PseudoConnector: A game of (non)lexical selection and production

Iva Šaban[1], Marco Marelli[1], Abhilasha A Kumar[2]

[1] *Department of Psychology, University of Milano-Bicocca, Italy*
[2] *Department of Psychology, Bowdoin College, USA*

Background: Efficient communication depends upon many cognitive processes which have to be coordinated. Communication allows humans to encode a message into signal that they think is the best cue for their recipient to infer intended meaning. On their side, receiver has to decode that signal back into a message. To achieve the successful communication process, both parties rely on words from their lexicon as a versatile tool of conveying meaning [1].

An important issue in verbal communication is that the correspondence between word and meaning is not entirely arbitrary. This is due to the functional pressure of language being easy to learn and understand on one side, and necessity of covering the variety of concepts on the other side [2, 3]. This form of non-arbitrariness, known as systematicity, can be extended beyond the boundaries of existing languages, and applies even to non-existing letter strings, which are not part of mental lexicon and supposedly lack conventional meaning (i.e., pseudowords). As adults, we are often exposed to novel words we have never encountered before [4], but have some intuition about their meaning, based on the form or context they appear in [5–8].

The present work aims to investigate how people use orthographic and semantic clues to infer and decode intended meaning, similarly to everyday communication. To do so, we developed a game inspired by *Connector* [9], in which players generate one-word clues that allow a teammate to pick out a target pair of real words from a grid. The present game, named *PseudoConnector*, extends this paradigm by focusing on novel (i.e., non-lexical) items, by incorporating both *speaker* (Experiment 1) and *listener* (Experiment 2) roles.

Method: The way people assign meanings to pseudowords could be explained by distributional semantic models (DSMs). These models use linguistic corpora to build a large base of word co-occurrence information. That is, they represent words as high-dimensional numerical vectors, and rely on assumption that word with similar meanings occur in similar contexts [10]. *Fasttext* [11, 12] is a DSM that can produce quantitative representations of the novel words by computing their semantic representation as the sum of the vectors of the n-grams that compose it. We used *fasttext* to compute semantic similarity for a set of pseudowords. We then manually selected a set of 14 high-similarity and 14 low-similarity pseudoword pairs, which together with two distracters formed a total of 28 game boards (see examples in Figure 1).

These boards were presented to speakers in Phase 1 who had to select two pseudowords that go together as a pair. After each selection, they were asked to rate how confident they are about their selection, on a 1-to-5 scale. In Phase 2, the same participants saw different combinations of pseudowords (*focal pair*), all from the boards presented in Phase 1. They were asked to produce a one-word English clue that best captures connections between the two pseudowords and would allow the other player to select that target pair. This phase consists of 56 trials. After each produced clue, participants were asked how likely they think the other player will guess the pair of pseudowords the clue word was referring to. In Experiment 2 (no data collected yet), listeners will be presented with 56 trials containing the same boards and clues produced by the corresponding speaker. Their task will be to guess the two pseudowords speakers were referring to when producing a clue and provide their confidence rating on a 1-to-5 scale (see Figure 2).

Results: At the moment of submission of this abstract, we conducted a pilot study (N=4), showing that speakers were able to correctly identify target pairs above chance level on both high-similarity and low-similarity pairs. They were also slightly faster and more confident on high-similarity relative to low similarity pairs. Analysis of the clue production in Phase 2 is yet to be performed.

Discussion: People seem to be able to form intuitive associations between novel words, use semantic clues to convey those associations and infer intended novel word pairs in a communicative context.

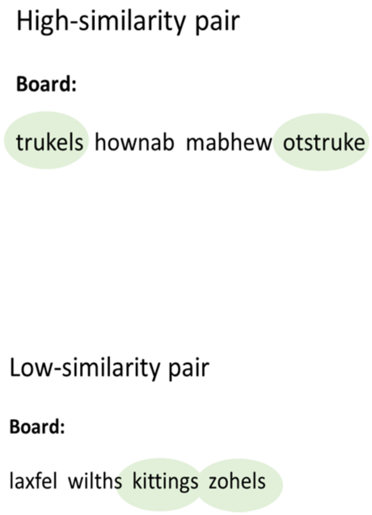

Figure 1: Examples of a board with high-similarity pseudoword pair (upper panel) and low-similarity pseudoword pairs (lower panel), marked in green (Phase 1).

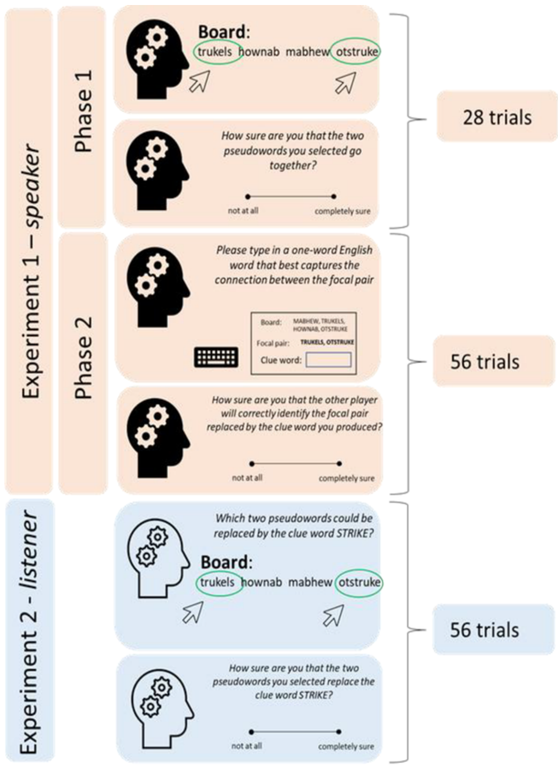

Figure 2: Overview of the study.

References

[1] Shannon, C., & Weaver, W. (1949). *The Mathematical Theory of Communication.* The University of Illinois Press.

[2] Dingemanse, M., Blasi, D., Lupyan, G., Christiansen, M., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. https://doi.org/10.1016/j.tics.2015.07.013

[3] Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1152–1164. https://doi.org/10.1037/a0027747

[4] Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–458. https://doi.org/10.1037/xhp0000159

[5] 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

[6] 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), 851–863. https://doi.org/10.1037/xge0001304

[7] Günther, F., & Marelli, M. (2020). Trying to make it work: Compositional effects in the processing of compound "nonwords". *Quarterly Journal of Experimental Psychology*, *73*(7), 1082–1091. https://doi.org/10.1177/1747021820902019

[8] 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

[9] Kumar, A., Steyvers, M., & Balota, D. (2021). Semantic Memory Search and Retrieval in a Novel Cooperative Word Game: A Comparison of Associative and Distributional Semantic Models. *Cognitive Science*, *45*(10), 13053. https://doi.org/10.1111/cogs.13053

[10] Harris, Z. (1954). Distributional Structure. Word. https://doi.org/10.1080/00437956.1954.11659520

[11] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/10.1162/tacl_a_00051

[12] Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). *Advances in Pre-Training Distributed Word Representations. Proceedings of the Eleventh International Conference on Language Resources and Evaluation.* LREC. https://doi.org/10.48550/arXiv.1712.09405