Application for the Poster Session in the RC21 Project Symposium

Abstract: German Support Verb Constructions in Context Embeddings [1]

Xinyao Lu [2]

In representing words with consideration of their contexts, context embedding models often show a strong ability in disambiguation. This case study attempts to explore the question, how do the embedding representations of German support verbs vary from those of the full verbs. In the first part of the poster I will display the practical skills used to search for support verb constructions from a corpus and give a quantitative analysis using the typical corpus linguistics methods. In the second part I will use different similarity measures to compare the representations and discuss about the proper interpretation.

1. Support verb constructions in German

Support verb construction (SVC, German: Funktionsverbgefuege) is a special kind of verbo-nominal collocation, in which the meaning of the verb is almost completely transferred to the noun and itself appears only as a support verb. Typical examples of SVCs in the German language are: in Kauf nehmen, (endure something uncomfortable) in Rechnung stellen, (ask for payment) and so on. It's important to notice that SVCs are more than collocations with a higher association-score, they are fixed semantic units.

2. Context embedding and disambiguation

In analysing the semantic field of the context, context word embedding assigns different vector representations to homographs. It's reasonable to assume that the support verbs in German SVCs also receive a different representation in comparison to the cases, as they are used as full verbs. (e. g. nehmen as support verb and as full verb). This study attempts to verify this assumption and, if possible, interpret the differences.

---

[1] This topic is inspired from the discussion in the seminar “Mehrwortausdrücke und Kollokationen aus computerlinguistischer Sicht” (Multiword Expressions and collocations from the perspective of computational linguistics) in the winter semester 24-25, lecturer: Besim Kabashi, Chair of Computational Corpus Linguistics

[2] Bachelor student in Computational Linguistics and German Studies at the FAU Erlangen-Nuremberg

3. Organizing the dataset

The first challenge for this study is how to carry out not only an exhaustive but also exact search for the SVCs from a corpus. I will suggest a solution combining Part-of-Speed Tagging and lexicographical resources, which could make the detecting process completely automatic and transparent. Then I will discuss, in which cases a manual close reading should still be necessary for the final decision. The dataset would thus be organized semi-automatically.

References:

[1] Grégoire Montavon, Wojciech Samek, Klaus-Robert Müller. *Methods for interpreting and understanding deep neural networks*, Digital Signal Processing, Volume 73, Pages 1-15, 2018.

[2] Stefan Th. Gries. *Frequency, Dispersion, Association and Keyness: Revising and tupleizing corpus-linguistic measures*, Amsterdam, Philadelphia: John Benjamins Publishing Company, 2024.

[3] Volker Harm. *Funktionsverbgefüge des Deutschen: Untersuchungen zu einer Kategorie zwischen Lexikon und Grammatik*, Berlin, Boston: De Gruyter, 2021.