An informational sweet spot at the intersection between sentence context and word form

Jana Hasenäcker (Cognitive Neuroscience Department, SISSA, Trieste, Italy)
Simona Amenta (Centre for Mind/Brain Science, University of Trento, Italy)
Davide Crepaldi (Cognitive Neuroscience Department, SISSA, Trieste, Italy)
Marco Marelli (Department of Psychology, University of Milan-Bicocca, Milan, Italy)
First author e-mail: janahasenaecker@gmail.com

One central issue in language processing is how we read and understand words in sentences. Research on sentence reading indicates that we expect upcoming words based on the information we already have: the time we need to read a word depends on the amount of (un)expected information that it conveys (e.g., Frank, 2013; Smith & Levy, 2013; Staub, 2015). Despite the preceding context, however, each single word has its own recognition dynamics: research on single word recognition shows that a word's orthography also impacts on the ease of processing. Readers have an expectation of the word meaning based on the letters it is made of (Marelli & Amenta, 2018; Marelli, Amenta, & Crepaldi, 2015). Some letter sequences are better cues for a word meaning than others, and this factor has been found to predict reliably word recognition latencies (Marelli & Amenta, 2018).

Hence, processing a word in a sentence can gather on at least two different sources of information, namely sentence context and word orthography.

In the present study, we jointly examine the impact of those two distinct, but potentially interacting, sources of information on eye movements during sentence reading. Information from sentence context was operationalized in terms of surprisal, capturing the degree of unexpectedness of a word given the preceding words. For this, we used the surprisal measures from Frank, Otten, Galli, and Vigliocco (2015). Information from word orthography was operationalized in terms of Orthography-Semantics Consistency (OSC; Marelli & Amenta, 2018), providing an estimation of how expected a meaning is given a certain orthographic string. This was computed as the frequency-weighted average semantic similarity between the meaning of a target word and the meanings of all its orthographic relatives. We then analyzed the eye tracking data by Frank, Monsalve, Thompson, and Vigliocco (2013) using generalized additive models to look at the effect of surprisal and OSC on gaze durations, first fixation durations, right-bounded time, and regression-path time.

The results were characterized by a non-linear interaction between surprisal and OSC in all of the eye tracking measures (except for first fixation duration). Overall, looking times were shorter for less surprising contexts. Interestingly, durations were also shorter for words that had intermediate values on both the surprisal and the OSC scales. This is illustrated by the green middle area in Figure 1 indicating shorter gaze durations. Indeed, when the sentence context is moderately constraining (medium surprisal), making many lexical items possible, as it happens for most communication, then word orthography becomes a precious source of information. Here, word forms that too loosely point to a meaning (very low OSC) do not help word recognition much: in this case, there is too much uncertainty on both sources. By contrast, words that too strongly point to one specific meaning (very high OSC) could interfere with words predicted from context if the prediction does not match exactly with the actual word. In line with a suggestion by Ferreira and Lowder (2016), intermediate levels of new information seem to be the ideal amount we can receive: a word form that helps getting to a meaning easily but still leaves room for uncertainty (as medium OSC does) gives the best boost to recognition. This also suggests that, for words with an intermediate amount of new information for the sentence, that is medium surprisal, readers tolerate some uncertainty about the precise identity of an orthographic form in favor of coarser sentence-level meaning integration (cf. Smith & Levy, 2010). We describe this as an "informational sweet spot" that gives just the right amount of information to the reader's eye, gathering both from sentence and word form—neither too little, so that it is trivial, nor too much, so that the reader is overwhelmed or must resolve potentially conflicting information before moving on to the next word.

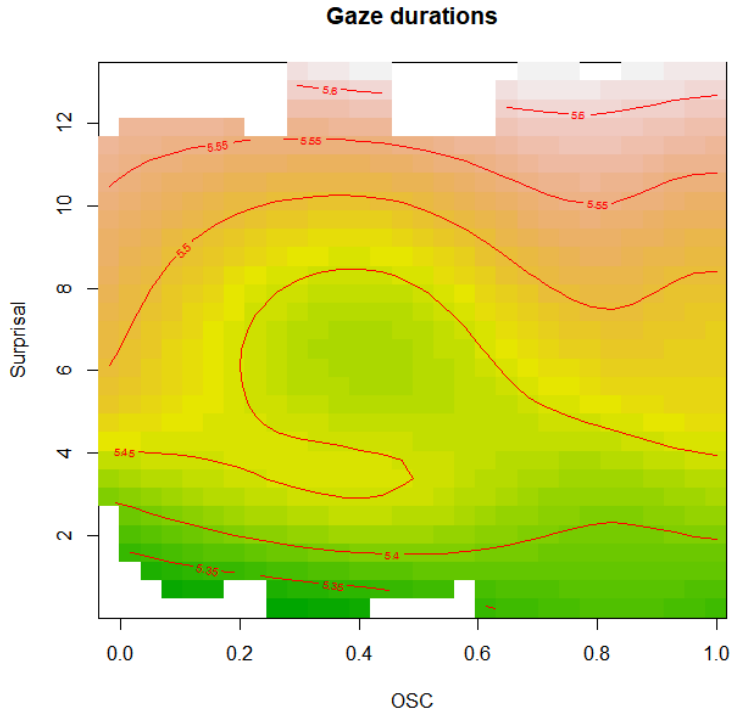

Figure 1. Tensor product smooth for the interaction of OSC (x-axis) and Surprisal (y-axis), color shades indicate different log-transformed gaze durations with green shades indicating shorter gaze durations and red shades indicating longer gaze durations.

References:

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. *Psychology of Learning and Motivation, 65*, 217–247.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive processing load in sentence comprehension. *Topics in Cognitive Science, 5,* 475–494.

Frank, S. L., Monsalve, I. F., Thompson, R. L., & Vigliocco, G. (2013). Reading-time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods, 45,* 1182-1190.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language, 140,* 1–11.

Marelli, M., & Amenta, S. (2018). A database of orthography-semantics consistency (OSC) estimates for 15,017 English words. *Behavior Research Methods, 50(4),* 1482–1495.

Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic Transparency in Free Stems: The Effect of Orthography-Semantics Consistency on Word Recognition. *Quarterly Journal of Experimental Psychology, 68(8),* 1571–1583.

Smith, N. J., & Levy, R. (2008). Optimal processing times in reading: A formal model and empirical investigation. *Proceedings of the 30th Annual Conference of the Cognitive Science Society,* pp. 595–600.

Smith, N. J., & Levy, R. (2010). Fixation durations in first-pass reading reflect uncertainty about word identity. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society,* pp. 1313–1318.

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass, 9(8),* 311–327.