3692	-3.069*	-.2902	.2977	-.975
Tipo de sujeito * Referencialidade (2p vs. 3p_não_hum)	-1.9140	.3781	-5.063*	-1.2766	.3220	-3.964*
Tipo de sujeito * Referencialidade (2p vs. expl)	-2.2443	.3817	-5.879*	-1.4645	.3370	-4.345*

No nível C1, observa-se um efeito principal de tipo de sujeito e uma redução do efeito de referencialidade, que apenas se mantém significativo na comparação entre sujeitos de 2.ª pessoa e expletivos, os dois extremos da hierarquia da referencialidade (cf. Tabela 4). Há, ainda assim, uma interação significativa entre tipo de sujeito e referencialidade em todas as comparações, com exceção da que envolve sujeitos humanos de 2.ª e 3.ª pessoas. O grupo C1, à semelhança do B2, prefere claramente sujeitos plenos a nulos em todas as condições (todos os $t \geq 4.201$), exceto na de sujeito expletivo, em que a diferença entre sujeito pleno e nulo apenas se aproxima de significância estatística (estimativa = .8814, EP = .4880, $t = 1.806$). Tal como no nível B2, no C1, observa-se um efeito de animacidade na aceitação de sujeitos nulos de 3.ª pessoa, sendo os [- humanos] mais aceites do que os [+ humanos] (estimativa = .7609, EP = .2050, $t = 3.713$). Não se encontram diferenças na aceitação de sujeitos nulos expletivos e [- humanos] (estimativa = .1714, EP = .2341, $t = -.732$).

Por fim, no nível C2, o desempenho dos aprendentes aproxima-se do dos falantes nativos de inglês. Tal como eles, os aprendentes exibem um efeito robusto de tipo de sujeito (cf. Tabela 4), aceitando sujeitos plenos e rejeitando nulos em todas as condições, incluindo a de sujeito expletivo (estimativa = 2.2736, EP = .4152, $t = 5.476$). Além disso, não apresentam efeitos de referencialidade, apenas se registando uma interação significativa entre este fator e o tipo de sujeito na comparação entre sujeitos de 2.ª pessoa e expletivos (Tabela 4). Esta interação decorre do facto de os aprendentes rejeitarem de forma mais categórica sujeitos nulos de 2.ª pessoa do que expletivos nulos (estimativa = -.5949, EP = .2816, $t = -2.113$). Este é o único ponto de divergência relativamente ao grupo de controlo. Tal como este grupo, e ao contrário do que se verifica nos níveis anteriores, os aprendentes com nível C2 não exibem diferenças na aceitação dos diferentes tipos de sujeitos nulos de 3.ª pessoa (+ humano vs. – humano: estimativa = -.2120, EP = .2249, $t = -.943$; - humano vs. expletivo: estimativa = -.3956, EP = .3436, $t = -1.151$).

Tabela 4

Resultados do Modelo Linear de Efeitos Mistos para os Grupos C1 e C2

	C1			C2		
	Estimativa	**Erro Padrão**	***t***	**Estimativa**	**Erro Padrão**	***t***
Tipo de sujeito	2.33092	.35538	6.559*	3.20370	.30814	10.397*
Referencialidade (2p vs. 3p_hum)	-.02257	.21922	-.103	.05704	.19519	.292
Referencialidade (2p vs. 3p_não_hum)	.33678	.26650	1.264	-.03467	.21316	-.163
Referencialidade (2p vs. expl)	.46132	.23387	1.973*	.31968	.22420	1.426
Tipo de sujeito * Referencialidade (2p vs. 3p_hum)	-.54262	.44683	-1.214	-.22520	.38991	-.578
Tipo de sujeito * Referencialidade (2p vs. 3p_não_hum)	-1.34458	.48126	-2.794*	-.61008	.39175	-1.568
Tipo de sujeito * Referencialidade (2p vs. expl)	-1.44772	.50190	-2.885*	-.93565	.42524	-2.200*

5. Discussão e conclusões

Os resultados do presente estudo demonstram que, ao contrário dos falantes nativos de inglês, os aprendentes portugueses de inglês L2 não evidenciam uma preferência clara por sujeitos pronominais plenos, nem rejeitam sistematicamente sujeitos nulos, o que está em linha com estudos anteriores que apontam para a existência de interferência persistente da L1 na aquisição de línguas de sujeito obrigatório por falantes de línguas de sujeito nulo (cf. Secção 2.2). O desempenho destes aprendentes é influenciado tanto pelo nível de proficiência na L2 quanto pelo fator referencialidade.

Relativamente à distinção entre sujeitos plenos e nulos, os nossos dados revelam que os aprendentes não são sensíveis a esta distinção quando os sujeitos são [-humanos] ou expletivos, que são justamente os mais próximos do extremo [-referencial] da HR. Este padrão de comportamento persiste para os sujeitos expletivos também nos níveis B2 e C1. Apenas no nível C2 se observa uma preferência clara por sujeitos plenos em detrimento de nulos, independentemente do seu valor referencial.

Quanto aos sujeitos nulos, observa-se que os aprendentes não têm um comportamento homogéneo nos diferentes níveis de proficiência. No nível B1, há um efeito da pessoa gramatical na aceitação de sujeitos nulos [+humanos]: os de 3.ª pessoa são mais aceites do que os de 2.ª pessoa, que ocupam uma posição mais próxima do extremo [+referencial] da HR. Esta diferença, no entanto, deixa de ser observada nos níveis seguintes, em que todos os tipos de sujeitos nulos [+ humanos] são rejeitados. No caso dos sujeitos nulos referenciais de 3.ª pessoa, os dados dos níveis B1 a C1 revelam um efeito de animacidade: os sujeitos nulos [-humanos] são mais aceites do que os [+humanos], que, de acordo com a HR, têm maior grau de referencialidade. Em nenhum dos níveis testados foram observadas diferenças significativas entre sujeitos nulos de 3.ª pessoa [-humano] e sujeitos expletivos.

Em conjunto, estes resultados indicam que o conhecimento de que o sujeito pronominal pleno é a única forma aceitável em inglês é adquirido tardiamente, havendo atrasos de desenvolvimento particularmente no caso dos sujeitos expletivos, localizados no polo [-referencial] da HR. Os resultados mostram ainda que, nas interlínguas, a eliminação de sujeitos nulos e a consequente substituição por sujeitos realizados ocorrem de forma gradual, segundo a seguinte escala: 2.ª pessoa > 3.ª pessoa [+ humano] > 3.ª pessoa [- humano] > expletivo. Assim, a reestruturação das gramáticas de L2 progride do extremo [+referencial] para o extremo [-referencial] da HR.

Uma questão suscitada por estes resultados diz respeito à possível relação entre a produtividade das formas nulas na L1 e a sua persistência na L2: será que as formas nulas mais difíceis de eliminar no inglês L2 são precisamente aquelas que são mais produtivas na L1? Por outras palavras, estará a influência da L1 na base do papel desempenhado pela referencialidade?

Em PE, os dados disponíveis indicam que os sujeitos nulos de 1.ª e 2.ª pessoas podem ocorrer na maioria dos contextos sem restrições, como no exemplo (3), enquanto sujeitos nulos de 3.ª pessoa só são legítimos se existir um referente no contexto situacional ou um antecedente linguístico acessível no contexto discursivo que permitam recuperar o antecedente do sujeito nulo (Lobo, 2013), como mostra o contraste entre (4a) e (4b). Além disso, a alternância entre sujeitos de 3.ª pessoa nulos e plenos não parece ser sensível à animacidade do antecedente: o sujeito nulo recupera antecedentes na posição de sujeito, enquanto o pleno retoma antecedentes na posição de objeto, independentemente dos traços de animacidade, como se mostra em (5) (Fiéis et al., 2022).

(3) Vou à praia.

(4a) ?Vai à praia.
(4b) O João está todo contente. Vai à praia.
(Lobo, 2013, p. 2331)

(5a) O menino$_i$ viu o brinquedo quando [-]$_i$ caiu da cadeira.
(5b) O menino viu o brinquedo$_i$ quando ele$_i$ caiu da cadeira.
(Fiéis et al., 2022, p. 9)

Deste modo, o desempenho dos aprendentes portugueses de inglês L2 com sujeitos nulos não pode ser explicado apenas com base na influência da L1. A HR parece orientar a reestruturação das interlínguas, mesmo quando os traços relevantes desta hierarquia, como a animacidade, não são determinantes para a alternância entre formas nulas e expressas na L1.

Em conclusão, os nossos resultados revelam semelhanças entre a mudança diacrónica (a partir do PB) e a aquisição de L2 (nos pares inglês L2 – PE L1). Nos dois casos, as mudanças que envolvem sujeitos nulos não ocorrem de forma abrupta, mas sim gradualmente, orientadas por um princípio subjacente comum — a referencialidade. Tal como predito pela HR e a Hipótese do Mapeamento Implicacional, quanto menor a referencialidade, maior a possibilidade de ocorrência de uma forma nula. Verifica-se ainda que, tanto na mudança diacrónica como na aquisição de L2, uma variante nula num determinado ponto da HR implica a possibilidade de variantes nulas à sua esquerda.

