

Semantic Pseudoword Creation

Rolando Bonandrini¹, Rolf Harald Baayen², Marco Marelli¹

¹ *Department of Psychology, University of Milan-Bicocca, Italy*

² *Quantitative Linguistics Lab, Seminar für Sprachwissenschaft, Eberhard Karls Universität Tübingen, Germany*

Background: To answer the semasiological question “what does X mean?”, we rely on the mechanisms that define how form maps onto meaning [1–3], regardless of whether the incoming word is familiar or novel [4–6]. Yet whenever we write or speak, rather than read or listen, we often have a specific idea (i.e., meaning) in mind which we need to convey. This constitutes an onomasiological problem (“how do I express X?”). Even though, in most circumstances this process entails choosing the right known word to express a given concept, humans are –crucially – able to produce novel words so that others can later identify their meaning [7]. Still –to date –no computational approach has been specifically proposed to produce novel word forms (i.e., not speech/spelling errors) starting from a given semantic input. In this study, we explored whether regularities between form and meaning estimated through Linear Discriminative Learning [8–12] can profitably be leveraged for novel word creation from predefined semantic information, and whether such novel words can profitably convey meaning to humans.

Method: To generate novel words, we developed a Semantic Pseudoword Creator (SPC) procedure (Figure 1). The SPC builds on the Discriminative Lexicon Model and Linear Discriminative Learning [8–12], which we used to estimate the mapping between orthography and meaning for existing words. We then used such mapping to derive (given a set of target semantic vectors) orthography vectors, which were eventually fed to a trigram sorting algorithm to produce novel words. Target vectors either corresponded to the meaning of a word, or to the average meaning of two words. We then conducted three online experiments to explore whether human participants were able to identify a generated pseudoword resulting from one single word prompt (i.e., a “pseudo-synonym”, Experiment 1), a pair of word prompts whose combined meaning produced a certain pseudoword (Experiment 2), and a generated pseudoword resulting from the combined meaning of a pair of word prompts (Experiment 3).

Results: Across all three experiments, participants proved to be able to identify the correct pseudoword(s)-prompt(s) association, and estimated semantic similarity between response options proved to explain human performance over and above surface similarity.

Discussion: These results suggest that it is possible to use the association between orthography and meaning to engineer novel words from a specific meaning. Indeed, data indicates that novel word forms created with the SPC approach can effectively convey meaning to human participants, regardless of whether meaning corresponds to an existing concept or rather a novel conceptual combination. Leveraging form-meaning mapping to generate novel words with a specific meaning may open new opportunities in psycholinguistic research, such as the possibility to create novel words belonging to specific syntactic and semantic categories, or the study of how uncharted areas of the semantic space may become lexicalized.

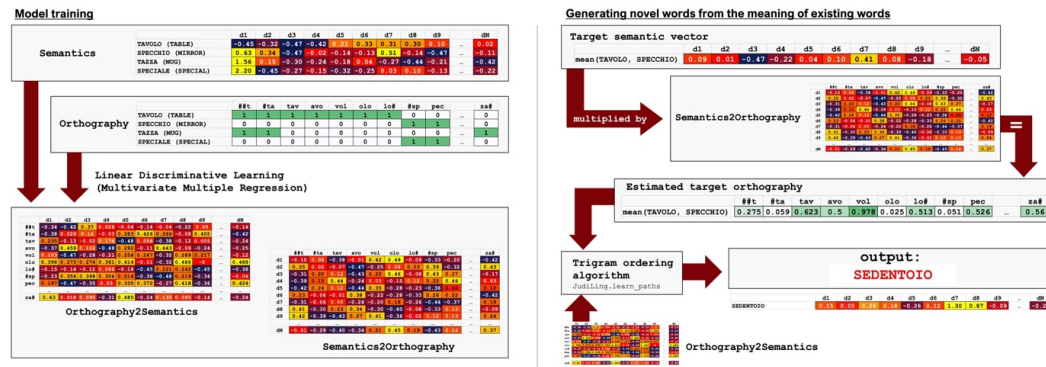


Figure 1: Schematic representation of the SPC procedure. Please note that, in the context of the DLM literature, the Orthography2Semantics matrix is known as *F*, while the Semantics2Orthography matrix is known as *G*.

References

- [1] Amenta, S., Marelli, M., & Sulpizio, S. (2017). From sound to meaning: Phonology-to-Semantics mapping in visual word recognition. *Psychonomic Bulletin & Review*, 24(3), 887–893. <https://doi.org/10.3758/s13423-016-1152-0>
- [2] 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. <https://doi.org/10.1080/17470218.2014.959709>
- [3] 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. <https://doi.org/10.3758/s13428-018-1017-8>
- [4] Hendrix, P., & Sun, 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(1), 157–183. <https://doi.org/10.1037/xlm0000819>
- [5] Gatti, D., Marelli, M., & Rinaldi, L. (2023). Out-of-vocabulary but not meaningless: Evidence for semantic-priming effects in pseudoword processing. *Journal of Experimental Psychology: General*, 152(3), 851. <https://doi.org/10.1037/xge0001304>
- [6] Bonandrini, R., Amenta, S., Sulpizio, S., Tettamanti, M., Mazzucchelli, A., & Marelli, M. (2023). Form to meaning mapping and the impact of explicit morpheme combination in novel word processing. *Cognitive Psychology*, 145, 101594. <https://doi.org/10.1016/j.cogpsych.2023.101594>
- [7] Pugacheva, V., & Günther, F. (2024). Lexical choice and word formation in a taboo game paradigm. *Journal of Memory and Language*, 135, 104477. <https://doi.org/10.31234/osf.io/7x8sb>
- [8] Baayen, R., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. *Complexity*, (1), 4895891.
- [9] Chuang, Y.-Y., & Baayen, R. (2021). Discriminative learning and the lexicon: NDL and LDL.
- [10] Heitmeier. (2025). *The Discriminative Lexicon: Theory, Implementation in the Julia Package JudiLing, and Applications*. Cambridge University Press.
- [11] Heitmeier, M., Chuang, Y.-Y., & Baayen, R. (2021). Modeling morphology with linear discriminative learning: Considerations and design choices. *Frontiers in psychology*, 12, 720713. <https://doi.org/10.3389/fpsyg.2021.720713>
- [12] Heitmeier, M., Chuang, Y.-Y., & Baayen, R. (2023). How trial-to-trial learning shapes mappings in the mental lexicon: Modelling lexical decision with linear discriminative learning. *Cognitive Psychology*, 146, 101598. <https://doi.org/10.1016/j.cogpsych.2023.101598>