

DFG project 'Similarity I' (UM100/1, 2012–2015) and 'Similarity II' (UM100/3, 2017–2020)

Carla Umbach

Zentrum für Allgemeine Sprachwissenschaft, Berlin

The linguistic expression of similarity

Part I: Project-related publications

a) In journals and reviewed collected volumes

Umbach, Carla, & Helmar Gust (2014) Similarity Demonstratives. *Lingua* 149, 74-93

Umbach, Carla (2014) Expressing similarity: On some differences between adjectives and demonstratives. *Proceedings of IATL 2013*, MIT Working Papers in Linguistics.

König, Ekkehard & Carla Umbach (2018) Demonstratives of Manner, of Quality and of Degree: A Neglected Subclass. In M. Coniglio, A. Murphy, E. Schlachter & T. Veenstra (eds.). *Atypical demonstratives: syntax, semantics and pragmatics*. Berlin, de Gruyter Mouton.

Umbach, Carla & Helmar Gust (2021) Grading similarity. In Löbner, S., Gamerschlag, T., Kalenscher, T., Schrenk, M., Zeevat, H. (Eds.) *Concepts, Frames and Cascades in Semantics, Cognition and Ontology*. Springer.

Umbach, Carla, Stefan Hinterwimmer & Helmar Gust (2021) German 'wie'-complements: Manners, methods and events in progress. *Natural Language and Linguistic Theory*.
<https://doi.org/10.1007/s11049-021-09508-z>

Gust, Helmar & Carla Umbach (2021) A qualitative similarity framework for the interpretation of natural language similarity expressions. In Lucas Bechberger, Kai-Uwe Kühnberger and Mingya Liu (eds). *Concepts in Action - Representation, Learning and Application*. Language, Cognition, and Mind. Springer.

Umbach, Carla & Stephanie Solt (to appear) Comparison via 'eher'. *Journal of Semantics*.

Umbach, Carla (to appear) Ways of classification: German *Art* and *Typ*. In Brems, Lieselotte; Davidse, Kristin; Mihatsch, Wiltrud; Hennecke, Inga; Kisiel, Anna; Kolyaseva, Alena (Eds.) *Type Noun Constructions in Slavic, Germanic and Romance Languages - Semantics and pragmatics on the move*. To appear in: De Gruyter Series Trends in Linguistics. Studies and Monographs.

b) Other publications

Umbach, Carla (2019) *Dasselbe* oder *das gleiche?* - warum wir dann doch dieselbe Pizza bestellen können wie der Gast nebenan. ZAS Jahrbuch 2018.

Umbach, Carla & Umut Özge (2019) Scalar and non-scalar comparison across categories: The case of Turkish equatives. Extended abstract, TbiLLC 2019 Symposium on Language, Logic and Computation, Batumi.

(a full list of publications is available at www.carla-umbach.de/publications.html)

DFG project 'Similarity I' (UM100/1, 2012–2015) and 'Similarity II' (UM100/3, 2017–2020)

The linguistic expression of similarity

Part III: Summary

In Cognitive Science, similarity has long been recognized as fundamental in explaining cognitive skills like perception, classification and learning. This leads to the expectation that similarity plays prominent role in language, too. Surprisingly, however, the notion of similarity is near to ignored in linguistics and semantics, which suggests that something might have been overlooked.

The similarity project started from the German expressions *so* ('such', 'like this') and *wie* ('like', 'how'). A semantic analysis is provided such that *wie* denotes the relation of similarity and that *so* is a demonstrative expressing similarity (instead of identity) between the thing pointed to and the referent of the linguistic phrase. The relation of similarity was spelt out in a feature-based framework that can be seen as an extension of the well-established degree semantic framework.

This account represents a major advance in making the concept of similarity available in semantics for the first time. It allows for a uniform analysis for the various uses of the demonstrative *so* and for the uses of *wie* in equative constructions as well as in expressing manner. And it turned out that these expressions do exactly what one would expect from a cognitive science point of view: sort things into kinds – *Anna hat so ein Auto wie Berta* ('Anna has a car like Berta's') means that Anna's car is of the same kind though not identical to Berta's car.