Assim, os resultados sugerem que a referencialidade pode orientar o desenvolvimento linguístico relativo à pronominalização em diferentes contextos. De forma mais ampla, os dados reforçam a proposta de Meisel (2011) no sentido de uma abordagem integrada do desenvolvimento linguístico, demonstrando que fenómenos tradicionalmente estudados de forma separada — como a aquisição de L2 e a mudança linguística — podem envolver mecanismos comuns de reestruturação gramatical.

Financiamento

Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e Tecnologia, I.P., no âmbito do projeto UID/03213 – Centro de Linguística da Universidade NOVA de Lisboa (CLUNL).

Referências

Al-Kasey, M., & Pérez-Leroux, A. T. (1998). Second language acquisition of Spanish null subjects. In Flynn, S., Martohardjono, G. & O'Neil, W. (Eds.), *The generative study of second language acquisition* (pp. 161–185). Routledge.

Ariel, M. (1990). *Accessing antecedents*. Routledge.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29–87). John Benjamins. https://doi.org/10.1075/hcp.8.04ari

Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics, 46*(2), 273–330. https://doi.org/10.1017/S0022226709990230

Ballester, E. P. (2013). Adult instructed SLA of English subject properties. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 58*(3), 465–486. https://doi.org/10.1017/S0008413100002668

Barbosa, P., Duarte, M. E., & Kato, M. (2005). Null subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics, 4*, 11–52.

Belletti, A., Bennati, E., & Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory*, *25*(4), 655–689. https://doi.org/10.1007/s11049-007-9020-6

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning, 33*(1), 1–17. https://doi.org/10.1111/j.1467-1770.1983.tb00983.x

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners tell us? *Studies in Second Language Acquisition, 33*(2), 247–271. https://doi.org/10.1017/S0272263110000722

Conselho da Europa. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.

Conselho da Europa. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume with new descriptors*. Council of Europe Publishing. https://www.coe.int/en/web/common-european-framework-reference-languages

Cyrino, S., Duarte, E., & Kato, M. (2000). Visible subjects and invisible clitics in Brazilian Portuguese. In Kato, M. & Negrão, E. (Eds.), *Brazilian Portuguese and the null subject parameter* (pp. 55–73). Iberoamericana.

Donaldson, B. (2022). Connecting language change with second language acquisition. In Geslin, K. (Ed.), *The Routledge handbook of second language acquisition and sociolinguistics* (pp. 174–185). Routledge.

Duarte, I., & Silva, M.C.F. (2016). The Null Subject Parameter and the structure of the sentence in European and Brazilian Portuguese. In W. Wetzels, J. Costa, & S. Menuzzi (Eds.), *The handbook of Portuguese linguistics*. Wiley Blackwell. https://doi.org/10.1002/9781118791844.ch13

Duarte, M. E. (1995). *A perda do princípio "evite pronome" no português brasileiro.* [Tese de doutoramento, Universidade Estadual de Campinas – Unicamp, Brasil].

Duarte, M. E. L., & Varejão, F. (2013). Null subjects and agreement marks in European and Brazilian Portuguese. *Journal of Portuguese Linguistics, 12*(2), 101–123.

Duarte, M. E., Mourão, G., & Mendonça, H. (2012). Os sujeitos de terceira pessoa: Revisitando Duarte (1993). In M. E. Duarte (Ed.), *O sujeito em peças de teatro (1833-1992): Estudos diacrônicos* (pp. 21–44). Parábola Ed.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. Studies. *Second Language Acquisition*, 27(2), 141–172. https://doi.org/10.1017/S0272263105050096
Ellis, R., Loewen, S., & Erlam, R. (2009). Implicit and explicit knowledge in second language learning, testing and teaching. In R. Ellis, S. Loewen, & R. Erlam (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Multilingual Matters.
Fiéis, A., Madeira, A., & Teixeira, J. (2022). Microvariation in the resolution of pronominal subjects in Romance: European Portuguese *vs*. Italian. *Isogloss. Open Journal of Romance Linguistics 8(5)*, 1–20. http://dx.doi.org/10.5565/rev/isogloss.220
Fukumura, K. & van Gompel, R. P. G. (2010). The effect of animacy on the choice of referring expression. *Language and Cognitive Processes*, *26*(10), 1472–1504. https://doi.org/10.1080/01690965.2010.506444
Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition,* 37(2), 269–297. https://doi.org/10.1017/S0272263114000850
Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition, 27*(2), 129–140. https://doi.org/10.1017/S0272263105050084
Judy, T. (2011) L1/L2 parametric directionality matters: More on the Null Subject Parameter in L2 acquisition. *EUROSLA Yearbook*, *1*, 165–190.
Judy, T., & Rothman, J. (2010). From a superset to a subset grammar and the semantic compensation hypothesis: Subject pronoun and anaphora resolution evidence in L2 English. In *BUCLD 34: Proceedings of the 34th annual Boston University Conference on Language Development* (pp. 197–208). Cascadilla Press.
Kato, M. A., & Duarte, M. E. L. (2014). Restrições na distribuição de sujeitos nulos no português brasileiro. *Veredas: Revista de Estudos Linguísticos*, *18*(1), 1–22.
Kim, J., & Nam, H. (2016). Development of implicit and explicit knowledge in L2 syntax: The role of age and working memory. *Language Learning, 66*(2), 355–389. https://doi.org/10.1111/lang.12150
Liceras, J. M. (1989). On some properties of the "pro-drop" parameter: looking for missing subjects in non-native Spanish. In Gass, S. M. & Schachter, J. (Eds.), *Linguistic perspectives on second language acquisition* (pp. 109–133). Cambridge University Press.https://doi.org/10.1017/CBO9781139524544.009
Lobo, M. (2013). Sujeito nulo: sintaxe e interpretação. In E. B. P. Raposo, M. F. B. Nascimento, M. A. C. Mota, L. Segura, & A. Mendes (Orgs.), *Gramática do português* (Vol. 2, pp. 2309–2335). Fundação Calouste Gulbenkian.
Lozano, C. (2006). Focus and split intransitivity: The acquisition of word order alternations in non-native Spanish. *Second Language Research*, *22*(2), 145–187. https://doi.org/10.1191/0267658306sr265oa
Meisel, J. (2011). Bilingual language acquisition and theories of diachronic change: Bilingualism as cause and effect of grammatical change. *Bilingualism: Language and Cognition*, *14*, 121–145. https://doi.org/10.1017/S1366728910000143
Morgado, S., Luegi, P., & Lobo, M. (2018). Efeitos de animacidade do antecedente na resolução de pronomes sujeito. *Revista da Associação Portuguesa de Linguística*, *4*, 190–205. https://doi.org/10.26334/2183-9077/rapln4ano2018a40
Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y
Pérez-Leroux, A. T., & Glass, W. R. (1999). Null anaphora in Spanish second language acquisition: Probabilistic versus generative approaches. *Second Language Research*, *15*(2), 220–249. https://doi.org/10.1191/026765899676722648
Prentza, A., & Tsimpli, I. (2013). The interpretability of features in second language acquisition: Evidence from null and postverbal subjects in L2 English. *Journal of Greek Linguistics, 13*, 323–365.

Rizzi, L. (1982). *Issues in Italian syntax*. Foris Publications.
Rothman, J. (2009). Pragmatic deficits with syntactic consequences?: L2 pronominal subjects and the syntax–pragmatics interface. *Journal of Pragmatics*, *41*(5), 951–973. https://doi.org/10.1016/j.pragma.2008.07.007
Rothman, J., & Iverson, M. (2007). The syntax of null subjects in L2 Spanish: Comparing two L1–L2 groups. In *BUCLD 31: Proceedings of the 31st Boston University Conference on Language Development* (pp. 531–542). Cascadilla Press.
Sorace, A., & Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, *22*(3), 339–368. https://doi.org/10.1191/0267658306sr271oa
Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning, 67*(4), 747–790. https://doi.org/10.1111/lang.12241
Teixeira, J. (2019). From a Romance null subject grammar to a non-null subject grammar: The syntax of pronominal subjects in advanced and near-native English. In I. Feldhousen, M. Elsig, I. Kuchenbrandt & M. Neuhaus (Eds.), *Romance languages and linguistic theory 15: Selected papers from 'Going Romance' Frankfurt 2016* (pp. 256–274). John Benjamins.
UCLA Institute for Digital Research and Education. (2011). *Dummy coding in regression*. UCLA Statistical Consulting Group. https://stats.oarc.ucla.edu/r/library/dummy-coding-in-regression/
White, L. (1985). The "pro-drop" parameter in adult second language acquisition. *Language Learning*, 35(1), 47–61. https://doi.org/10.1111/j.1467-1770.1985.tb01014.x
White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press.

