Integrating Frame Semantics in Lexical Substitution Tasks to Improve Lexical Precision

Ana Luiza Vianna [1], Mikaela Martins [1], Eduardo Cortes [2], Sandro Rigo [1], Rafael Kunst [1]
[1]UNISINOS University, [2] Federal University of Rio Grande do Sul
alvianna@unisinos.edu.br, mikaelalm@unisinos.br, egcortes@inf.ufrgs.br, rigo@unisinos.br, rafaelkunst@unisinos.br

Keywords: Frame Semantics, Lexical Substitution, Natural Language Processing

Due to the recent boom in technology development, we have witnessed in the past decades, doors have been opened for enterprises that link language and technology. In some cases, it might be necessary for companies to adapt the language used in software for it to be useful in certain domains. Thus, it is necessary to conceptually represent these domains. In this sense, this paper aims to investigate how the use of semantic frames as a contextual information layer could help the lexical prediction task in a company's terminology base. Frame Semantics (Fillmore, 1982; 1985) is characterized as an empirical model for semantic analysis. Petruck (2001) describes the frame as a central structure that contributes to the organization of the lexicon as well as to semantically related words. As Faber (2012) points out, based on the precepts of Frame-based Terminology, if words evoke frames and can be seen as a point of access to knowledge, terms can also have this function, showing that frames help understand specialized knowledge. One way to analyze the terms and search for information regarding the frames and their contextual aspects is through FrameNet (Ruppenhofer, 2016), in which we can connect the linguistic analysis and the computational tasks. Therefore, the first step of our methodology was to analyze the used dataset, consisting of SemEval-2007 Task 10 (McCarthy, 2007) and Concepts-In-Context (CoInCO) (Kremer, 2014) data. It presented several instances, each one containing a sentence, the context, a target term, and a list of predictions of lexical substitution terms for the target term. To add the frame layer to the computational model, we selected the frames Employee_scenario and Employer_scenario from FrameNet that were related to the company's scenario. To support lexical prediction experiments, we searched for words related to the scenarios in the dataset. From the computing side, the prediction task in the dataset improved with the frames layer, providing high-quality and wider information as contextual information. Computational models, Transformers (cf. Vaswani, 2017), use the main terms from the semantic frames, improving the results from 38.7% to 41.3% in relation to recall metric, which measures the model's capacity to predict the relevant terms considering the totality of terms. From the linguistic side, the knowledge structure and its information help the user to understand the terminology and its context. The addition of the frame semantics layer resulted in the creation of a scenario that considers the context, the meaning of words, and the semantic information, among other linguistic resources to attribute adequate senses to the lexical units. In relation to the dataset analysis, we observed that the lexical substitution is a complex task regarding annotation, because there are several alternative terms for a target term in a specific context, which makes it difficult to annotate all possible alternatives. Therefore, the evaluation can be impaired, once even if the models predict a correct alternative, it is considered an incorrect answer since it was not annotated by the linguist.

References

Arefyev, Nikolay, et al. 2022. *Always keep your target in mind: Studying semantics and improving performance of neural lexical substitution*. arXiv preprint arXiv:2206,11815.

Boas, Hans. 2013. *Frame Semantics and translation*. In: A. Rojo and I. Ibarretxte- Antunano (eds.), Cognitive Linguistics and Translation. Berlin/New York: Mouton de Gruyter.

Deng, L., Liu, Y. 2018. *A Joint Introduction to Natural Language Processing and to Deep Learning*. Springer, Singapore.

Faber, P. 2012. *A cognitive linguistics view of terminology and specialized language*. Berlin, Boston: De Gruyter Mouton.

Fillmore, Charles. 1982. *Frame semantics*. In: The Linguistic Society of Korea (Eds.). Linguistics in the Morning Calm. Seoul: Hanshin. 111–37.

Fillmore, Charles. 1985. *Frames and the semantics of understanding*. Quaderni di Semantica. Vol.6, N.2. 222–254.

Kremer, Gerhard, et al. (2014). W*hat substitutes tell us-analysis of an "all-words" lexical substitution corpus*. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics.

Liu, Z., Han, X. 2018. *Deep Learning in Knowledge Graph*. In: Deng, L., Liu, Y. (eds) Deep Learning in Natural Language Processing. Springer, Singapore. In: Deng, L., Liu, Y. (eds) Deep Learning in Natural Language Processing. Springer, Singapore.

Petruck, Miriam. 2001. *Frame Semantics*. Berkeley: University of California.

Ruppenhofer, Josef; Ellsworth, Michael; Petruck, Miriam; Johnson, Christopher; Baker, Collin; Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf

McCarthy, Diana, and Roberto Navigli. *Semeval-2007 task 10: English lexical substitution task*. Proceedings of the fourth international workshop on semantic evaluations (SemEval-2007).