Interpreting similarity using attribute spaces and generalized measure functions

CARLA UMBACH[1] & HELMAR GUST[2]
([1]Zentrum für Allgemeine Sprachwissenschaft (ZAS) Berlin,
[2]Institut für Kognitionswissenschaft, Universität Osnabrück)

Similarity is well-known to be a core concept of human cognition, e.g., in categorization and learning. Therefore, expressions of similarity in natural language are of special interest: How to account for their meaning including the results on similarity in Cognitive Science and Artificial Intelligence without abandoning truth-conditional semantics?

In this paper we will suggest a way to connect truth-conditional semantics to conceptual structures by generalizing the notion of measure functions known in degree semantics (Kennedy 1999) from the one-dimensional to the many-dimensional case. Generalized measure functions map individuals to points in multi-dimensional attribute spaces spanned by contextually relevant concepts. Similarity is then spelt out as indistinguishability with respect to a given set of attributes.

The approach suggested in this paper offers a path to include conceptual information into a truth-conditional account of semantics arguing, at the same time, that this path has already been established in degree semantics where degrees are included into a truth-conditional account of gradable adjectives.

Expressions of similarity are, e.g., German *ähnlich*/English *similar* and German *so*/ English *such*, (*like) this*. We start from the deictic use of German *so* occurring, e.g., in nominal and adjectival phrases, cf. (1).

(1) a. (speaker pointing at a person):
So groß ist Anna. 'Anna is this tall.'
b. (speaker pointing at a car):
So ein Auto hat Anna. 'Anna has such a car / a car like this.'

We assume that the target of the pointing gesture is the individual the speaker points at (instead of, e.g., a property or a kind).[1] The demonstrative *so* is interpreted as a three-place predicate SIM(x, y, F), relating two individuals *x* and *y* and a set of relevant features of comparison *F*, cf. (2).

(2) a. [[so groß]] = $\lambda x.\ \text{SIM}\ (x, x_{target}, \{height\})$
b. [[so ein Auto]] = $\lambda Q.\ \exists x.\ \text{SIM}\ (x, x_{target}, F)\ \&\ car(x)\ \&\ Q(x)$

In the adjectival case, there is a single feature of comparison given by the adjective's meaning – height in (2a).[2] In the nominal case, multiple features of comparison must be taken into account which are only implicitly given by the concept represented by the noun and are due to contextual restrictions. Moreover, while adjectival dimensions usually relate to ratio scales, nominal dimensions may relate to scales of various types

[1] See Umbach & Gust (to appear) for an in depth discussion of the target of the pointing gesture is and the role of kinds in the interpretation of *so/such* as compared to Carlson (1980).
[2] We consider only dimensional adjectives like *groß/tall*. Dimensional adjectives involve one dimension within a single comparison even if they are multi-dimensional in the sense of Sassoon (2011).

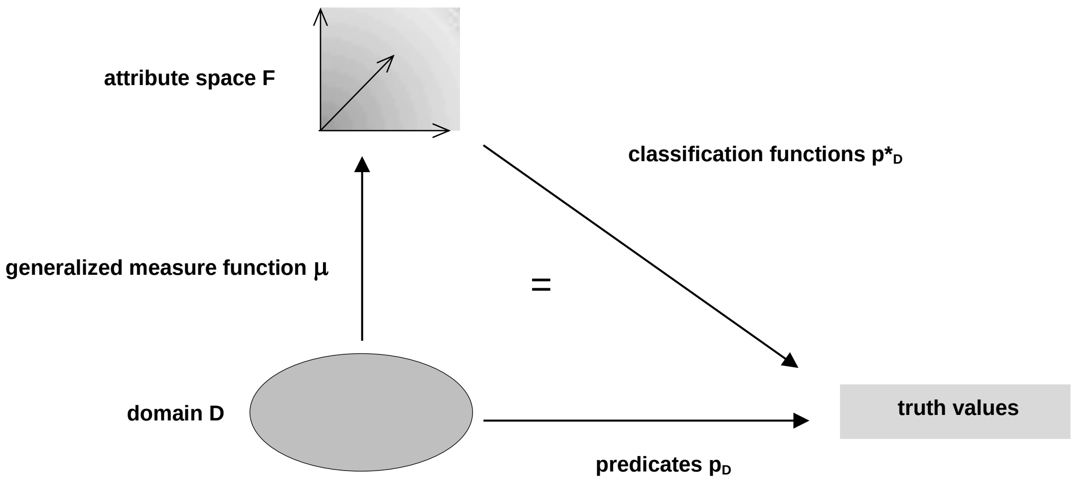

Figure 1: The combination of the domain of individuals D and an attribute space F

– ratio, ordinal, or even nominal. Setting these differences aside, the notion of adjectival measure functions can straightforwardly be generalized to the nominal case: While adjectival measure functions map individuals to degrees in a single dimension, cf. (3a), generalized measure functions map individuals point-wise into multi-dimensional attribute spaces, cf. (3b).

(3) a. $\mu_{HEIGHT}: U \rightarrow \Re$
b. $\mu_{CAR}: U \rightarrow \langle DRIVE, HP, \ldots \rangle$, where $\mu_{CAR}(x) = \langle \mu_{DRIVE}(x), \mu_{HP}(x), \ldots \rangle$ and $\mu_{DRIVE}(x) \in \{DIESEL, GAS, \ldots\}$, $\mu_{HP} \in \Re$, ...

Multi-dimensional attribute spaces are given by a set *F* of dimensions associated with a set *C*(*F*) of classification functions defined on their points. These classification functions approximate natural language predicates on a conceptual level yielding corresponding truth values (modulo fuzzy membership). For example, a classification function *high-powered** associated with the horsepower dimension in the example above is subject to the constraint in (4). The role of classification functions is two-fold. First, while generalized measure functions take individuals to points in attribute spaces, classification functions take these points back to regular predicates such that the diagram in fig. 1 commutes. From this point of view, they warrant the integration of attribute spaces into truth-conditional semantics.

(4) high-powered*(μ_{HP} (x)) iff high-powered (x)

Secondly, classification functions determine the level of granularity: Similarity is defined such that two individuals are similar with respect to a set of relevant features iff the classification functions yield the same result when applied to corresponding points in the attribute space, cf. (5) (where *C*(*F*) is the set of classification functions associated with the dimensions in *F*).

(5) sim(x, y, F) iff $\forall p^* \in C(F)$: $p^*(\mu_F(x)) = p^*(\mu_F(y))$

The similarity relation in (5) for a fixed *F* corresponds to the indistinguishability notion of similarity in rough set theory (Pawlak 1998), which is an equivalence relation. This implementation is adequate for the interpretation of *so/such* (although symmetry may be discussed). The interpretation of the adjectives *similar*/*ähnlich* will require a slightly different relation (cf. Tversky's 1977 contrast model) which can, however, easily be defined in multi-dimensional attribute spaces as suggested above.

Umbach, C., Gust, H. (to appear) Similarity demonstratives. *Lingua.*

Carlson, G. N. (1980) Reference to Kinds in English. Garland, New York & London.

Gärdenfors , P. (2000) Conceptual Spaces. MIT Press.

Kennedy, C. (1999) Projecting the adjective. Garland Press, New York.

Pawlak, Z. (1998) Granularity of knowledge, indiscernibility and rough sets. *Proceedings of the IEEE International Conference on Fuzzy Systems*, 106-110.

Sassoon, G. (2011) Sassoon, G. (2011) Adjectival vs. nominal categorization processes. *Belgian Journal of Linguistics* 25, 104-147.

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