Predictability and surprisal as approximators of information status

Andrew Dyer (Language Science and Technology, Saarland University)
andrew.dyer@uni-saarland.de

Language model surprisal is often used as a rough measure of the difficulty of processing language (Goldstein et al., 2022; Wilcox et al., 2023), with more surprising tokens held to correspond to more difficult or contentful units of speech. This is often extended to “novel and unexpected” information (Xu and Futrell, 2024), with the implicit linking hypothesis that new information is more surprising. This posits a link to information status: the givenness or newness of entities and mentions in discourse (Chafe, 1976), which is itself known to affect the effort in processing words and sentences (Asahara, 2017).

Information status captures what speakers find predictable given previous context (Prince, 1981). This is mirrored by the learning objective of language models- maximising predictability of upcoming tokens- and the attention to long-range context in transformer-based models. The implicit topic-modeling in such models also mirrors the view that given information corresponds to that which is topical (Givón, 1983). This accounts for bridging references in discourse, where entities not previously introduced are nonetheless unsurprising due to their semantic link with previous context (Clark, 1977; Clark and Haviland, 1977). From this perspective, it is credible that sufficiently context-aware language models’ surprisal values could approximate the information status of referents and mentions in discourse as experienced by human interlocutors.

On the other hand, this view contrasts with the more explicit view of information status usually evident in the design of information status and coreference corpora, whereby referents in discourse are explicitly assigned an information status attribute by the receiver at each mention in the discourse depending strictly on whether they have been mentioned, be that categorical (Gundel, Hedberg, and Zacharski, 1993) or gradient (Arnold and Griffin, 2007).

Despite the interest in these conflicting views, there has as yet been no direct corpus-based study of the extent to which language model surprisal is correlated with, or is predictive of, information status- and vice-versa. A finding that language models approximate information status would be both a contribution to the debate on the nature of information status representation and effects (Arnold, 2016); and support for the practical approach of using language-model surprisal as a quantitative measure or stand-in for information status.

To study this, we will measure the link between information status and transformer based language model surprisal on the English and Portuguese portions of CiepInf (Dyer et al., 2024) a parallel multilingual corpus annotated for information status and coreference. We will compare the performance of a set of language models with different parameter sizes, architectures and context-sizes. We aim to shed light on the extent to which information status correlates with, or predicts, surprisal, and vice-versa; and the extent to which surprisal informs the actual forms of language use in new and given mentions. If such a pattern of interaction is found, we will have some more evidence to support the predictability-based view of information status.

Arnold, Jennifer E. (2016). “Explicit and Emergent Mechanisms of Information Status”. en. In: Topics in Cognitive Science 8.4, pp. 737–760. ISSN: 1756-8765. DOI: 10.1111/tops.12220. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/tops.12220 (visited on 09/25/2024).
Arnold, Jennifer E. and Zenzi M. Griffin (May 2007). “The effect of additional characters on choice of referring expression: Everyone counts”. eng. In: Journal of Memory and Language 56.4, pp. 521–536. ISSN: 0749-596X. DOI: 10.1016/j.jml.2006.09.007.
Asahara, Masayuki (Nov. 2017). “Between Reading Time and Information Structure”. In: Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation. Ed. by Rachel Edita Roxas. The National University (Philippines), pp. 15–24. URL: https: //aclanthology.org/Y17-1006.
Chafe, Wallace L. (1976). “Givenness, contrastiveness, definiteness, subjects, topics, and point of view”. In: Subject and Topic. Ed. by Charles N. Li. New York: Academic Press, pp. 25–55.
Clark, Herbert H (1977). “Bridging”. English. In: Thinking: Readings in Cognitive Science. Cambridge: Cambridge University Press, pp. 411–420.
Clark, Herbert H and Susan E Haviland (1977). “Comprehension and the Given-New Contract”. en. In: Discourse Production and Comprehension. DISCOURSE PROCESSES: ADVANCES IN RESEARCH AND THEORY 1. Norwood, New Jersey: ABLEX PUBLISHING CORPORATION, pp. 1–38.
Dyer, Andrew et al. (Dec. 2024). “A Multilingual Parallel Corpus for Coreference Resolution and Information Status in the Literary Domain”. In: Proceedings of the 22nd International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2024). Hamburg, Germany: Association for Computational Linguistics.
Givón, Talmy (1983). “Topic continuity in discourse: An introduction”. In: Topic continuity in discourse: A quantitative cross-language study. Goldstein, Ariel et al. (Mar. 2022). “Shared computational principles for language processing in humans and deep language models”. In: Nature Neuroscience 25.3. Publisher: Nature Publishing Group, pp. 369–380. ISSN: 1546-1726. DOI: 10.1038/s41593-022-01026-4. URL: https://www.nature.com/articles/s41593-022-01026-4 (visited on 09/23/2024).
Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski (1993). “Cognitive Status and the Form of Referring Expressions in Discourse”. In: Language 69.2. Publisher: Linguistic Society of America, pp. 274–307. ISSN: 0097-8507. DOI: 10.2307/416535. URL: https://www.jstor.org/stable/416535 (visited on 08/20/2024).
Prince, Ellen F. (1981). “Toward a taxonomy of given-new information”. In: Syntax and semantics: Vol. 14. Radical Pragmatics. Ed. by P. Cole. New York: Academic Press, pp. 223–255.
Wilcox, Ethan G. et al. (2023). “Testing the Predictions of Surprisal Theory in 11 Languages”. In: Transactions of the Association for Computational Linguistics 11, pp. 1451–1470. DOI: 10.1162/tacl_a_00612. URL: https://aclanthology.org/2023.tacl-1.82.
Xu, Weijie and Richard Futrell (Mar. 2024). “Syntactic dependency length shaped by strategic memory allocation”. In: Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP. Ed. by Michael Hahn et al. St. Julian’s, Malta: Association for Computational Linguistics, pp. 1–9. URL: https://aclanthology.org/ 2024.sigtyp-1.1 (visited on 03/26/2024).