UNION, INTERSECTION AND COMPARISON OF LINGUISTIC STRUCTURES

A.I. Stukachev, C. Lyu, T. Tsao

Sobolev Institute of Mathematics
Acad. Koptyuga avenue 4
Novosibirsk State University
Pirogova street 1
630090 Novosibirsk Russia
e-mail: aistu@math.nsc.ru

Classical notions and techniques from model theory are used in mathematical linguistics since Montague seminal works [1], with precise formal representations of language structures, both syntactical and semantical. However, in modern computational linguistics machine learning and statistical methods generate linguistic data structures (datasets) of different type. In particular, these structures do not always satisfy the requirements of formal correctness.

We choose an example of formal mathematical definition of a linguistic data structure and introduce operations of union, intersection and comparison for such structures, formalized in algebraic and model-theoretic terms. We describe a modification algorithm for correct merging of linguistic datasets that guarantee the semantic consistency. Also, the operation of intersection of linguistic datasets can be introduced in a natural way, forming a lattice.

Another natural and important question for linguistic structures is a comparison of relevance or similarity. In distributional language models, cosine similarity is a method for measuring semantic similarity of words based on their vector representations. We define a method for measuring semantic similarity for texts and linguistic datasets, by generating corresponding graph structures and calculating the graph editing distance.

References

[1] D.R. Dowty, Introduction to Montague Semantics, D. Reidel Publishing Company (1989)

[2] A.S. Burnistov, A.I. Stukachev, Generalized computable models and Montague semantics, *Studies in Computational Intelligence*, 1081 (2023), 107-124.

[3] A.S. Burnistov, A.I. Stukachev, Computable functionals of finite types in Montague semantics, *Siberian Electronic Mathematical Reports*, v. 21, N2 (2024), pp. 1460-1472..

[4] A.I. Stukachev, U.D. Penzina, Skolem functions and generalized quantifiers for negative polarity items semantics, *Lecture Notes in Networks and Systems*, 1198 (2025), 123-132.