(RE)CONTO: Instrumento de avaliação de narrativas orais para crianças falantes de Português Europeu

(RE)CONTO: Tool to assess oral narratives in European Portuguese-speaking children

Stéphanie Vaz[1], *Maria Lobo*[1], *Marisa Lousada*[2]

[1] Centro de Linguística da Universidade NOVA de Lisboa – NOVA/FCSH
[2] Universidade de Aveiro/ RISE-Health

Abstract

Narrative competencies are determinant for the academic success of children and adolescents. The assessment of these competencies – including comprehension and production – is important to complement the diagnosis of children with language impairment, and to plan interventions in clinical and educational contexts. We present in this paper a tool to assess narrative competencies developed for children aged 6 to 12 speakers of European Portuguese. The tool comprises telling and retelling tasks, as well as comprehension questions (literal, inferential and critical comprehension). We report the results obtained by a sample of 521 children with typical development from eight regions of Portugal and by an exploratory group of 17 children with Developmental Language Disorder, considering the type of task, macrostructural and microstructural properties, group and age range. Additionally, we present some psychometric properties of the tool, which attest to its validity and reliability.

Keywords: language assessment, narrative competencies, macrostructure, microstructure.

Resumo

As competências narrativas são determinantes para o sucesso académico de crianças e adolescentes. A avaliação destas competências (incluindo a compreensão e a produção) é importante para estabelecer o diagnóstico de crianças com perturbações da linguagem, bem como para planear a intervenção em contextos clínicos e educativos. Apresenta-se neste artigo um instrumento de avaliação de competências narrativas desenvolvido para crianças entre os 6 e os 12 anos falantes de português europeu. O instrumento inclui tarefas de conto e de reconto, bem como perguntas de compreensão (literal, inferencial e crítica). São reportados resultados obtidos por uma amostra de 521 crianças com desenvolvimento típico de oito regiões de Portugal, bem como de um grupo de 17 crianças com Perturbação do Desenvolvimento da Linguagem, considerando o tipo de tarefa, parâmetros macroestruturais e microestruturais, e faixa etária. Apresentam-se ainda algumas propriedades psicométricas do instrumento, que comprovam a validade e fiabilidade do instrumento.

Palavras-chave: avaliação da linguagem, competências narrativas, macroestrutura, microestrutura.

1. Introdução

A avaliação de competências narrativas orais fornece dados importantes para diagnosticar perturbações da linguagem com alterações na compreensão e na produção e para planear a intervenção em contextos clínicos e educativos. Apesar de haver vários instrumentos estandardizados de avaliação de competências narrativas para outras línguas, não existe nenhum instrumento para o português europeu (PE), como mostra a revisão sistemática levada a cabo por Viana et al. (2017). Neste artigo, apresentamos um instrumento de avaliação de competências narrativas desenvolvido para o PE, destinado a crianças entre os 6 e os 12 anos, bem como as propriedades psicométricas do instrumento que comprovam a sua validade e fiabilidade.

Nesta primeira secção, fazemos uma revisão sobre a relevância da avaliação da competência narrativa, sobre os diferentes parâmetros de avaliação e sobre os instrumentos de avaliação de narrativas disponíveis para outras línguas. Na secção 2., apresentamos detalhadamente o instrumento original que desenvolvemos para o PE, caracterizamos a amostra de participantes e os procedimentos usados na recolha e na codificação e cotação dos dados. Na secção 3, explicitamos os resultados obtidos e as principais medidas de validade e de fiabilidade. Na secção 4, discutimos os resultados e sistematizamos as principais conclusões.

1.1. Relevância da avaliação de competências narrativas

Ainda que seja difícil chegar a uma definição de narrativa suficientemente restrita, mas também suficientemente flexível para dar conta dos textos que habitualmente se agrupam nesta categoria, diferentes autores convergem na ideia de que as narrativas são relatos de sequências de eventos encadeados temporalmente ordenados (Rudrum, 2005; Silva, 2015). São muitos e variados os textos que se classificam como narrativos por terem predominantemente características de uma narrativa. Como discute Silva (2015), a designação "narrativo" pode ser usada para classificar quer os textos globalmente, de acordo com critérios fundamentalmente de natureza cognitiva e funcional, quer partes de um texto que pode ser ou não predominantemente narrativo – veja-se a noção de "sequência narrativa" introduzida por Jean-Michel Adam (Adam, 1992), que se aplica a subpartes de um texto. Entre os textos globalmente caracterizados como sendo narrativos encontram-se relatos de vida pessoal, relatos ficcionais ou relatos de factos históricos, que podem ser moldados de acordo com parâmetros socioculturalmente estabelecidos, como os que definem os géneros memória, epopeia, fábula, parábola, romance histórico, romance policial, entre muitos outros (Adam, 1992). A literatura para a infância contém numerosos exemplos de textos narrativos, geralmente ficcionais. As crianças estão, pois, expostas a textos narrativos desde muito cedo e são confrontadas com textos narrativos de complexidade variável ao longo do seu percurso escolar (Pereira & Azevedo, 2005).

As orientações curriculares do ensino básico português (veja-se as *Aprendizagens Essenciais* para a disciplina de Português, disponibilizadas pela Direção-Geral de Educação) preveem que se trabalhe quer aspetos da produção, quer da compreensão de narrativas. As competências desenvolvidas oralmente podem ser transpostas para a escrita, contribuindo para a produção de textos com maior qualidade e para uma melhor compreensão da leitura (Borges, 2015; Pereira, 2010). Na verdade, a capacidade de compreender e de produzir narrativas orais bem estruturadas e linguisticamente elaboradas tem mostrado ser preditora de competências de literacia (Wellman et al., 2011).

Em contextos clínicos, a avaliação da produção de narrativas orais pode ser um meio complementar de diagnóstico de crianças com perturbações da linguagem, permitindo distinguir crianças com e sem perturbações da linguagem (Cain & Oakhill, 2006; Kraljevic et al., 2020; Norbury & Bishop, 2003; Paul & Smith, 1993). Adicionalmente, a intervenção com base em narrativas orais pode ser uma ferramenta terapêutica para auxiliar crianças com perturbação da linguagem (Bishop, 2009; Favot et al., 2021; Petersen, 2011; Westby, 2005), seguindo modelos de avaliação dinâmica (de Villiers & de Villiers, 2010), ou até para trabalhar aspetos emocionais e criativos (Sterponi, 2019). A importância da avaliação de competências narrativas no contexto clínico é reforçada por resultados como os do estudo de Feagans e Applebaum (1986), que mostram que a

capacidade de compreender e recontar uma narrativa é um preditor de sucesso académico mais forte do que as competências sintáticas e de vocabulário em crianças com perturbações da linguagem.

Tendo isto em conta, é inequívoca a importância de avaliar as competências de produção e de compreensão de narrativas quer em contextos escolares, quer em contextos clínicos. Na secção 1.2., detalham-se as diferentes dimensões habitualmente consideradas na avaliação de competências narrativas.

1.2. **Parâmetros de avaliação de competências narrativas**

Como já referido, entre as competências narrativas, podem considerar-se competências de produção, mas também competências de compreensão. Quando se considera a produção, é habitual avaliar-se aspetos macroestruturais e microestruturais (Hughes et al., 1997; van Dijk, 1980; Westby, 2005).

Os aspetos macroestruturais dizem respeito à forma como a narrativa está estruturada. Há vários modelos de análise da macroestrutura narrativa (Gagarina et al., 2012), mas podem considerar-se alguns elementos que geralmente caracterizam uma narrativa. Uma narrativa clássica começa com uma abertura, que geralmente introduz a situação inicial, estabelece a localização espácio-temporal em que decorrem os eventos e apresenta a personagem principal. À situação inicial, geralmente caracterizada por um estado de equilíbrio, segue-se um evento que desestabiliza a situação inicial. Finalmente, a situação problemática é resolvida e é reestabelecido o equilíbrio. No desfecho da narrativa, repõe-se a situação de equilíbrio e pode apresentar-se uma lição, como é característico das fábulas (Adam, 1992).

A macroestrutura de um texto narrativo pode ser mais simples, contendo apenas um episódio ou mais complexa, contendo vários episódios (situação inicial – complicação – resolução), e pode integrar apenas sequências de tipo narrativo ou integrar também sequências com outras características, como as sequências descritivas, em que se caracteriza uma personagem ou um espaço, ou as sequências dialogais, em que se estabelece uma conversação entre personagens, ou até sequências argumentativas, em que se fundamenta uma opinião (Adam, 1992). De facto, muitos textos narrativos contêm, para além das sequências narrativas predominantes, sequências descritivas ou dialogais. A análise da macroestrutura de uma narrativa deve, pois, atender a estes diferentes elementos que integram a narrativa, desde a abertura até ao desfecho, considerando os vários episódios e as várias sequências que a integram (Gagarina et al., 2012).

