A cute horgous meets a scary timfil: how do we interpret novel words in context?

Ngoc-Anh Tran[1], Kazimierz Garstecki[2], Giovanni Cassani[2]

[1] Dep. of Methodology and Statistics [2] Dep. of Cognitive Science & Artificial Intelligence (Tilburg University)

g.cassani@tilburguniversity.edu

Background. Studies on form-meaning mappings in the lexicon have highlighted many systematic relations between the surface form of words and what they mean [1, 2, 3, 4]. Recent studies further showed how prime-target similarity and semantic neighbourhood density influence lexical decision latencies also for pseudowords [5, 6], blurring the boundary between words and pseudowords. From a language learning perspective, such a blurred boundary makes sense: every word a person knows has been a pseudoword during development, and many valid words are pseudowords to many speakers, who might encounter them and have to quickly shape a semantic representation. Our work examines the interplay between semantic connotations conveyed by the word form itself [7] and by the sentence context in which the word form is first introduced [8, 9]. We focus on valence, i.e., how positive or negative a stimulus is perceived, building on valence ratings collected for isolated English pseudowords [10]. This abstract details the methodology and analysis plan we will follow.

Materials. In our behavioural experiment, each trial will consist of a free association (FA) task and a self-paced reading (SPR) task (order counterbalanced across subjects (see Figure 1)). The key manipulation involves the valence of the target (pseudo)word, as gauged from word [11] and pseudoword [10] ratings, and the valence of the sentence in which they will appear. Our targets consist of 40 pseudowords from [10] and 40 words from [12] (13 negative, 14 neutral, and 13 positive for both, determined from ratings when available and from definitions when not). We then created 40 sentences (13 negative, 14 neutral, 13 positive, determined based on the valence of the content words in the sentence [11]. For each participant, we sample 20 targets (10 words, 10 pseudowords: three positive, four neutral and three negative in each group), embedded in 20 sentences (seven positive, six neutral, and seven negative). (Pseudo)word and sentence valence will be fully crossed.

Hypotheses. We hypothesise that negatively-rated pseudowords (e.g., *horgous*) will elicit more negatively valenced associates than positively-rated pseudowords (e.g., *timfil*), even when embedded in a sentence with positive valence (and vice versa for positively-rated pseudowords). Moreover, reading times for the words immediately following the target pseudoword are expected to be longer when the pseudoword's valence conflicts with the sentence's valence, which would indicate processing costs.

Analysis plan. We will use thesauri as well as distributed semantic representations to estimate the i) valence and ii) semantic coherence of associates produced for a (pseudo)word. Then, we will compare how they change between participants who produced associates before seeing the target in context and participants who first saw the pseudoword in context, to study the influence that the sentence exerted on the target (pseudo)word's associations. Moreover, we will analyse reading times on the word immediately following the target. Associates' coherence and valence, and RTs will be analysed using mixed models: the key predictors will be (pseudo)word and sentence valence. We will further control for word length, plausibility, orthographic neighbourhood density, and orthographic overlap with the likeliest word to occur in the sentence context. Finally, we will consider the surprisal of the contextualised target (pseudo)words in different Large Language Models to investigate whether these models capture form-meaning mappings in context.

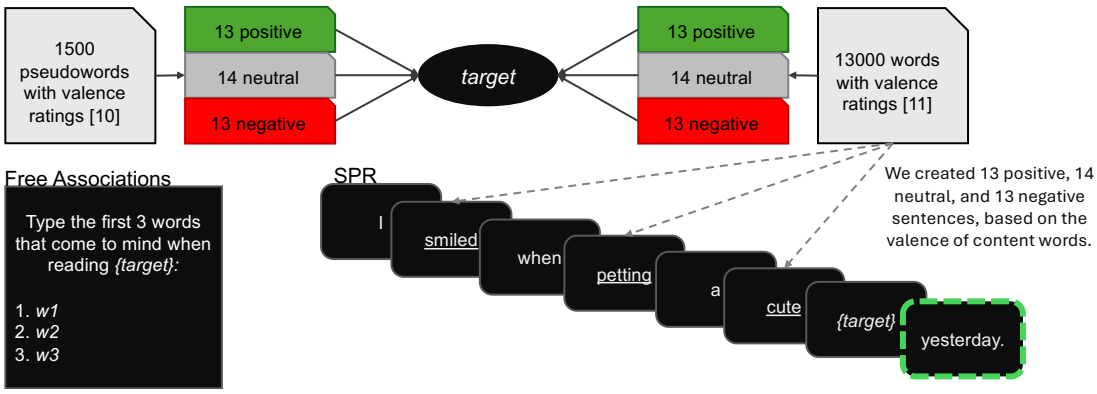

Figure 1: Experimental design: given a target, half the subjects will produce associates right before the Self-Paced Reading task; the other half half will first perform the SPR task and then produce associates right after, to compare associates produced for isolated and contextualised targets. We analyse the reading time at the first post-target word (dashed green contour) and the valence and semantic coherence of associates.

References

[1] Sidhu, D. M. (2025). Sound symbolism in the lexicon: A review of iconic-systematicity. *Lang. Linguist. Compass*, *19*(1), e70006.

[2] Ćwiek, A., Fuchs, S., Draxler, C., Asu, E. L., Dediu, D., Hiovain, K., Kawahara, S., Koutalidis, S., Krifka, M., Lippus, P., et al. (2022). The bouba/kiki effect is robust across cultures and writing systems. *Philos. Trans. R. Soc. B*, *377*(1841), 20200390.

[3] Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *PNAS*, *113*(39), 10818–10823.

[4] Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled/r/is associated with roughness, linking sound and touch across spoken languages. *Sci. Rep.*, *12*(1), 1035.

[5] Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *J. Exp. Psychol. Learn. Mem. Cogn.*, *47*(1), 157.

[6] Gatti, D., Marelli, M., & Rinaldi, L. (2023). Out-of-vocabulary but not meaningless: Evidence for semantic-priming effects in pseudoword processing. *J. Exp. Psychol. Gen.*, *152*(3), 851.

[7] Gatti, D., Rodio, F., Rinaldi, L., & Marelli, M. (2024). On humans'(explicit) intuitions about the meaning of novel words. *Cogn.*, *251*, 105882.

[8] Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *J. Exp. Psychol. Learn. Mem. Cogn.*, *27*(1), 225.

[9] Savic, O., Unger, L., & Sloutsky, V. M. (2022). Exposure to co-occurrence regularities in language drives semantic integration of new words. *J. Exp. Psychol. Learn. Mem. Cogn.*, *48*(7), 1064.

[10] Gatti, D., Raveling, L., Petrenco, A., & Günther, F. (2024). Valence without meaning: Investigating form and semantic components in pseudowords valence. *Psychon. B. Rev.*, *31*(5), 2357–2369.

[11] Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behav. Res. Methods*, *45*, 1191–1207.

[12] Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 english lemmas. *Behav. Res. Methods*, *51*, 467–479.