1. Reading modality: voice and eyes as source for cognitive processes

Reading processes have been studied in a psycho-linguistic perspective and a lot is known about what is involved (Perfetti, 1992). Interrelated processes and behaviors in reading have been repeatedly studied in silent reading and reading aloud. The data dependent on reading modality inform us in different ways about online linguistic processing, comprehension and fluency.

Reading aloud provide data from reading speech, where speed, hesitations and prosody constitute a rich source for inferring and interpreting ongoing linguistic integration. The required speech articulation creates a lag between visual input perception and its production (Benjamin & Schwanenflugel, 2010).

In silent reading, eyes can be a good window to capture cognitive processes. Fixations, progressions and regressive saccades, that is, the eye movements patterns can inform us about processing costs when dealing with materials with different linguistic and visual properties (Rayner, John Polushkin, 2005).

What about combined reading data from voice and eyes? Will they be more informative when isolated?

Eye–voice span (EVS) is a construct that has been exploited tentatively from Bussel to today (Inhoff et al., 2011; Laubrich & Kleig, 2016, a.p.). Its purpose is to grasp the dynamic cognitive labor involving linguistic representations, perceptual and cognitive processes required for comprehension when reading aloud. EVS means the distance between eyes and voice, knowing that when the production of a given word begins, the eyes are ahead 2 to 3 words or around 500ms. The EVS amplitude is variable considering the reader experience and processing troubles triggered for local linguistic properties of print input, such as lexical properties.

Our previous research on reading

Costa, Falé and Luegi are developing a set of experiments: on oral reading per se with speech analysis (Costa, 1992), on silent reading using eye-tracking (Luegi, 2006; Costa, Mateus & Luegi, 2009); and more recently reading aloud with simultaneous register of speech and eye movements (Falé, Costa & Luegi, 2016). We aim to study the impact of lexical, syntactic and visual properties in processing costs. Considering the data from speech reading and from eye-tracking, we found that eye movements are more sensitive to linguistic properties and text complexity than reading speech; sometimes the eyes respond to linguistic properties that voice apparently ignores.

Current research question:

To what extent can the lexical complexity affect the EVS?

Hypothesis

Considering that lexical properties, such as word length, syllable structure, word stress and frequency, impact on visual word recognition and lexical access, and assuming that EVS should be sensitive to lexical properties, we predict that:
- phonological complex words, as a result of number and syllable type and word stress, should have an effect on EVS, shortening it.
- less frequent phonological complex words should reduce more the EVS amplitude.

3. Experiment

Design and materials

Two tests, each with 30 target words, avoiding positions in the end of the line, punctuation marks, and contiguity between targets. The 60 words, with 3 or 4 syllables, were distributed over three levels of complexity (around 20 per each level), taking into account:
- phonological properties, such as number and syllable type, and stress type, following the hierarchy based on syllable types frequencies in European Portuguese (Vigário & Falé, 1994; Vigário et al., 2006).
- word frequency, according the Multifunctional Computational Lexicon of Contemporary Portuguese (C1 C2 C3)

Participants

17 Portuguese adult native female speakers, university students, proficient readers.

Procedure

Participants read two texts and were asked to read aloud at their own pace, trying to understand. After reading each text, participants answered a multiple-choice questionnaire, thus ensuring a reading comprehension task.

Eye movements were recorded with a SMI VIEW X® HI-SPEED system, at a 120Hz speed, and sound was recorded with a Logitech® Webcam Pro 9000.

Stimuli were divided, for presentation, into two blocks of text, font in size 22, Courier New, with two paragraphs spacing between rows, in a 27-inch screen.

Independent variables
- Phonological complexity at three levels: C1 ≤ C2 ≤ C3
- Phonological complexity plus frequency at three levels: C1 ≤ C2 ≤ C3

Dependent variables

Eye movement analysis

First fixation (FF) – first time a word is fixated regardless if it is fixated one or more times; reveals specific processes of visual word recognition (VWR).

First pass (FP) – all the time spent in the fixation of a word before moving the eyes to right or left regions; reveals aspects involved in WM and lexical access.

Speech analysis

Word Production Duration (WPD) – the time spent on the production of a word; should reveal lexical access and any disturbance coming from word recognition, phonological mapping, lexical access and possible articulatory problems.

EVS considering phonological complexity and frequency

Considering the results from eye movements and word production, statistical analysis was performed only on target words classified in 3 levels of complexity as a result of phonological and frequency properties.

Onset-EVS

- EVS is higher in C1 than in C2 (Est. = –0.105; SE = 0.116; t-ratio = 1.11; p = 0.27; 0.048;)
- There are no differences between C1 and C3 (Est. = –0.207; SE = 0.207; t-ratio = –0.75; p = 0.11; 0.082).

Onset_offset-EVS

- EVS is higher in C1 than in C2 (Est. = .78.58; SE = .71.42; t-ratio = .26.0; p = 0.012;)
- EVS is higher in C1 than in C3 (Est. = -.10.44; SE = .11.64; t-ratio = -.67.12; p = .001; 0.001).

5. Conclusions and next step

The temporal lag we obtained in onset-EVS can show the effect of some formal lexical properties: the eyes, in a first fixation, perceive quickly visual information that trigger neural and cognitive processes in specific cortical visual areas, the mapping between letters and sounds, and the planning of the motor processes that lead to the articulatory word phonetic form.

The temporal lag obtained by onset_offset-EVS must integrate and be reactive to all the formal lexical information, including meaning and frequency of the word, and even its relations with the mental test model that is being built during reading. When the articulation begins, it is informed by all lexical properties of the previous fixated word.

Results show a quasi non effect of onset_offset-EVS derived from complexity: simpler words allow a large EVS than C2, and not different from C3. Comparing C2 and C3, we can say that as the complexity increases, lag decreases.

When the voice is informed by eyes over all the lexical information and if the word is complex, voice holds the eyes and does not let them move forward, otherwise the working memory goes into overload and there will be loss of crucial information for understanding. In turn, to wait for the voice, eyes must do longer fixations and reformulations. To show the eyes, the voice commits disfluencies such as vowel lengthening [i[hibi[te]] or truncations [E(h)i[hibi[te]]).

And here is our next step in research: deepen what we know about cognitive processes in reading by the prosody and the dynamics of eye pattern.

References


The authors would like to thank Luís J. Lopes, Ana Sofia Almeida, Priscila Mendes and Renato Lopes for their support and contribution.

This work was supported by Fundação para a Ciência e a Tecnologia (UID/CLC/50022/2013) and FCT Fundação para a Ciência e a Tecnologia (SFRH/BD/97798/2014).


(2) Universidade Aberta