A microestrutura de um texto narrativo corresponde à forma como são expressas as ideias através de mecanismos linguísticos. Num texto narrativo, serão importantes as diferentes construções que servem para a localização e sequenciação temporal dos eventos, como expressões adverbiais temporais e aspetuais, orações subordinadas temporais, tempos verbais, bem como as expressões que marcam relações de causalidade entre os eventos. A microestrutura contempla ainda os mecanismos de coesão referencial, que permitem estabelecer relações anafóricas entre as diferentes referências a uma mesma entidade ao longo do texto. Para além disso, podem também ser incluídos indicadores de complexidade linguística, como a densidade oracional e a diversidade lexical, e de produtividade linguística, como a extensão medida em número de palavras ou de orações (Gagarina et al., 2012).

Em contexto de avaliação, a produção de narrativas pode ser desencadeada de várias formas: através de uma tarefa de conto, ou de uma tarefa de reconto, em ambos os casos, com ou sem o suporte de imagens. A investigação tem mostrado que o tipo de tarefa e o tipo de suporte influenciam a qualidade das narrativas orais ou escritas produzidas pelas crianças (Cruz, 2011; Roch et al., 2016; Shapiro & Hudson, 1991; Silva & Spinillo, 2000; Spinillo, 1993; Veloso, 1999; e.o.). Diferentes estudos mostram que os desempenhos das crianças tendem a ser melhores em tarefas de reconto, ainda que isso possa levar a alguma sobrevalorização do seu conhecimento linguístico. O tipo de suporte (imagem única ou sequência de imagens ou ausência de imagens) também tem influência nas produções. A presença de imagens ajuda a criança a reconstruir a narrativa e funciona como auxiliar de memória na tarefa de reconto, mas pode também ter como efeito levar a criança a especificar menos as suas produções. Assim, a escolha da tarefa e do suporte é um passo importante na construção de um instrumento de avaliação de competências narrativas.

Finalmente, pode-se incluir também a compreensão, considerando diferentes processos, entre os quais estão a compreensão literal, a compreensão inferencial e a compreensão crítica (Giasson, 2005; Viana et al., 2010; e.o.). A compreensão literal diz respeito à capacidade de interpretar informação explícita; a compreensão inferencial diz respeito à capacidade de interpretar informação que não está explícita, mas que pode ser deduzida a partir de pistas linguísticas e do conhecimento do mundo; a compreensão crítica diz respeito à capacidade de formar um juízo (sobre atitudes de personagens, por exemplo) e de entender qual a lição que é transmitida pela narrativa, que pode estar explícita ou implícita no texto.

1.3. **Instrumentos de avaliação de competências narrativas**

Existem vários instrumentos de avaliação de competências narrativas para diferentes línguas e para o inglês em particular, que se distinguem pelas faixas etárias abrangidas, pelas tarefas que propõem (conto ou reconto, com ou sem suporte de imagens), pelas dimensões que contemplam (apenas produção ou também compreensão) e pela forma como analisam e codificam os desempenhos das crianças.

Um dos instrumentos que foi mais usado na investigação sobre o desenvolvimento infantil contempla a história *Frog, where are you*?, sequência de imagens sem texto que ilustra uma série de peripécias vividas por uma criança que procura por uma rã (Berman & Slobin, 1994). Para o português, os estudos de Batoréo (1996) e de outros investigadores (Monteiro et al., 2016; Silva, 2012; e.o.) usaram duas sequências, de seis imagens cada, ilustrando a "história do gato" e "a história do cavalo" (Hickman, 1982). Estes instrumentos, ainda que largamente usados, não se encontram estandardizados.

Entre os instrumentos estandardizados e usados em contexto clínico encontram-se os seguintes: *Bus Story Test* (Renfrew, 1969); *TNL - Test of Narrative Language*, *TNL2* (Gillam & Pearson, 2004, 2017); *ERRNI – Expression, reception and recall of narrative instrument* (Bishop, 2004); *ENNI – Edmonton Narrative Norms Instrument* (Schneider et al., 2005); *DELV-NR – Diagnostic Evaluation of Language Variation – Norm Referenced* (Seymour et al., 2005).[1]

Todos os instrumentos anteriormente descritos têm como procedimento comum a obtenção de um *corpus* de produções de fala através de narrativas em tarefa de reconto de histórias, com ou sem apoio de imagens. Quando não existem imagens, a criança apenas tem de escutar uma história narrada pelo avaliador e proceder ao seu reconto. As faixas etárias da aplicabilidade das provas centram-se entre os 3 e os 15 anos. A codificação e análise diferem de prova para prova, no entanto, as observações incidem sempre sobre a estrutura ou esquema narrativo, a análise linguística, a análise dos recursos de coesão e a análise das hesitações.

Mais recentemente, foi desenvolvido o instrumento *LITMUS - MAIN - Language Impairment Testing in Multilingual Settings: Multilingual Assessment Instrument for Narratives* (Gagarina et al., 2015), concebido para ser usado com crianças bilingues em cada uma das suas línguas. Apesar de não estar ainda normalizado, este instrumento tem sido amplamente usado internacionalmente, havendo versões disponíveis para muitas línguas.

Não existe, contudo, nenhum instrumento validado para o PE, adaptado para a faixa etária de crianças de primeiro e segundo ciclo do ensino básico (Viana et al., 2017).

2. **Metodologia**

2.1. **Amostra**

Participaram no estudo 521 crianças com desenvolvimento típico (DT) e 17 crianças com Perturbação do Desenvolvimento da Linguagem (PDL), de oito regiões de Portugal continental e insular (Aveiro, Açores,

[1] Para uma caracterização mais detalhada de cada um destes instrumentos, consulte-se Vaz (2024).

Guarda, Leiria, Lisboa, Madeira, Portalegre, Santarém),[2] com idades entre os 6 e os 12 anos. Nas Tabelas 1 e 2, apresenta-se a distribuição dos participantes por idade e por região, para cada um dos grupos.

Tabela 1

Número de Participantes de Cada Grupo por Região

	DT	PDL
Aveiro	67	--
Açores	79	--
Guarda	61	--
Leiria	82	--
Lisboa	49	13
Madeira	66	--
Portalegre	61	--
Santarém	56	4
Total	521	17

Tabela 2

Número de Participantes de Cada Grupo por Faixa Etária

	DT	PDL
6 anos	77	7
7 anos	87	1
8 anos	94	2
9 anos	93	2
10 anos	73	2
11 anos	84	2
12 anos	13	1
Total	521	17

Foram definidos critérios específicos para a participação no estudo, assegurando a homogeneidade do grupo de crianças com PDL. O (RE)CONTO foi aplicado, de forma exploratória, a este grupo, seguindo os seguintes critérios de seleção:

- Diagnóstico de PDL;
- Inteligibilidade de fala suficiente para permitir ao avaliador compreender o conteúdo da narrativa oral (palavras ou enunciados);
- Ser falante monolingue de português europeu;
- Ter idade cronológica entre 6 e 12 anos;
- Frequentar o 1.º ou 2.º ciclo do ensino básico.

[2] Estavam previstas recolhas em mais regiões, que tiveram de ser canceladas devido ao surto pandémico de covid-19.

2.2. **Desenvolvimento do instrumento**

Após uma fase de pilotagem (Vaz et al., 2020), foi fixada a versão final do instrumento, o protocolo de aplicação e as grelhas de codificação. A versão final do instrumento, a que se deu o nome de (RE)CONTO, contém três sequências de imagens coloridas correspondentes a três histórias de complexidade linguística e narrativa variável. As imagens foram desenhadas por uma ilustradora profissional, com experiência na área de instrumentos de avaliação da linguagem. O instrumento contempla a H1 – História do pássaro e da minhoca; a H2 – História dos elefantes espertalhões; e a H3 – História da fada arco-íris. A primeira é uma história original; a segunda é adaptada do livro *Chibos Sabichões* (González, 2016); a terceira é traduzida e adaptada do livro espanhol *El hada del arco iris* (Yebra, 2015). Na Figura 1, apresentamos a sequência de imagens da H1.

Figura 1

Sequência de Imagens (a Preto e Branco) da H1[3]



A H1 foi usada associada a tarefas de conto e de reconto; a H2 e a H3 foram usadas apenas associadas à tarefa de reconto. Para cada história, foram construídas questões de compreensão, incluindo compreensão literal, inferencial e crítica. As histórias construídas para a versão de reconto diferiam entre si na complexidade narrativa: a H1 tem uma sequência narrativa mais clássica, com uma situação inicial, uma complicação e uma resolução, uma segunda complicação e resolução e um desfecho; a H2 integra uma sequência descritiva e várias sequências dialogais e tem uma estrutura de eventos mais repetitiva; a H3 integra sequências descritivas e tem uma estrutura narrativa menos linear. Do ponto de vista linguístico, a H3 é a mais complexa, integrando uma maior densidade de orações complexas e uma maior diversidade lexical.

