From quiz show to mental lexicon: Sub-Lexical patterns inform the semantics of unknown words

Simona Amenta[1], *Marco A Petilli*[2], *Luca Borgonovo*[2], *Marco Marelli*[2]

[1] *Department of Informatics, Systems and Communication, University of Milano-Bicocca, Italy*
[2] *Department of Psychology, University of Milano-Biocca, Italy*

Background: How do we understand the meaning of unknown words? It has been estimated that adult speakers of a given language encounter a new word every two days [1]. In recent years, research on novel word processing highlighted how individuals rely on sub-lexical information (i.e., recurring character n-grams) to infer meaning [2–4]. This research has however mostly focused on made-up words, especially devised to resemble legal words of a given language, but without corresponding lexical or semantic representations. While these items represent an ideal case to study unknown word understanding, there is also evidence that made-up words and existing words differ significantly on a distributional level [5]. To overcome this limitation, and to work with ecological linguistic material, in the current study we turned to rare words (i.e., low frequency), that while part of the lexicon, are unlikely to be known by most speakers. Moreover, instead of relying on classic experimental paradigms (e.g., lexical decision, semantic priming), we analysed materials from a quiz show, where contestants had to guess the correct definition of a rare word (see below). Our aim was to investigate whether sub-lexical information systematically guides the inference of unknown word meanings in natural lexical uncertainty.

Method: Our dataset included linguistic materials and contestants' responses taken from a popular Italian quiz show. In the game, contestants are presented with a rare Italian word and a series of definitions (e.g., "A type of chair"), shown one at a time. For each definition, they must decide whether it is correct or not. If they reject it, another definition is presented. This continues until they select a definition. If they select the correct one, they win; if the selected definition is incorrect, another contestant takes the turn and continues with the remaining definitions (Figure 1). We harvested data from 1,247 episodes, with a total of 2,177 words and 13,652 definitions. Word embeddings for each target word and its corresponding definitions were induced with an off-the-shelf FastText model [6], allowing us to estimate semantic representations based on sub-lexical information of the targets. Cosine similarity (i.e., semantic similarity) between a target word and each corresponding possible definition was the main predictor of our analyses. We first fitted a logit mixed-effects model to test whether the cosine similarity predicted the probability of selecting that definition (i.e., a Yes response). We then fitted a mixed-effect regression model to investigate whether cosine similarity predicted response times and whether this effect differed across response types (hits, correct rejections, false alarms, and wrong rejections). In both models, we controlled for orthographic overlap between words and definitions, to ensure effects were not driven by surface similarity.

Results: Results indicate that, the more a definition is semantically closer to the estimated meaning of the rare word, the more likely participants are to select it when it is correct. Moreover, we observe a general effect on response times: irrespective of the type of response, participants take longer to respond to definitions that are semantically closer to the estimated meaning of the rare word (Figure 2).

Discussion: Our results show that, when confronted with an unknown word, individuals exploit semantic information activated by sub-lexical units to infer its meaning. Importantly, individuals seem to be able to map the induced semantics onto that of possible definitions, and to use the resulting semantic similarity to guide their choices: when the similarity is higher, individuals spend more time evaluating the definition before responding. This pattern suggests that semantic similarity functions as a graded decision signal during lexical inference. Our results are in line with previous research on novel word processing which showed how language-model-induced semantic representations of made-up words align with human intuitions about their possible meaning [3, 4]. Crucially, we extend this literature by showing that the same mechanism operates in ecologically valid settings, supporting the interpretation

of existing but unknown words.

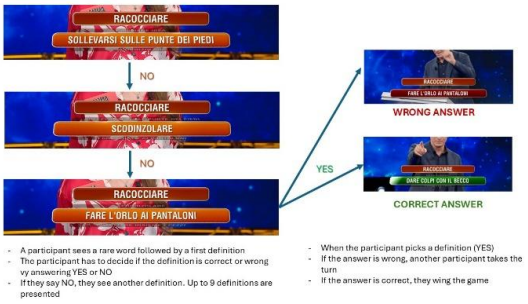

Figure 1: Game flow.

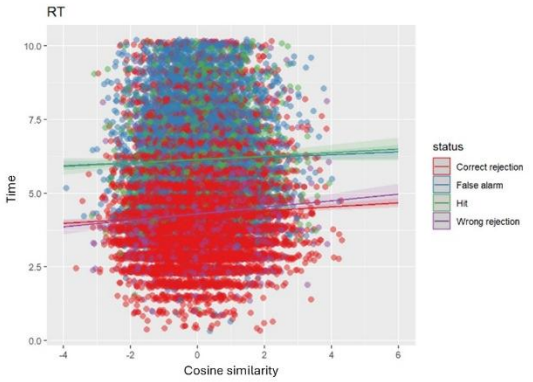

Figure 2: Relationship between cosine similarity and response time (RT) across response status. Each point represents a single trial. Lines show model-predicted effects from a linear mixede ffects regression predicting RT from cosine similarity, response status (hits, correct rejections, false alarms, and wrong rejections), and their interaction. Shaded areas represent 95% confidence intervals.

References

[1] Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How Many Words Do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age. *Frontiers in Psychology*, *7*, 1116. https://doi.org/10.3389/fpsyg.2016.01116

[2] 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

[3] Varda, A. D., Gatti, D., Marelli, M., & Günther, F. (2024). Meaning beyond lexicality: Capturing pseudoword definitions with language models. *Computational Linguistics*, *50*(4), 1313–1343.

[4] Gatti, D., Rodio, F., Rinaldi, L., & Marelli, M. (2024). On humans' (explicit) intuitions about the meaning of novel words. *Cognition*, *251*, 105882. https://doi.org/10.1016/j.cognition.2024.105882

[5] Günther, F., & Marelli, M. (n.d.). The compositionality of novel compounds.

[6] 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