There are two results that were unforeseen in the application phase. First, there is a link from the notion of similarity to findings in the area of genericity and of manner modification (see section 1). It strengthens the idea of *so* / *wie* as a means of (ad-hoc) kind formation and opens a window into constraints on concept formation as discussed in cognitive science. Secondly, while the generalized similarity analysis of equative comparison sketched in the proposal is straightforward for German, there are data from Turkish showing that such a generalization would not be adequate. Turkish is evidence that there are in fact two different strategies of equatives, a degree-based and a similarity-based one, independent of the grammatical category of the parameter. This led to an important theoretical insight: the established degree-based approaches and the similarity approach should not be seen as competing semantic theories. Instead, a successful analysis of equative comparison has to provide access to degrees as well as similarity, which is allowed in the framework developed in the project.

The linguistic expression of similarity

Part II: Final report

Introduction

The project *Similarity I + II* was aimed at an in-depth investigation of linguistic expressions of similarity. In Cognitive Science, similarity has long been recognized as fundamental in explaining cognitive skills like perception, classification and learning. Following Tversky, "similarity [...] serves as an organizing principle by which individuals classify objects, form concepts, and make generalizations." (Tversky 1977, p. 327). Quine likewise argues that "... there is nothing more basic to thought and language than our sense of similarity; our sorting of things into kinds." (Quine 1969, p. 116). Surprisingly, however, in linguistics similarity has rarely been considered as a semantic notion.

The similarity project started from the German expressions *so* ('such', 'like this') and *wie* ('like', 'how'). The former is a demonstrative and is directly referential (in the sense of Kaplan 1989), but unlike regular demonstrative (e.g., *dies/this*) it occurs a modifier in nominal, verbal and adjectival phrases posing the problem how to reconcile its demonstrative characteristics and its modifying function. In the project, an analysis was developed according to which *so* is a *similarity demonstrative* expressing similarity (instead of identity) between the target the speaker points to and the referent of the phrase. For example, if the speaker points to a car uttering the sentence *So ein Auto hat Anna* ('Anna has a car like this') she means to say that Anna's car is similar, though not identical, to the car pointed at.

The expression *wie* denotes similarity, too, but in contrast to *so* there is no deictic component. *Wie* occurs in equative comparison constructions, thereby suggesting a similarity account of equatives – *Annas Auto ist (so) wie Bertas Auto*. ('Anna's car is like Berta's car) means that Anna's car is similar to Berta's. This view entails that the standard marker *wie* is the core semantic component of equatives. Beyond comparison, *wie* is used as a manner question word, as in *Wie hat Anna das Fahrrad repariert?* ('How did Anna repair the bike?'). Ruling out ambiguity the challenge lies in explaining the link between similarity and the notion of manner.

The core results of the similarity projects include

- 1) the interpretation of *so* as a similarity demonstrative;
- 2) a generalized account of German equative comparison based on the standard marker *wie* expressing similarity;
- 3) reconstruction of the notion of manner by similarity in subordinated clauses;
- 4) a semantic framework based on generalized measure functions spelling out the notion of similarity via (variably grained) identity of features;
- 5) studies on the commonalities and differences of expressions of similarities, in particular *so* vs. *ähnlich* vs. *gleich*, and *derselbe* vs. *der gleiche*, and also classification as expressed by the nouns *Art* and *Typ*;
- 6) in cooperation with the project *Degree Attenuators* (Solt): a semantic analysis of the adverb *ehler*.

The results are detailed in the sections below.

Compared to the application proposals of Similarity I and II, each part of the project includes a result not foreseen in advance. First, the use of *so* as a similarity demonstrative turned out to be restricted to features of comparison depending on the (concepts expressed by the) modified nouns or verbs. This finding points to results in the area of genericity and of manner modification opening a window into cognitive mechanisms of concept formation. The second surprise was the realization that in Turkish two different strategies of equative comparison indicated by two different standard markers are used in parallel. This is evidence that degree-based approaches and the similarity approach should not be seen as competing theories. Instead, the formal framework should provide access to degrees as well as similarity, and the semantic interpretation should be compositionally derived from the meaning of the standard marker.

1. Similarity demonstratives

(In this section, the results of the first part of the similarity project will be summarized. A more detailed report is given in see sect. 1a of the follow-up application for the second part).

German *so* is a genuine demonstrative, that is, directly referential, but at the same time serves as a cross-categorical modifier combining with verbal, nominal and also adjectival expressions, thereby posing the problem how to reconcile its demonstrative characteristics and its modifying function. The analysis developed in the project is such that *so* is a *similarity demonstrative*: When using *so* deictically the speaker points to individuals or events, as is the case with standard demonstratives like *dies* ('this'). Unlike the latter, however, the relation between the target of the pointing gesture and the referent of the demonstrative phrase is not identity (as is in built in Kaplanian semantics) but instead similarity. This interpretation applies across categories (nominal, verbal, adjectival), and to deictic as well as anaphoric uses. Consider the examples in (1). In (a), Anna's car is said to be similar to the car pointed at. In (b), Anna's manner of dancing is said to be similar to that in the dancing event pointed at. Finally, in (c) Anna's height is said to be similar to the height of the person the speaker is pointing at.