Na Tabela 3, apresentam-se alguns indicadores de complexidade linguística dos textos que acompanham cada uma das histórias. Como se pode ver, a H3 apresenta, em geral, uma maior complexidade linguística, traduzida num maior número de *types*, numa maior extensão média de enunciado, numa maior proporção de orações subordinadas por total de enunciados e numa maior diversidade de orações subordinadas. A H2, ainda que seja mais longa, com um maior número de *tokens*, é mais repetitiva, o que se traduz num menor rácio *type/token*, e tem uma menor proporção de orações subordinadas por total de enunciados. Ainda assim, é a H2 que contém um maior número de orações subordinadas.

[3] As imagens originais são apresentadas a cores.

Tabela 3

Indicadores de Complexidade Linguística para as Três Histórias na Versão de Reconto

	H1	H2	H3
Número de *types*	158	180	185
Número de *tokens*	292	514	363
Rácio *type-token*	0,541	0,350	0,510
Extensão média de enunciado	11.792	11.065	13.630
Proporção de orações subordinadas / total enunciados	0,58	0,44	0,67
N.º de enunciados com mais do que uma o. subordinada	1	3	3
N.º de orações subordinadas	14	20	16
Diversidade de tipos de orações subordinadas	7	7	9

2.3. **Procedimentos**

Todas as crianças foram avaliadas individualmente, tendo sido aplicadas: 1) uma prova de avaliação da linguagem – *GOL-E – Grelha de avaliação da Linguagem – nível escolar* (Sua-Kay & Santos, 2014) para despistar eventuais problemas de linguagem e correlacionar com os resultados no (RE)CONTO; 2) a prova de memória *RAVLT – Rey Auditory Verbal Learning Test* (Teste de Aprendizagem Auditivo-Verbal de Rey) (Rey, 1964), para verificar se o desempenho na tarefa de reconto está dependente das capacidades de memória; 3) o *(RE)CONTO*, incluindo uma tarefa de conto, uma tarefa de reconto e perguntas de compreensão para cada história. O grupo PDL só teve como tarefa o conto da H1 e o reconto da H2, e as respetivas perguntas de compreensão.

Na tarefa de conto, a investigadora mostrava a sequência de imagens da H1e pedia à criança para contar a história com o suporte das imagens. Na tarefa de reconto, a criança ouvia a história gravada ao mesmo tempo que ia vendo a sequência de imagens. Depois, a investigadora pedia à criança para recontar a história com o suporte das imagens e, no fim, colocava as questões de compreensão. A tarefa de reconto foi aplicada com as três histórias a diferentes grupos de crianças. No total, obtiveram-se 907 narrativas (conto e reconto) produzidas pelas crianças.

Todas as produções das crianças foram gravadas em áudio com consentimento prévio dos pais/educadores, e o projeto e materiais foram submetidos e aprovados pela Comissão de Ética da NOVA/FCSH, pelo Ministério de Educação e pela Direção Regional de Educação da Madeira e dos Açores.

O instrumento foi aplicado individualmente, nas respetivas escolas, em salas com condições adequadas para o efeito (privacidade, silêncio).

2.4. **Codificação e análise**

As histórias narradas pelas crianças foram transcritas em formato .chat, seguindo os critérios estabelecidos no manual de transcrição do *Child Language Data Exchange System* (CHILDES) (MacWhinney, 2000). O processo de transcrição teve como princípio fundamental a divisão do contínuo sonoro em enunciados distintos, permitindo a análise individualizada de cada unidade discursiva. Para potenciar a fiabilidade dos resultados, houve necessidade de uniformizar critérios e criar etiquetas especificas de anotação (cf. Vaz, 2024).

Para efeitos de investigação, as transcrições foram anotadas em linhas de codificação permitidas pelo programa quanto à presença e tipo de orações subordinadas e coordenadas, presença e tipo de conetores, presença de desvios de gramaticalidade. Nos desvios de gramaticalidade considerados, encontram-se erros de concordância nominal ou verbal, problemas com flexão regular ou irregular, uso inadequado de tempo ou modo, problemas de retoma pronominal, problemas no uso de artigos.

Após avaliação por um grupo de peritos e pilotagem com profissionais da área da terapia da fala, linguística e educação, chegou-se a uma versão simplificada de grelha de codificação da prova, que contempla parâmetros macroestruturais e parâmetros microestruturais. Nos parâmetros macroestruturais, codificados como 0 ou 1, incluía-se a presença/ausência de abertura da narrativa, sequência de eventos e desfecho da narrativa, bem como introdução de personagens e localização espácio-temporal, e sequências de eventos específicos de cada narrativa. Nos parâmetros microestruturais, com cotações entre 0 e 2, incluíram-se: número de orações subordinadas finitas; complexidade de estruturas de subordinação tendo em conta o grau de encaixe; número de conetores temporais; número de conetores causais; coesão referencial; coesão temporal; desvios de gramaticalidade. Todas as cotações foram especificadas com base em critérios de desempenho descritos no manual de apoio à classificação da prova. Procurou-se que os critérios fossem de fácil codificação e adaptados a profissionais que podem não ter um domínio muito aprofundado de metalinguagem gramatical, fornecendo exemplos de respostas válidas e inválidas para cada parâmetro.

Além dos parâmetros que faziam parte da grelha de codificação, para cada narrativa produzida pela criança, procedeu-se, através do uso das ferramentas de análise automáticas disponíveis no CLAN, à análise da extensão média de enunciado, contagem de *types*, de *tokens*, e de rácio *type-token*. Estes dados estão disponíveis em Vaz (2024). No presente artigo, teremos em conta apenas as classificações devolvidas pelas grelhas de codificação.

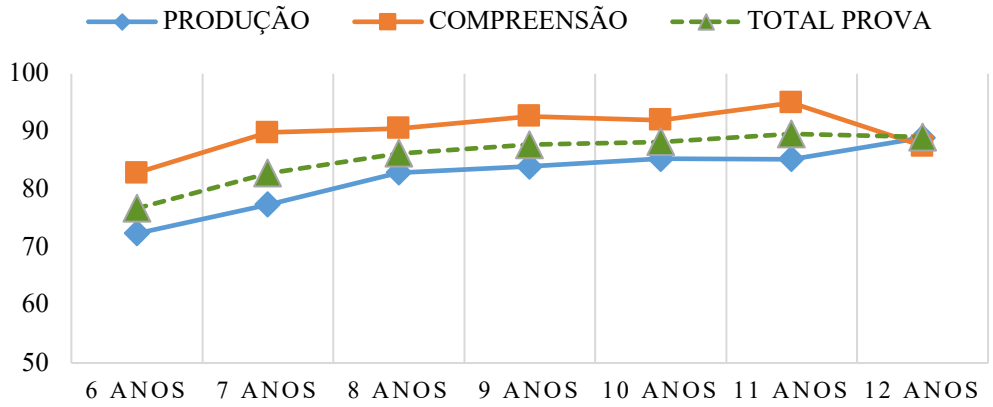
3. Resultados

3.1. Resultados globais

Na Figura 2, apresentam-se os resultados globais que o grupo com DT obteve na prova, bem como os resultados nas tarefas de produção e de compreensão, por faixa etária:

Figura 2

Resultados Globais Médios Obtidos pelo Grupo DT



Como se observa na figura 2, há uma tendência para a progressão nos valores obtidos, ainda que de forma moderada, com exceção do grupo de 12 anos, o que é explicável pela reduzida dimensão da amostra neste subgrupo.

3.2. **Análise macroestrutural e microestrutural**

Nas Figuras 3 e 4, apresentam-se os valores médios obtidos na tarefa de reconto de cada uma das histórias nos parâmetros macroestruturais e microestruturais, respetivamente. Como se pode observar, é mais notória a progressão nos parâmetros microestruturais, sendo a H2 aquela em que os valores tendem a ser mais baixos em ambos os parâmetros.

