

# Syntax-Semantics Interface: A Plea for a Deep Dependency Sentence Structure

**Eva Hajičová**

Charles University

Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Prague, Czech Republic

hajicova@ufal.mff.cuni.cz

In collaboration with Václava Kettnerová, Veronika Kolářová, Markéta Lopatková, Jarmila Panevová, and Dan Zeman (and with technical support of Jiří Mírovský)

The aim of the contribution is to bring arguments for a description for natural language that (i) includes a representation (i) of a deep (underlying) sentence structure and (ii) is based on the relation of dependency. Our argumentation rests on linguistic considerations and stems from the Praguian linguistic background, both with respect to the Praguian structuralist tradition as well as to the formal framework of Functional Generative Description and to the experience with building the Prague Dependency Treebank. The arguments, of course, are not novel but we will try to gather and report on our experience when working with deep syntactic dependency relations in the description of language; the basic material will be Czech but multilingual comparative aspects will be taken into account as well.

Speaking about a “deep” sentence structure, a natural question to ask is how “deep” this linguistic structure is to be. Relevant in this respect is the differentiation between ontological content and linguistic meaning. Two relations will be discussed in some detail and illustrated on examples from Czech and English, namely the relation of synonymy and that of ambiguity (homonymy). The relation of synonymy will be specified as an identity of meaning with respect to truth conditions and it will be demonstrated how this criterion may help to test sentences and constructions for synonymy. The relation of ambiguity will be exemplified by two specific groups of examples, one concerning surface deletions and the necessity to reconstruct them in the deep structure, and the other group involving the notion of deep order of sentence elements with examples related to the phenomenon of information structure.

The necessity to distinguish surface and deep structure has led to several proposals of a multi-level description of language, both in the domain of theoretical linguistics and in the domain of annotation schemes of language corpora, such as LFG or CCG. We will describe in a nutshell the Prague Dependency Treebank, focusing on the deep (so-called tectogrammatical) level of annotation.

After some observations on the history of the dependency-based syntactic relations, attention will be focused on two basic topics, namely the issue of headedness and the notion of valency. We will outline an approach to the distinction between arguments and adjuncts and their semantic optionality/obligatoriness based on two operational criteria and we will demonstrate on the example of several Czech valency dictionaries how a dependency-based description brings together grammar and lexicon.

Among the many challenges that still await a deeper analysis, two will be briefly characterized, namely the phenomenon of projectivity and the representation of coordination.

To summarize, we argue that both attributes of our approach, namely “deep” and “dependency-based” are important for a theoretical description of language if this description is supposed to help to reflect the relation between form and meaning, that is, when it is supposed to serve as a basis for language understanding. Despite undisputable recent progress in NLP which relies more on computational methods than linguistic representations or features, we believe that for true understanding, having an adequate theory is worth the effort.

*This work has been supported by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071) and by the project GA17-07313S of the Grant Agency of the Czech Republic.*