

Similarity-based classification

Carla Umbach & Helmar Gust

Classification of objects can be expressed linguistically in basically two ways, either by ascribing a (nominal or verbal) predicate to the object, or by comparing the object to some other entity that is already known to the addressee. When ascribing a predicate to an object, the object is classified by the help of the predicate. In (1a), for example, we are told that Anna's height is 1,80m. When comparing the object to a second one, it is classified by similarity in the relevant respect. Thus, in (1b) we only learn that Anna and Marie are similar with respect to height and we have to infer Anna's height from what we know about Marie. Likewise, in (2a) the property of the student in question is explicitly mentioned while (2b) it has to be inferred from what we know about Marie. Finally, in (3a) we are told that Anna went to Vienna via Prague, while in (3b) we only learn that there is some respect with respect to which Anna's trip to Vienna was similar to Marie's trip.

- | | | | |
|-----|----|--|--|
| (1) | a. | Anna ist 1,80m groß. | 'Anna is 1,80m' |
| | b. | Anna ist so groß wie Marie. | 'Anna is as tall as Marie' |
| (2) | a. | Anna ist eine Studentin mit beachtlichen Mathekenntnissen. | 'Anna is a student with notable Math skills.' |
| | b. | Anna ist eine Studentin wie Marie. | 'Anna is a student like Marie.' |
| (3) | a. | Anna ist über Prag nach Wien gefahren. | 'Anna went to Vienna via Prague.' |
| | b. | Anna ist so nach Wien gefahren wie Marie. | 'Anna took the same route to Vienna as Marie.' |

While comparative constructions involving gradable adjectives have been discussed in detail (e.g., Bierwisch 1986, Kennedy 1999), there are few approaches going beyond the adjectival domain. Moreover, although similarity-based classification is regarded as a basic cognitive ability of human agents and has been studied at length in Cognitive Psychology as well as Artificial Intelligence (cf. Tversky 1977, Hahn & Chater 1998, Gärdenfors 2000, Aamodt & Plaza 1994), the question of how similarity is exploited in natural language is rarely discussed in linguistics.

This paper aims at a unified semantic analysis of comparative constructions including gradable as well as non-gradable cases. We will focus on the semantics of German *wie*-phrases and provide a transparent interpretation starting from the hypothesis that the basic function of *wie* consists in turning an individual-denoting expression (like a proper name) into a property. The formal framework will be adopted from AI approaches to similarity and classification (Saquer & Deogun 2001), which facilitate the core problem of how to spell out the role of the respect (or dimension) of similarity.

The general idea of this paper consists in providing evidence that natural languages possess a mechanism to construct ad hoc categories exploiting the utterance situation – e.g. *wie* phrases in German – which can be used in parallel to regular linguistic predicates. This supports the well-known claim that rule-based and similarity-based mechanisms co-exist in human cognition.

Bierwisch, M. (1989) The semantics of gradation. In M. Bierwisch & E. Lang (eds) *Dimensional adjectives*. Springer, Berlin, 71–262.

Kennedy, Chr. (1999) *Projecting the adjective: The syntax and semantics of gradability and comparison*. Garland Press, New York.

Hahn, U. & Chater, N. (1998). Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable? *Cognition*, 65:197–230.

Tversky, Amos (1977) Features of Similarity. *Psychological Review*, 84-4, 327-352.

Gärdenfors, Peter (2000): *Conceptual Spaces: The Geometry of Thought*

Aamodt, A., Plaza E. (1994): *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*. AICom - Artificial Intelligence Communications, IOS Press, Vol. 7: 1, pp. 39-59.

Saquer, J., Deogun, J., S. (2001), *Concept Approximations Based on Rough Sets and Similarity Measures*. *Int. J. Appl. Math. Comput. Sci.* Vol. 11, No 3, pp. 655-674