Determinants of scanpath regularity in reading

Abstract

In reading, the eyes often deviate from a straight left-to-right trajectory. Research in several fields identified three factors contributing to the irregularity of scanpath patterns. These factors include: age of reader, length of words, and syntactic complexity. However, so far no study has investigated the relative contributions of these factors and their interactions. In this study, we use a novel method for analyzing scanpath patterns to fill this gap.

Background & Goals

• Short words are skipped more often than longer words. Skipping of a short word is often followed by a regressive eye movement to the skipped word (Vitu & McConkie, 2000).
• Words that are hard to integrate into the sentence context can induce regressive eye movements (see Rayner, 1998, and Clifton, Staub, Rayner, 2007).
• Older readers skip and regress more often, while younger readers have a more regular left-to-right gaze trajectory (Klieg, Grabner, Rolfs, Engbert, 2004).
• These effects have so far mostly been studied using word-based measures like the probability to skip a word or to regress from it. Scanpaths have received little attention, in part due to a lack of appropriate analytical tools.
• Therefore, the goal of this study is to investigate the joint scanpath effects of age, word-length, and syntactic complexity using a new measure that is tailored for the analysis of spatio-temporal eye movement patterns (von der Malsburg & Vasishth, 2011).

Predictions:

1. Sentences with many short words elicit more irregular scanpath patterns.
2. Sentences that are linguistically more difficult to process elicit more irregular scanpaths.
3. Scanpaths recorded from older readers are more irregular than scanpath by young readers.

Material

• Potsdam Sentence Corpus (Klieg et al., 2004): 144 simple German sentences (mean number of words: 7.9, sd: 1.4).
• Participants from diverse socioeconomic backgrounds. Young readers on average 23.8 years old (sd: 7.9), old readers 68.8 years (sd: 5.6).
• Sentence difficulty was quantified using a variant of surprisal theory (Boston, Hale, Vasishth, Klieg, 2011) and cue-based parsing theory (Lewis & Vasishth, 2005).

Method

• For each sentence in the corpus, multi-dimensional scaling (Kruskal, 1968) and Gaussian mixture models (Fraley & Raftery, 2002, 2006) were used to calculate a density function over scanpaths.
• The regularity of a scanpath was quantified as its neighborhood density.
• Sentences that elicit more diverse scanpath patterns have lower scanpath density (fig. 2).
• A linear mixed model was used to model the density of scanpaths (i.e. regularity) as a function of (i) average word length in a sentence, (ii) age of reader, difficulty of the sentences according to (iii) surprisal theory, and (iv) cue-based retrieval theory, and (v) all two-way interactions of length, age, surprisal, and retrieval.

Summary

This is, to our knowledge, the first demonstration of the use of a scanpath measure to test predictions for scanpath regularity in reading. We found effects of word length, age of reader, surprisal and of retrieval difficulty. Beyond these anticipated results, we found that surprisal and retrieval effects were weaker in older readers. This interaction of age and syntax can be interpreted in terms of (i) a shift towards more risky and top-down driven sentence processing, or (ii) problems with executive control in older readers (cf. Wotschak & Klieg, 2012). Since all sentences were relatively simple, these results also show that the scanpath measure used here is sensitive even to subtle effects.