

Discourses as units of knowledge in the light of neural language models. Refinement of the theory of discursive space

Rafal Maciag¹

¹Institute of Information Studies, Jagiellonian University, Poland

Abstract

In recent years, and even months, a rapid development of NLP solutions can be observed. This technology allows one to define deeper semantic inferences in the text based on the idea of neural language models (NLMs). Neural language models (NLMs) are containers of knowledge. The relationship between language and knowledge has been extensively reflected in research in the form of the so-called discourse analysis. Based on Michel Foucault's concept of discourse, especially the text from 1971 (Foucault, 1971), knowledge model was proposed named discursive space, in which discourses as instances of knowledge travel trajectories in a multidimensional dynamical space (Maciag, 2022). The idea presented in the paper assumed that it is possible to isolate semantic structures more complex than the semantic units used so far, i.e. tokens, which are based on words and their relationships in sentences. Such structures are discourses, i.e. linguistic (semantic) structures with a higher degree of abstraction than the sentences they consist of. Therefore, one should search for higher-order units (discourses) composed of lower-order semantic units (words) and their relations in sentences. It would be a repetition of the embedding technique used in NLM, but transferred to a higher semantic level, the aim of which is to create a set of vectors describing discourses. By analyzing the mutual position of the indicated discourses in the corpus of texts, a discursive linguistic model would be created. The introduction of a variable in the form of time, i.e. the construction of a dynamical discursive model, would fulfill the assumptions of discursive space.

Keywords: *natural language processing; neural language models; knowledge; discourse; discursive space.*

Foucault, M. (1971). *L'ordre du discours: Leçon inaugurale au Collège de France prononcée le 2 décembre 1970*. Paris: Gallimard.

Maciag, R. (2022). Theory of Knowledge Based on the Idea of the Discursive Space. *Philosophies*, 7(4), 72. doi: 10.3390/philosophies7040072