

# What we are talking about and what we are saying about it

Eva Hajičová

Charles University  
Institute of Formal and Applied Linguistics  
Malostranské nám. 25, 118 00 Prague, Czech Republic  
hajicova@ufal.mff.cuni.cz

**Abstract.** In view of the relationships between theoretical, computational and corpus linguistics, their mutual contributions are discussed and illustrated on the issue of the aspect of language related to the information structure of the sentence, distinguishing "what we are talking about" and "what we are saying about it".

## 1 Introduction

The name of the research domain of Computational Linguistics seems to be self-explanatory; however, there has always been a dispute what exactly 'computational' means (especially from the point of view of the relation between its theoretical and applied aspects and from the point of view of its supposedly narrowing scope due to the prevalent use of statistical methods). In addition, with the expansion of the use of computers for linguistic studies based on very large empirical language material, and, consequently, with the appearance of an allegedly new domain, corpus linguistics, a question has emerged what is the position of corpus linguistics with regard to computational linguistics.

After a summary of some of the issues related to the problem of 'how many linguistics there are' (Sect. 2), we briefly sketch in which respects the different 'linguistics' can mutually contribute to each other (Sect. 3). The main objective of our paper is to illustrate on an example of a linguistically based multi-layered annotation scenario (Sect. 4) and of a selected linguistic phenomenon, namely the information structure of the sentence (Sect. 5.1), how linguistic theory can contribute to a build-up of an integrated scenario of corpus annotation (Sect. 5.2) and, in the other direction, how a consistent application of such a scenario on a large corpus of continuous texts can provide a useful feedback for the theory (Sect. 5.3). In Section 6, some conclusions will be drawn from the personal experience with working with the given theory and scenario.

## 2 How many linguistics?

If the terms *computational linguistics* and *corpus linguistics* are understood rather broadly, as covering those domains of linguistics that are based on the