Evidence for Concept Types from a Semantic-Pragmatic Copus Study

CHRISTIAN HORN & NICOLAS KIMM
(Heinrich Heine University Düsseldorf)

In this paper we present the results of a corpus-linguistic analysis of Löbner's (2011) semantic distinction between four different conceptual noun types. Löbner argues that conceptual types of nouns and NPs basically differ with respect to two inherent referential properties, i.e. relationality [±R] and uniqueness [±U]. Whereas [±R] refers to the distinction between one- and n-place predicates, [±U] relates to whether the conceptual type is constructed as referring to only one referent in a given context. The cross-classification of the values of these properties leads to the four conceptual types: sortal concepts ('SCs', [–R][–U]; *tree*, *stone*), relational concepts ('RCs', [+R][–U]; *sister*, *window*), individual concepts ('ICs', [–R][+U]; *sun*, *pope*, proper names), and functional concepts ('FCs', [+R][+U]; *roof*, *head*). Grammatically, the four concept types are seen as congruent with certain modes of determination: In accordance with their [+R] property, FCs and RCs are seen as predisposed for possessive use. Due to their [+U] property, ICs and FCs are expected to have a predisposition for definite determination. Löbner expects for each concept type congruent uses to be more frequent than incongruent uses.

We investigate these expectations corpus-linguistically by testing whether the grammatical use of the nouns statistically reflects the respective concept type. The first step is to assess the concept type of the nouns in our text collection on the basis of a systematic annotation procedure. We then analyze concept types from two perspectives: (a) with their modes of determination (making up 'linguistic congruency') and (b) in their pragmatic use (making up 'pragmatic congruency'). In accordance with Grice's maxim of quantity (cf. Grice 1975), we expect that a possessor argument can be dropped if the speaker assumes that the hearer can retrieve it from the context of utterance, and one such type are associative anaphors. Such uses would have impact on the frequency of possessive uses of [±R] uses. For the study we present, we focus on nominal associative anaphors (NAA) defined by the following conditions (cf. Kimm & Horn 2011):

(i) The referent of an anaphoric NP is determined by associating it with a referent previously introduced in the discourse (often called the "anchor" or "antecedent").
(ii) The anchor is given by an NP.
(iii) The reference to the anchor must be successful.
(iv) The anaphoric NP may be used definite or indefinite.
(v) Both the anaphoric NP and the anchor do not co-refer.

An example for an NAA is given in (1).

(1) Mary passes a house. The roof is on fire.

The FC *roof* is used linguistically incongruent since the required possessor *house* is not given explicitly in the NP. Instead, it is given in the first sentence as the antecedent

of the associative anaphor *the roof*. We consider this a case of pragmatic congruency. Subsuming the data for linguistic and pragmatic congruency, our hypothesis is that the results for overall congruency reflect the distinction between the four different concept types.

Methodologically, we analyze a German corpus since German provides a rich inventory of explicit grammatical marking. We present (i) annotation guidelines for both the classification of concept types and NAAs, and (ii) results of studies carried out on both aspects. The guidelines for the annotation of concept types comprise three parts. In the first part, we analyse the nouns in their respective context of utterance and identify the given meaning variant. Here we also exclude idiomatic uses. The second part addresses the separate semantic identification of the referential properties [±R] and [±U] of the meaning variants so that for each of them its concept type can be derived. In the third part for each noun token its respective grammatical use is annotated. The annotated NPs allow for a statistical analysis with respect to linguistic congruency. Consequently, these NPs form the basis for the second part of our empirical analysis, the annotation of NAAs. The annotation guidelines for this task are derived from the conditions given above: Each definite or indefinite NP is classified whether its reference is determined by associating it with a referent that is introduced in the discourse by another NP, the anchor. If the reference to the anchor is successful and both NPs do not co-refer, the first NP is classified as an NAA.

We will show that the data for linguistic congruency reflect the semantic distinction of concept types with statistical significance. However, they show that only 30% of the [+R] concepts in the corpus occur in possessive constructions resulting in a low amount of linguistic congruency with respect to the [±R] distinction. The integration of pragmatic congruency enhances the amount of overall congruency and provides further evidence for the distinction of concept types.

Grice, Paul (1975): Logic and Conversation. In Peter Cole & Jerry L. Morgan (eds.), Syntax and Semantics, Volume 3. 41–58.

Kimm, Nicolas & Christian Horn (2011): Nominal Associative Anaphors – A Text-Based Analysis at the Semantics-Pragmatics Interface. In: Iris Hendrickx, Sobha Lalitha Devi, António Branco, and Ruslan Mitkov (Eds.): Anaphora and Reference Resolution. Lecture Notes in Artificial Intelligence 7099. Springer. Berlin. 108–118.

Löbner, Sebastian (2011): Conceptual Types and Determination. In: Journal of Semantics.