Title: **THE SQUIRREL TEST: COMPETING FORMS FOR NEW WORDS IN INTERACTION**

Authors: Jiayi Zhou & Mădălina Zgreabăn

Abstract

There is little work on why some competing forms of neologisms become more used than others, if humans prefer neologisms similar to phonologically and morphologically known words, or vice versa. If large language models (LLMs) possess Theory of Mind (ToM) is also still a debated topic in natural language processing. To bridge these gaps, we present a research proposal on the 'squirrel test', a research experiment aimed to test both humans and LLMs on their preference of neologisms. The test examines the preferences of humans' and LLMs' in four conditions: listener with explicit interlocutor, listener with implicit interlocutor, speaker with explicit interlocutor and speaker with implicit interlocutor, having the following research questions: Is there any preference from the speakers' and listeners' sides with explicit or implicit participants? Do LLMs have similar preferences with those of humans?

The preferences of the participants will be investigated w.r.t. five different types of neologisms, i.e., acronyms, lexical borrowings, portmanteaus, morphological derivations and arbitrary coinages, with the same meaning and number of syllables in English.

Preferences of speakers and listeners with an implicit or explicit interlocutor, as well as their reaction times will be recorded. We expect speakers to choose more familiar forms, while listeners more explicit forms, due to pragmatic theories of information inference and the principle of less effort. Speakers and listeners in the explicit interlocutor condition are expected to choose intermediate forms. LLMs will be also tested, having their surprisal measured. We expect no preference of LLMs for certain neologisms and, consequently, no surprisal, under the hypothesis that LLMs do not have ToM. Contrary results could offer evidence of ToM in LLMs.

The results would offer insights into humans' preferences, LLMs' alignment with them, and LLMs' ToM abilities. These aspects prove important for LLMs performance in human-computer interaction settings, and for better understanding general language trends w.r.t. neologisms.

Selected References

Bach, K., & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

Cook, P. (2012). Using social media to find English lexical blends. In Proceedings of the 15th EURALEX International Congress (EURALEX 2012), pages 846–854, Oslo, Norway.

Ferrer i Cancho, R., & Sole, R. V. (2003). Least effort and the origins of scaling in human language. Proceedings of the National Academy of Sciences of the United States of America, 100(3), 788–791.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. Science, 336(6084), 998-998.

Guz, W. (2012). Are nonce words really deviant, context-dependent, and unlexicalizable. Sound structure and sense: studies in memory of Edmund Gussmann, 223-238.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. Science, 336(6084), 1049–1054.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic Typology and Efficient Communication. Annual Review of Linguistics 4(1):109–128.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. Cognition, 141, 87-102.

Lei, S., Yang, R., & Huang, C. R. (2021). Emergent neologism: A study of an emerging meaning with competing forms based on the first six months of COVID-19. Lingua, 258, 103095.

Li, B., Zhu, Z., Thomas, G., Xu, Y., & Rudzicz, F. (2021). How is BERT surprised? Layerwise detection of linguistic anomalies. arXiv preprint arXiv:2105.07452.

Michaelov, J. A., Bardolph, M. D., Van Petten, C. K., Bergen, B. K., & Coulson, S. (2023). Strong Prediction: Language model surprisal explains multiple N400 effects. Neurobiology of Language, 1-71.

Misra, K., Ettinger, A., & Rayz, J. T. (2020). Exploring BERT's Sensitivity to Lexical Cues using Tests from Semantic Priming. arXiv preprint arXiv:2010.03010.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. Behavioral and Brain Sciences, 13(4), 707–784.

Ryskina M., Rabinovich, E., Berg-Kirkpatrick, T., Mortensen, D., & Tsvetkov, Y. (2020). Where New Words Are Born: Distributional Semantic Analysis of Neologisms and Their Semantic Neighborhoods. In Proceedings of the Society for Computation in Linguistics 2020, 367–376, New York. Association for Computational Linguistics.

Schmid, H. J. (2008). New words in the mind: Concept-formation and entrenchment of neologisms.

Shannon, C. E., & Weaver, W. (1949). The Mathematical Theory of Communication. University of Illinois Press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press.

Zu, Y. (2022, October). Research on Intelligent Construction of China English Network New Words Database Based on Adjacent Entropy Recognition Algorithm. In 2022 IEEE 4th Eurasia Conference on IOT, Communication and Engineering (ECICE) (pp. 381-385). IEEE.