Hierarchical Semantic Analysis of Japanese Sentences of the form "*Y*-ga *X*-kara (*Z*-ni) nigeru": Combining a manual corpus analysis and psycholinguistic experiments

Keiko Nakamoto (Bunkyo University) and Kow Kuroda (NICT, Japan)

The first purpose of this paper was to add new evidence for the validity of conceptual analysis procedure called Hierarchical Semantic Frame Network Analysis (HFNA) proposed by Kuroda et al. (2005). The second one was to show the effectiveness of the new approach to conceptual analysis we proposed on the basis of comparison among sets of semantic frames obtained from HFNA. We targeted a set of situations described by sentences of the form "*Y*-ga *X*-kara (*Z*-ni) nigeru," meaning from "*Y* runs away from *X*" to "*Y* gets lost from *X*".

For the first purpose, a corpus analysis and psycholinguistic experiments were conducted in the same way as in Nakamoto et al. (2005). In the corpus analysis, all sentences including *nigeru* (i.e., inflected forms like *nigete*, *nigeta*) were collected from a Japanese newspaper corpus (JENAD: Utiyama and Isahara 2003). This set of roughly 200 sentences were carefully analyzed to identify the set of concrete situations against which they are understood. The situations identified this way were equated with "semantic frames" at a finer-granularity (Fillmore, et al. 2003). Roughly speaking, we defined semantic frames as "schemas of situations" (cf. Gentner 2005). They can be seen as a restricted form of Idealized Cognitive Models (Lakoff 1987), but they were all represented as combinations of semantic features such as [+animate(*X*)] ("*X* is animate") or [+intentional(*X*)] ("*X* did something intentionally), whose value can be a continuous ranging from "0" (Very false) to "1" (Very true). This was done to ensure the compatibility with one of the two psychological experiments we did.

As a result of corpus analysis, 17 frames were identified and 31 semantic features necessary for distinguishing among these frames identified. To test the validity of the corpus analysis, two psycholinguistic experiments (Card Sorting Task and Feautre Rating Task) were conducted. For both experiments, materials were created so that each frame has three unambiguous instances. In the Card Sorting Task, participants were required to sort the material sentences freely so that the sentences in similar senses were grouped together. This task was conducted to test if the intuitive grouping/sorting done by unbiased non-linguists was close enough to the manual analysis we did. In the Feature Rating Task, a different set of participants were asked to "rate" the appropriateness of a set of characteristics (like "*X* intended to capture *Y*") against each of the sentences we used in Card Sorting. Rating scale is from 0 (Very false) to 1 (Very true). This was done in order to ensure that the semantic features/properties that are assumed to characterize each of the semantic frames/situations were adequate enough to distinguish among them. Multivariate analyses on these behavioral data revealed that the laypersons' semantic intuition was roughly equivalent to the structure of HFNA we obtained from the corpus analysis, thereby supporting the validity of the approach proposed in the framework of FOCAL.

For the second purpose, we made pairs of examples of "*X-ga Y-kara nigeta*" and "*Y-ga X-wo osotta* (roughly meaning '*X* attacked *Y*'), and judged the acceptability of these examples. Since some sort of a "threat" from *Y* is a prerequisite for *X*'s escape, it was expected that the comparison would reveal the relationship between the situations denoted by these two verbs, *osou* and *nigeru*. The results of the comparison indicated that the direct correspondence between them was very limited. Many sentences of the *X-ga Y-kara nigeta* cannot be directly converted into *Y-ga X-wo osotta*: there are subtle selectional restrictions such as "Harm-causer of the *osou* situation must be explicitly dangerous."

In the final section, we argued that the effectiveness of corpus analyses and psycholinguistic experiments for the research of semantic aspects of language. We claimed that such empirical approaches can overcome limitations of the rather "traditional" approaches in Cognitive Linguistics based on (uncontrolled, usually theory-biased) intuitive analysis of artificially made examples.