How Linguistics Can Contribute to LLM Development, and Vice-Versa: Insights from Fine-Tuned Models' Polysemy Processing

Hanjung Lee
hanjung@skku.edu
Sungkyunkwan University

This talk investigates how contextualized language models can contribute to the study of lexical polysemy, and how linguistic analysis can in turn reveal the representational limits and biases of such models. The empirical focus is on the polysemy of two English change-of-state verbs, *break* and *freeze*, whose senses range from concrete physical meanings to highly abstract and metaphorical extensions. In particular, the study addresses a notoriously problematic issue in theoretical semantic analysis: the problem of sense distinctions and boundaries in polysemy.

Using 1,188 labeled corpus instances extracted from COCA, four Transformer-based encoder models—BERT-base, BERT-large, RoBERTa-large, and ALBERT—were fine-tuned for meaning-class classification. The best-performing model, RoBERTa-large, achieved strong overall performance, with accuracy and macro F1 scores of approximately 0.87. However, performance varied considerably across sense categories. Some senses, such as *decoding*, *preservation*, *disclosure*, and *bodily harm*, were classified with high accuracy, whereas others, including *economical freezing*, *change*, and *emotional/mental freezing*, were substantially more difficult to distinguish.

Through detailed error analyses of semantic relationships among confusable senses and recurrent representational biases in model prediction, this study identifies the major semantic and distributional factors that sharpen or blur sense boundaries in verbal polysemy, while also revealing persistent tendencies of contextualized language models to converge toward prototypical, semantically generalized, or distributionally dominant interpretations.

The first part of the error analysis shows that model misclassifications reflect systematic semantic relationships between gold and predicted labels, including metaphorical extension, semantic subcategories, semantic overlap, context-dependent polysemy, and contextual implication. These relations differ in the strength of semantic connectivity they involve. Strongly connected senses—especially those linked through overlap, metaphorical extension, or shared superordinate structure—are more likely to be confused, suggesting that sense boundaries become blurred as semantic connectivity increases. These findings support a continuum-based view of polysemy (Haber and Poesio 2021, 2024) and highlight the importance of the strength of semantic connectivity in shaping sense distinctions and boundaries (Petersen and Potts 2023; Lee, to appear).

The second part of the analysis examines the direction of model prediction errors. The findings reveal recurrent representational biases in the model's learned semantic space, including semantic generality bias, prototypicality bias, dominant scenario fixation bias, and concretization bias. These biases help explain why the model converges on particular confusable senses rather than others. Notably, semantic generality bias and dominant scenario fixation account for the majority of observed biased errors, suggesting that statistical-distributional attractors may exert a stronger influence on model prediction than cognitively central semantic prototypes (in the sense of Tyler and Evans 2003).

The talk concludes by arguing for a reciprocal relationship between linguistics and LLM development. Contextualized language models provide powerful tools for investigating context-sensitive variation in meaning, but linguistic analysis is essential for diagnosing how models narrow, distort, or homogenize subtle semantic variation. In this sense, linguistics can help guide the development of more reliable, interpretable, and semantically sensitive language technologies (Grieve et al. 2025; Lee, to appear).

Selected References

Grieve, Jack, Sara Bartl, Matteo Fuoli, Jason Grafmiller, Weihang Huang, Alejandro Jawerbaum, Arika Murakami, Marcus Perlman, Dana Roemling & Bodo Winter. 2025. The sociolinguistic foundations of language modeling. *Frontiers in Artificial Intelligence* 7:1472411.

Haber, Janosch & Massimo Poesio. 2021. Patterns of lexical ambiguity in contextualized language models. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia & Scott Wen-tau Yih (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2021*, Punta Cana, Dominican Republic, 2663-2676. Association for Computational Linguistics.

Haber, Janosch & Massimo Poesio. 2024. Polysemy—Evidence from linguistics, behavioral science, and contextualized language models. *Computational Linguistics* 50(1). 351–417.

Lee, Hanjung. To appear. *Polysemy in Semantics with Contextualized Language Models: Distribution, Boundaries and Interpretation of Polysemous Senses* (Language and Computers: Studies in Digital Linguistics). Berlin and Boston: De Gruyter Brill.

Petersen, Erika & Christopher Potts. 2023. Lexical semantics with large language models: A case study of English "break". In Andreas Vlachos & Isabelle Augenstein (eds.), *Findings of the Association for Computational Linguistics: EACL 2023*, Dubrovnik, Croatia, 490-511. Association for Computational Linguistics.

Tyler, Andrea & Vyvyan Evans. 2003. *The Semantics of English Prepositions. Spatial Scenes, Embodied Meaning and Cognition*. Cambridge: Cambridge University Press.