QUALICO 2016

Trier 24-28 August

Abstracts

Universität Trier

Quantitative Analysis of Syntactic Dependency in Czech

Radek Čech [1], Ján Mačutek[2], Michaela Koĺcová[2] and Markéta Lopatková[3]
[1]University of Ostrava, Czech Republic; [2]Comenius University, Slovak Republic; [3]Charles University Prague, Czech Republic

A hierarchical structure of a sentence can be expressed by a dependency grammar formalism (Mel'čuk, 1998; Hudson, 2007). This formalism describes the structure of a sentence in a form of a tree graph; nodes of the graph represent words, while links between nodes represent syntactic relationships between words. Within the approach, there is a syntactic function assigned to each word in a sentence, e.g., predicate, subject, object, attribute; this kind of syntactic function is referred to as analytical function here (Bejček et al., 2013).

The aim of our talk is to present results of quantitative analysis of dependency characteristics of particular analytical functions. For each word in a syntactically annotated corpus (Bejček et al. 2013), a dependency frame is derived first. The dependency frame consists of all analytical functions assigned to its directly dependent words. For instance, from the sentence

Children	love	green	apples
subject	predicate	attribute	object

it is possible to derive two dependency frames. Particularly, the predicate love has the frame [subject; object] because the words Children and apples are directly dependent on the word love according to the dependency grammar formalism; analogically, the object apples has the frame [attribute]. Further, a list of unique dependency frames (with frequency characteristics) is set up for each analytical function and for each basic word form (i.e. the lemma of a word). Based on an expectation that syntactic relationships between analytical functions are ruled by some general mechanisms (cf. Köhler, 2012), we set up hypotheses as follows: (1) there is a regular distribution of dependency frames in general in a language; (2) there is a regular distribution of dependency frames of each analytical functions; differences among distributions of particular analytical functions are caused by their specific syntactic properties; differences are manifested by different distributional models or different parameters of thesame model; (3) the more frequent the analytical function is, the more dependency frames it has; (4) the more particular lemmas occur within the analytical function, the more dependency frames the analytical function has.

To test the hypotheses, a Czech syntactically annotated corpus – the Prague Dependency Treebank 3.0 is used (Bejček et al., 2013).

The results can be interpreted as a generalization of the approach presented by Čech et al. (2010) which is focused on dependency properties of predicates.

Keywords: Syntactic dependency; dependency frame; probability distribution

REFERENCES

Bejček, E., et al. (2013). *Prague Dependency Treebank 3.0*. Prague.

Čech, R., Pajas, P., & Mačutek, J. (2010). Full Valency. Verb Valency without Distinguishing Complements and Adjuncts. *Journal of Quantitative Linguistics*, *17*(4), 291–302.

Hudson, R. (2007). *Language Networks. The New Word Grammar*. Oxford: Oxford University

Press.
Köhler, R. (2012). *Quantitative syntax analysis* [Quantitative Linguistics (QL), 65]. Berlin / New York: de Gruyter.
Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. Albany, NY: Sunny Press.