

Co-Pilots for Frame Semanticists

Tiago Timponi Torrent¹, Arthur Lorenzi Almeida² & Mark Turner³

¹ Federal University of Juiz de Fora | FrameNet Brasil | CNPq, tiago.torrent@uff.br

² Federal University of Juiz de Fora | FrameNet Brasil, arthur.lorenzi@estudante.uff.br

³ Case Western Reserve University | Red Hen Lab, turner@case.edu

Keywords: Frame Semantics, FrameNet, AI, Lutma, Frame Creation, Large Language Model

A Large Language Model (LLM) is a text-generation system trained on vast amounts of text. To generate text, the model must have developed a way to deal implicitly with the structures of languages so that it mimics the conceptual patterns underlying languages, at least in the eyes of a cooperative human comprehender. The model succeeds only to the extent that the human user judges its output to be appropriate. In this talk, we do not focus on whether LLMs have any understanding of language or not. Rather, we discuss whether the LLM can be used as a tool for the researcher to do linguistics. Therefore, this talk reframes the discussion about LLMs, from what they can do with language to what they can do for linguists. Namely, we aim to answer the following questions: Can LLMs be prompted to help the linguist analyze cognition and language? Can LLMs leverage whichever representations they build for correlations between linguistic items so as to express conceptual and linguistic structures?

Creating an entry for FrameNet is extraordinarily laborious (Fillmore et al., 2003). Although more than a thousand conceptual frames—and the lexical units evoking them—have been characterized in FrameNet, many more still need to be built. To facilitate this process, tools such as Lutma—“a collaborative, semi-constrained, tutorial-based tool for contributing frames and lexical units to the Global FrameNet initiative.” (Torrent et al., 2022) been built. In this context, we further specify our initial questions as follows: Can an LLM assist the researcher who is attempting to expand FrameNet? Can an LLM assist the user of existing frame creation tools?

To answer those questions, we conduct an exploratory study focusing on a particular LLM—ChatGPT, developed by OpenAI. The possibilities for ChatGPT as an eliciting interlocutor for the linguist are extensive. The conversational interface allows for continuous dialogue with the system, which can present text in a variety of formats; write code potentially useful for frame-building tools; and, when informed of its mistakes, respond repeatedly and inventively to requests to correct them.

We present routines for prompting ChatGPT to assist linguists in frame building. We also analyze the limitations of the model and some of its potential risks. Results indicate that ChatGPT can propose structures similar to frames and frame elements; propose subframes and super-frames, including inheritance structure; list lexical units evoking a newly created frame; and provide examples and analyses.

Our pilot study found that ChatGPT has serious limits and is no substitute for the trained frame semanticist. It can make mistakes in proposing (a) attribute frames that lack a core frame element that indicates the attribute that is associated with the entity; (b) entity frames whose list of lexical units is structured not by consistent lexicographic properties but rather by an ontology that derives from a check-list theory of meaning as composed of combinations of criterial properties; (c) event frames where core frame elements are missing, while other frame elements are inappropriately included in the frame even though they are a frame on their own.

References

- Fillmore, Charles J., Miriam R.L. Petruck, Josef Ruppenhofer & Abby Wright. 2003. Framenet in Action: The Case of Attaching. *International Journal of Lexicography* 16(3). 297–332. doi:10.1093/ijl/16.3.297. <https://doi.org/10.1093/ijl/16.3.297>.
- Torrent, Tiago Timponi, Arthur Lorenzi, Ely Edison Matos, Frederico Belcavello, Marcelo Viridiano & Maucha Andrade Gamonal. 2022. Lutma: A frame-making tool for collaborative FrameNet development. In *Proceedings of the 1st workshop on perspectivist approaches to nlp @Irec2022*, 100–107. Marseille, France: European Language Resources Association. <https://aclanthology.org/2022.nlperspectives-1.13>.