A method to obtain detailed constructional data using ontology of argument NPs

Abstract

This study proposes a method to describe the structures of linguistic constructions by analyzing the ontological specification of their argument NPs. What motivates this attempt is a doubt concerning the validity of implicit assumption that it is enough to stipulate part-of-speech plus thematic-role labels, such as NOUN+AGENT, to arguments of verbs in describing constructions (cf. Goldberg 1995). A construction is considered as one coherent concept having features of a ‘Gestalt’. Consequently, components of a construction are not necessarily rigidly modularized, but rather are inseparably attached to the whole structure. It is highly possible, therefore, that arguments of one verb are not fully compatible with arguments of other verbs even if they seemingly share identical thematic functions. Thus the schemas of argument NPs of AGENT-touch-OBJECT construction and that of AGENT-hit-OBJECT, for instance, should be given more detailed analysis and distinguished at some deeper level.

With the above problem as the background, we experientially conducted the following procedure with a relatively small-sized corpus and propose the method as a promising approach to the problem. First, instances of certain grammatical constructions are collected and argument NPs are identified using a dependency parser. Next, hierarchical ontology data of the NPs are obtained using the WordNet database (Fellbaum 1998). Finally a mathematical procedure of Formal Concept Analysis (Ganter et al. 2005) is applied to each set of NPs occupying the same slot in the construction. The last process produces a tree-like structure of concepts that virtually represents the range of possible semantic attributes of the argument.

The significance of the present study resides in the following respects. First, it not only makes possible the description of linguistic constructions of a much higher granularity, but it also helps verify and eventually support the preceding, rather coarse, description of linguistic constructions. Second, it shows a way to extract constructional data semi-automatically from a given corpus. There is an important limitation to this method: it requires a corpus already annotated with semantic data for avoiding the possible problem regarding polysemy of words. However, there are a few ongoing projects offering such a corpus. We use one such corpus SemCor (Shi & Mihalcea 2005) as a suitable testing material.

References