Figura 3

Valores Médios Obtidos nos Parâmetros Macroestruturais na Tarefa de Reconto de Cada História no Grupo DT

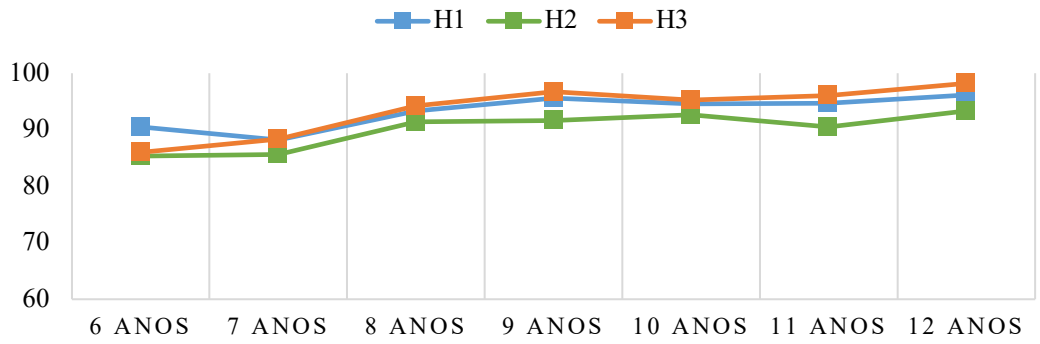
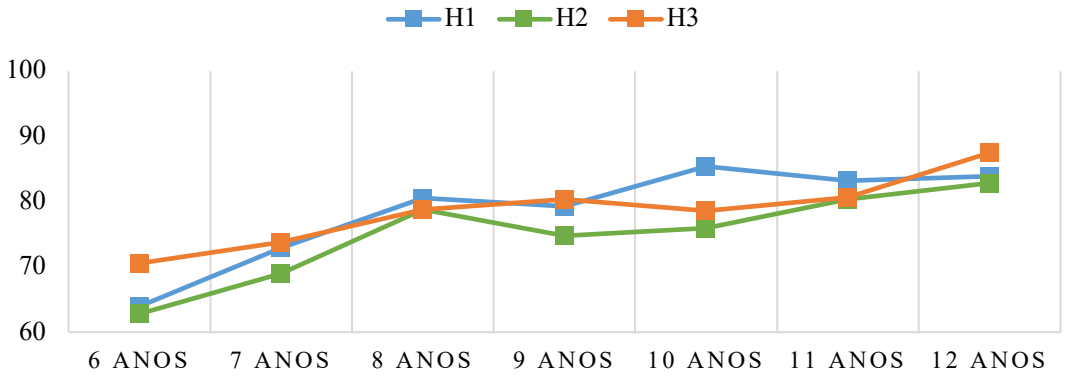


Figura 4

Valores Médios Obtidos nos Parâmetros Microestruturais na Tarefa de Reconto de Cada História no Grupo DT

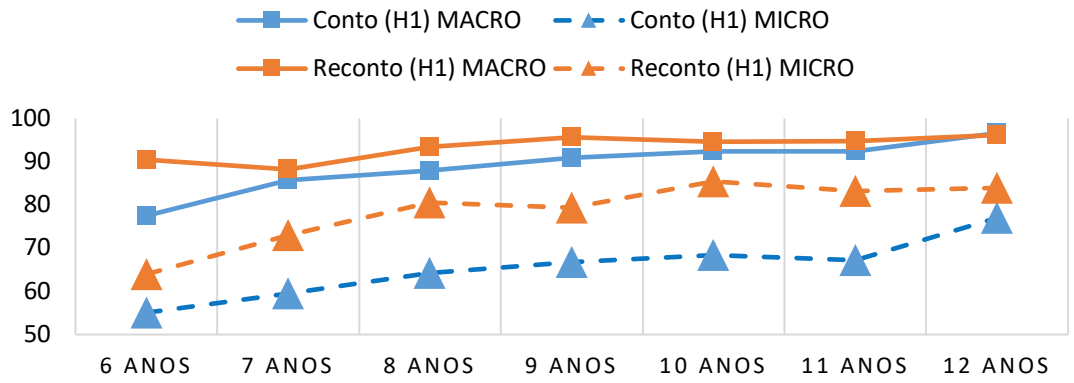


3.3. **Comparação entre tarefa de conto e de reconto**

De forma a verificar a influência da tarefa no desempenho das crianças, procedeu-se ainda à comparação entre os valores obtidos na produção da H1 em função do tipo de tarefa – conto ou reconto (Figura 5).

Figura 5

Valores Médios Obtidos na Tarefa de Conto e de Reconto da H1 no Grupo DT por Faixa Etária



Os resultados apresentados na Figura 5 mostram que os valores obtidos na tarefa de reconto são sempre superiores aos que foram obtidos na tarefa de conto, sendo a diferença bastante mais expressiva nos parâmetros de análise microestrutural. Estes resultados reforçam a relevância da tarefa de reconto para a intervenção e para a estimulação da produção de estruturas linguísticas mais complexas. A prova (RE)CONTO pode, assim, servir para estabelecer um diagnóstico, mas também para estimular a criança a produzir narrativas mais elaboradas.

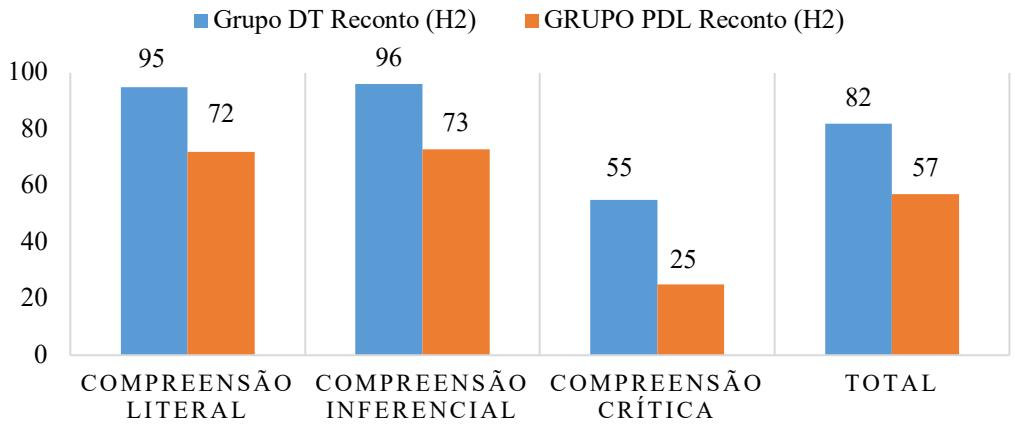
3.4. **Comparação entre grupo DT e grupo PDL**

Finalmente, comparamos os resultados exploratórios obtidos com o grupo PDL com os do grupo DT, de forma a verificarmos a eficácia da prova na identificação de crianças com dificuldades linguísticas e a facilidade de aplicação a grupos com desempenhos linguísticos comprometidos.

Na Figura 6, apresenta-se a percentagem de acerto de cada grupo na tarefa de compreensão da H2, discriminando cada subtipo de compreensão – literal, inferencial e crítica.

Figura 6

Acerto nas Questões de Compreensão da H2 por Grupo

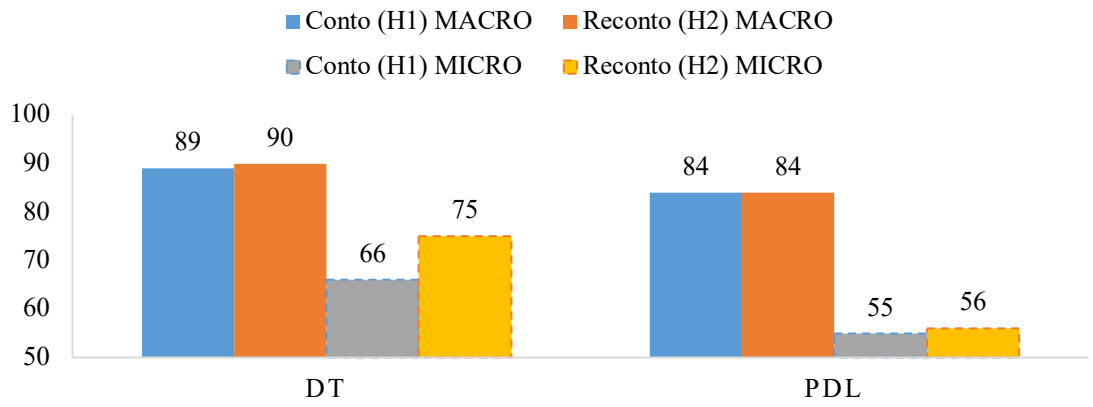


Como se pode observar, o desempenho do grupo PDL é significativamente inferior ao do grupo DT, em particular nas questões que envolviam compreensão crítica, isto é, em que se pedia à criança que dissesse que lição se poderia retirar da história.

Na Figura 7, apresentam-se os valores médios de acerto obtidos por cada um dos grupos na tarefa de conto da H1 e de reconto da H2, considerando os parâmetros de análise macro e microestruturais.

Figura 7

Valores Médios Obtidos nas Tarefas de Conto e Reconto nos Grupos DT e PDL



A comparação entre os grupos mostra-nos que os dois grupos se distinguem sobretudo nos parâmetros microestruturais, em que os valores obtidos pelo grupo PDL são consistentemente mais baixos do que os do grupo DT.

3.5. **Propriedades psicométricas**

As imagens, histórias, grelhas de análise do instrumento (RE)CONTO e o respetivo manual foram avaliados por um painel de sete peritos que incluía linguistas, terapeutas da fala e profissionais da área da didática, tendo obtido um excelente índice de validade de conteúdo global: 0,94.

A correlação entre os resultados obtidos pelas crianças no (RE)CONTO e os obtidos no teste de avaliação da linguagem GOL_E permitiram avaliar a validade concorrente do instrumento, tendo-se obtido uma correlação positiva significativa entre os instrumentos, com um nível de confiança de 95%.

Adicionalmente, a validade convergente foi analisada através da comparação com o RAVLT (Teste de Aprendizagem Verbal de Rey), permitindo compreender a relação entre a memória e o desempenho na tarefa de reconto. No grupo DT (n=526), encontraram-se correlações fracas, mas significativas, entre a Curva de Aprendizagem ($r = 0{,}166$, $p < 0{,}05$) e o Índice de Interferência Retroativa ($r = 0{,}089$, $p < 0{,}05$) com o reconto. Contudo, no grupo PDL (n=17), nenhuma correlação foi significativa, possivelmente devido à reduzida dimensão da amostra. Estes resultados sugerem que a memória tem alguma influência no desempenho narrativo, mas não é o único fator determinante. As competências linguísticas e organizacionais podem desempenhar um papel relevante, refletindo a complexidade da relação entre memória e linguagem (Cowan, 2008; Baddeley, 2000; Tulving, 1983).

A consistência interna do instrumento foi verificada utilizando o Alpha de Cronbach, que obteve um valor de 0.76, indicando uma consistência interna adequada. Para testar a fiabilidade interavaliador, foi realizado um estudo com três terapeutas da fala e dois avaliadores de linguística, obtendo-se um coeficiente Kappa superior a 0.87, refletindo uma excelente concordância entre os avaliadores nas tarefas e nas análises macro e microestruturais.

4. **Discussão**

Os resultados obtidos com o instrumento (RE)CONTO reforçam a relevância da avaliação das competências narrativas, corroborando estudos anteriores que destacam a relação entre a produção de narrativas e o desenvolvimento da literacia (Wellman et al., 2011). Como evidenciado na análise dos dados, observou-se uma progressão moderada nos valores globais obtidos pelos participantes, o que está em linha com investigações que demonstram uma evolução gradual das competências narrativas ao longo da idade (Pereira & Azevedo, 2005). Contudo, a irregularidade na progressão dos resultados em determinados subgrupos etários, nomeadamente no grupo de 12 anos, sugere a necessidade de considerar a dimensão da amostra e outros fatores contextuais que possam influenciar o desempenho.

Relativamente à análise macroestrutural e microestrutural, os nossos resultados evidenciam uma evolução mais marcada nos parâmetros microestruturais, tal como mostrado em estudos prévios que indicam que a maturação das competências linguísticas mais finas, como a coesão referencial e a complexidade sintática, tende a ser mais notória em crianças com desenvolvimento típico da linguagem (Gagarina et al., 2012). A menor complexidade das narrativas na H2 (História dos Elefantes Espertalhões) pode estar relacionada com fatores como a familiaridade com o conteúdo da história ou a exigência cognitiva necessária para integrar diferentes elementos narrativos, o que corrobora estudos que apontam para a influência do conteúdo e do tipo de tarefa na produção narrativa (Roch et al., 2016; Silva & Spinillo, 2000).

A comparação entre as tarefas de conto e de reconto de histórias confirmou que as crianças obtêm melhores desempenhos na tarefa de reconto, especialmente ao nível da microestrutura, o que já havia sido documentado na literatura (Cruz, 2011; Shapiro & Hudson, 1991). Estes resultados reforçam a importância da tarefa de reconto não apenas como um meio de avaliação, mas também como um recurso pedagógico e de intervenção clínica para estimular o desenvolvimento de narrativas mais completas. A presença de uma estrutura narrativa previamente apresentada parece facilitar a organização do discurso e incentivar a utilização de construções linguísticas mais complexas, um resultado que se alinha com estudos que evidenciam o papel do suporte visual

e do modelo narrativo na produção infantil (Spinillo, 1993; Veloso, 1999). Além disso, as diferenças verificadas entre os participantes sugerem que a competência narrativa pode ser influenciada por fatores individuais, como o nível de exposição a práticas narrativas em ambiente familiar e escolar.

A análise comparativa entre o grupo DT e o grupo PDL revelou que as crianças com perturbação apresentam dificuldades mais acentuadas nos parâmetros microestruturais e na compreensão crítica. Estes resultados demonstram que crianças com dificuldades linguísticas tendem a apresentar menor riqueza lexical, menor complexidade sintática e dificuldades na organização narrativa (Cain & Oakhill, 2006; Norbury & Bishop, 2003; Paul & Smith, 1993). O facto de a diferença entre os grupos ser mais evidente nos aspetos microestruturais reforça a ideia de que estas crianças podem beneficiar de estratégias de intervenção focadas no reforço da coesão referencial e na utilização de estruturas complexas, como proposto por Bishop (2009) e Petersen (2011). Nesse sentido, a inclusão de tarefas específicas de reconto e de modelação narrativa pode revelar-se uma abordagem eficaz para a promoção da aprendizagem de estruturas linguísticas mais complexas em crianças com dificuldades de linguagem.

Para além destas questões, o *corpus* construído para o presente estudo, composto por 907 narrativas orais transcritas e anotadas segundo as normas CHAT, representa um contributo valioso para futuras investigações no domínio do desenvolvimento narrativo infantil. A distinção entre as 874 produções de crianças com desenvolvimento típico e as 33 produções de crianças com perturbação da linguagem permite não apenas uma análise detalhada das diferenças linguísticas entre os grupos, mas também a identificação de possíveis marcadores linguísticos de perturbações do desenvolvimento da linguagem. O estudo da subordinação e das construções sintáticas mais complexas revelou-se particularmente relevante neste contexto, reforçando a ideia de que a produção de estruturas de subordinação e de orações com mais do que uma subordinada constitui um indicador mais robusto de desenvolvimento linguístico do que outras estruturas complexas. Estes resultados poderão ter implicações diretas na formulação de critérios diagnósticos e na definição de objetivos terapêuticos mais precisos.

Por fim, as propriedades psicométricas do instrumento (RE)CONTO demonstram a sua robustez enquanto ferramenta de avaliação. A elevada validade de conteúdo, a consistência interna adequada e a excelente fiabilidade interavaliador garantem que este instrumento pode ser utilizado com confiança na avaliação de competências narrativas orais. No entanto, os resultados relativos à validade concorrente e convergente sugerem que o desempenho narrativo é influenciado por múltiplos fatores além da memória, o que vai ao encontro de investigações que destacam a complexidade dos processos envolvidos na produção e compreensão de narrativas (de Villiers & de Villiers, 2010; Sterponi, 2019). Além disso, a aplicabilidade clínica do (RE)CONTO poderá beneficiar de estudos longitudinais que acompanhem a evolução das competências narrativas das crianças ao longo do tempo e investiguem a relação entre o desempenho narrativo e outras competências linguísticas e cognitivas.

Assim, os resultados deste estudo reforçam a importância da avaliação das competências narrativas em crianças com e sem dificuldades linguísticas, demonstrando o potencial do (RE)CONTO como instrumento de avaliação de narrativas orais para o português europeu em crianças dos 6 aos 12 anos de idade. Estudos futuros poderão centrar-se na ampliação da amostra e na análise de variáveis adicionais que possam influenciar o desempenho narrativo, tais como o contexto socioeconómico, a frequência de exposição a narrativas e o impacto de intervenções pedagógicas e terapêuticas específicas. A investigação nesta área poderá contribuir significativamente para o desenvolvimento de estratégias mais eficazes na promoção das competências narrativas em contexto escolar e clínico, garantindo um apoio mais adequado às crianças que apresentam dificuldades no desenvolvimento linguístico.

Agradecimentos

Deixamos um agradecimento especial aos diretores de agrupamentos e professores que acreditaram na relevância desta investigação, e às crianças e suas famílias, que foram essenciais para a realização deste estudo, bem como a todos os

colaboradores e peritos que participaram na validação do instrumento (RE)CONTO e na normalização da prova de memória *RAVLT*.

Agradecemos ainda ao Eduardo Castro e ao João Oliveira Santos o apoio imprescindível no processo de transcrição, codificação e análise de dados.

Financiamento

Este trabalho foi financiado pela FCT – Fundação para a Ciência e a Tecnologia, I.P. através da bolsa de doutoramento atribuída a Stéphanie Vaz (PD/BD/131387/2017), bem como do projeto UID/03213 – Centro de Linguística da Universidade NOVA de Lisboa (CLUNL) e no âmbito do CINTESIS, Unidade de I&D (UIDB/4255/2020 e UIDP/4255/2020) e RISE (LA/P/0053/2020).

Referências

Adam, J.-M. (1992). *Les textes : Types et prototypes*. Nathan.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Batoréo, H. (1996). *Expressão do espaço no português europeu: Contributo psicolinguístico para o estudo da linguagem e cognição* [Tese de doutoramento, Universidade de Lisboa].

Berman, R. A., & Slobin, D. I. (Eds.). (1994). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press.

Bishop, D. V. M. (2004). *Expression, reception, and recall of narrative instrument – ERRNI*. Harcourt Assessment.

Bishop, D. V. (2009). Curing the pathology of oral language: Where will we put our efforts? *Child Development*, *80*(3), 745–749.

Borges, I. (2015). *Expressão de tempo e construção de narrativa* [Tese de mestrado, Instituto Politécnico de Setúbal]. Repositório da UNL. http://hdl.handle.net/10362/20744

Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, *76*(4), 697–708. https://doi.org/10.1348/000709905X69807

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory*? Progress in Brain Research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Cruz, M. (2011). *As histórias como recurso para o desenvolvimento da linguagem oral* [Relatório de prática de ensino supervisionada, Universidade do Algarve]. Sapientia – Repositório Científico da UAlg. http://hdl.handle.net/10400.1/3096

de Villiers, P. A., & de Villiers, J. G. (2010). Assessment of language acquisition. *WIREs Cognitive Science*, *1*(2), 230–244. https://doi.org/10.1002/wcs.30

Favot, K., Carter, M., & Stephenson, J. (2021). The effects of oral narrative intervention on the narratives of children with language disorder: A systematic literature review. *Journal of Developmental and Physical Disabilities*, *33*(4), 489–536. https://doi.org/10.1007/s10882-020-09763-9

Feagans, L., & Appelbaum, M. I. (1986). Validation of language subtypes in learning disabled children. *Journal of Educational Psychology*, *78*(5), 358–364. https://psycnet.apa.org/doi/10.1037/0022-0663.78.5.358

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN – Multilingual assessment instrument for narratives. *ZAS Papers in Linguistics*, *56*. Zentrum für Allgemeine Sprachwissenschaft.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of narrative abilities in bilingual children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 243–276). Multilingual Matters. https://doi.org/10.21832/9781783093137-011

Giasson, J. (2005). *La lecture : De la théorie à la pratique*. De Boeck & Larcier.

Gillam, R. B., & Pearson, N. A. (2004). *Test of narrative language – TNL*. WPS.
Gillam, R. B., & Pearson, N. A. (2017). *Test of narrative language – TNL2* (2.ª ed.). WPS.
González, O. (2016). *Chibos sabichões*. Kalandraka.
Hickman, M. (1982). *The development of narrative skills: Pragmatic and metapragmatic aspects of discourse cohesion* [Tese de doutoramento, Universidade de Chicago].
Hughes, D., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Thinking Publications.
Kraljevic, J., Hrzica, G., & Gorup, I. (2020). A comparative macrostructural analysis of narrative discourse in children with typical language development and children with developmental language disorder. *Journal for General Social Issues*, *29*(3), 453–470.
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3.ª ed.). Lawrence Erlbaum Associates.
Monteiro, P., Costa, A., & Batista, A. (2016). Parâmetros para análise de narrativas orais em crianças com e sem perturbação da linguagem expressiva. *Revista da Associação Portuguesa de Linguística*, *1*, 621–650. https://doi.org/10.26334/2183-9077/rapln1ano2016a26
Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, *38*(3), 287–313. https://doi.org/10.1080/13682031000108133
Paul, R., & Smith, R. L. (1993). Narrative skills in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech, Language, and Hearing Research*, *36*(3), 592–598. https://doi.org/10.1044/jshr.3603.592
Pereira, L. Á., & Azevedo, F. (2005). *Como abordar... a escrita no 1º ciclo do ensino básico*. Areal Editores.
Pereira, D. (2010). *Compreensão leitora e literatura infantil: Práticas no 1º ciclo do ensino básico* [Tese de mestrado, Universidade da Beira Interior]. uBibliorum – Repositório Digital da Universidade da Beira Interior. http://hdl.handle.net/10400.6/1948
Petersen, D. (2011). A systematic review of narrative-based language intervention with children who have language impairment. *Communication Disorders Quarterly*, *32*, 207–220. https://doi.org/10.1177/1525740109353937
Renfrew, C. (1969). *The bus story test: A test of narrative speech*. Speechmark Publishing.
Rey, A. (1964). *L'examen clinique en psychologie*. Presses Universitaires de France.
Roch, M., Florit, E., & Levorato, C. (2016). Narrative competence of Italian–English bilingual children between 5 and 7 years. *Applied Psycholinguistics*, *37*, 49–67. https://doi.org/10.1017/S0142716415000417
Rudrum, D. (2005). From narrative representation to narrative use: Towards the limits of definition. *Narrative*, *13*(2), 195–204. https://doi.org/10.1353/nar.2005.0013
Schneider, P., Dubé, R. V., & Hayward, D. (2005). *Edmonton narrative norming instrument – ENNI*. University of Alberta.
Seymour, H. N., Roeper, T. W., & de Villiers, J. (2005). *Diagnostic evaluation of language variation — Norm-referenced test (DELV-NR)*. Pearson.
Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion on young children's picture-elicited narratives. *Developmental Psychology*, *27*(6), 960–974. https://psycnet.apa.org/doi/10.1037/0012-1649.27.6.960
Silva, M. E. L., & Spinillo, A. G. (2000). A influência de diferentes situações de produção na escrita de histórias. *Psicologia: Reflexão e Crítica*, *13*(3), 337–350. https://doi.org/10.1590/S0102-79722000000300003
Silva, M. da E. (2001). *O desenvolvimento da competência narrativa: Uma análise de narrativas orais e escritas produzidas por sujeitos de 6, 7, 9, 11 e 14 anos* [Tese de mestrado, Universidade Nova de Lisboa].

Silva, M. L. S. (2012). *Aquisição/aprendizagem do português europeu em contexto plurilingue na Suíça: Aspetos morfossintáticos* [Tese de mestrado, Universidade Aberta]. Repositório Aberto – Repositório Institucional da Universidade Aberta. http://hdl.handle.net/10400.2/2596

Silva, P. N. da. (2015). Narrativo: Modo, género, tipo de texto ou tipo de sequência? In *Encontro Nacional da Associação de Professores de Português, 11, Évora, 2015 – "Literatura e gramática [Em linha]: um diálogo infinito: atas"*. APP. http://hdl.handle.net/10400.2/4515

Spinillo, A. G. (1993). Era uma vez... e foram felizes para sempre: Esquema narrativo e variações experimentais. *Temas de Psicologia*, *1*, 67–87.

Sterponi, L. (2019). Narratives and the development of emotional and imaginative capacities. *Frontiers in Psychology*, *10*, 1457.

Sua Kay, E., & Santos, M. E. (2014). *Grelha de observação da linguagem – Nível escolar (GOL-E)* (2.ª ed.). Escola Superior de Saúde do Alcoitão.

Tulving, E. (1983). *Elements of episodic memory*. Clarendon Press.

van Dijk, T. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Erlbaum.

Vaz, S. D., Lobo, M., & Lousada, M. (2020). Avaliação de narrativas orais em crianças falantes de português europeu (PE): Um teste piloto. *Revista da Associação Portuguesa de Linguística*, *7*, 368–384. https://doi.org/10.26334/2183-9077/rapln7ano2020a22

Vaz, S. D. (2024). *Instrumento de avaliação de narrativas orais para crianças falantes de português europeu* [Tese de doutoramento, Universidade Nova de Lisboa].

Veloso, T. (1999). *A compreensão de narrativas e a produção do discurso* [Tese de mestrado não publicada]. Universidade do Minho.

Viana, F. L., Ribeiro, I. da S., Fernandes, I., Ferreira, A., Leitão, C., Gomes, S., Mendonça, S., & Pereira, L. (2010). *Aprender a compreender torna mais fácil o saber*. Almedina.

Viana, F. L., Silva, C., Ribeiro, I., & Cadime, I. (2017). Instrumentos de avaliação da linguagem: Uma perspetiva global. In M. J. Freitas & A. L. Santos (Eds.), *Aquisição de língua materna e não materna: Questões gerais e dados do português* (pp. 333–357). Language Science Press.

Wellman, R. L., Lewis, B. A., Freebairn, L. A., Avrich, A. A., Hansen, A. J., & Stein, C. M. (2011). Narrative ability of children with speech sound disorders and the prediction of later literacy skills. *Language, Speech, and Hearing Services in Schools*, *42*(4), 561–579. https://doi.org/10.1044/0161-1461(2011/10-0038)

Westby, C. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157–232). Allyn & Bacon.

Yebra, A. (2015). *El hada del arco iris* (11.ª ed.). Editorial Bruño.