- (1) a. (speaker pointing to a car in the street):
So ein Auto hat Anna auch. 'Anna has such a car / a car like this, too.'
- b. (speaker pointing to someone dancing):
So tanzte Anna gestern auch. 'Yesterday, Anna danced like this, too.'
- c. (speaker pointing to a person):
So groß ist Anna auch. 'Anna is this tall, too.'

In the nominal and the verbal case, *so* phrases can be seen as creating similarity classes corresponding to ad-hoc generated kinds (cmp. Barsalou 1983). This is evident when combined with the noun *Art* ('kind') – *so ein Auto* is equivalent to *diese Art von Auto* ('this kind of car'), and *so tanzen* is equivalent to *diese Art zu tanzen* ('this kind of dancing'). Adjectival cases, however, should not be considered as kinds: *so groß* is equivalent to *diese Größe* ('this size') but not to *diese Art von groß sein* ('this kind of being tall'). Thus the similarity analysis of *so* agrees with the analyses of English (anaphoric) *such* and Polish *tak* put forward by Carlson 1980 and also by Anderson & Morzycki 2015 in assuming that – in the nominal and the verbal case – there is reference to kinds. However, different from these analyses it is argued kinds denoted by *so*-phrases are not previously established but instead ad-hoc generated with the help of the similarity relation contributed by the lexical meaning of *so*.

The notion of similarity has been criticized in the literature as being trivial without specifying features of comparison (Goodman 1972). This entails that similarity has to be implemented as a three-place relation: $SIM(x,y,F)$ where x and y are entities to be compared (of the appropriate type) and F

represents a set of features of comparison (contributed by the context). It turned out in the course of the project that the features of comparison are severely restricted depending on the lexical meaning of the parameter. This is demonstrated in the examples in (2), in which the demonstrative *so* in the second sentence is supposed to pick up a property of an antecedent in the first sentence. In (2a), being a diesel as well as being a Japanese car are easily picked up, leading to the interpretation that Berta has a Japanese car and a diesel, respectively. In (2b), however, being new does not qualify as a feature of comparison – the second sentence cannot be understood as saying that Berta has a new car.

- (2) a. Anna hat einen Diesel / ein japanisches Auto. Berta hat auch so ein Auto (nämlich einen Diesel/ ein japanisches Auto).
 b. Anna hat ein neues Auto. Berta hat auch so ein Auto (*nämlich ein neues Auto).
 'Anna has a diesel / a Japanese car / a new car. Berta has such a car, too (namely a diesel / a Japanese car / a new car).'

Analogous effects have been observed in the area of genericity (Carlson 2010, Greenberg 2003) and concept formation (Prasada & Dillingham 2006), and also for adverbial event modifiers (Maienborn & Schäfer 2011). A series of experimental studies was conducted collecting naturalness ratings for nominal and verbal stimuli. Design and evaluation were supported by Prof. Britta Stolterfoht (Tübingen). This finding has been confirmed in a series of experimental studies and establishes an important link between linguistic ways of kind formation and insights about concept formation in cognitive psychology. The experiments have been reported at international conferences, a summary of the results is included in König & Umbach (2018) and a full paper with Prof. Stolterfoht is in preparation.

Beyond German *so*, there are in many other languages demonstrative expressions occurring as modifiers of nominal, verbal and/or adjectival phrases. In König & Umbach (2018) these expressions are called '*demonstratives of manner, quality and degree*' (in short MQD demonstratives) and are examined from a typological point of view. In the similarity project the cross-linguistic perspective focused, in particular, on Turkish (in cooperation with Prof. Umut Özge, METU, Ankara) and also on Kambaata (in cooperation with Dr. Yvonne Treis, CNRS-LLACAN).

2. A generalized account of German equative constructions

In degree semantics, equatives are considered nearly exclusively from the perspective of comparatives (*Anna is taller than Berta*), and are assumed to differ from comparatives only in expressing a non-strict (\geq) rather than a strict order. However, while comparatives are (mostly) scalar, equatives can be scalar as well as non-scalar. In German, equatives are uniform across categories in employing *wie* as a standard marker in scalar as well as non-scalar comparison. This suggests a uniform and, more importantly, compositional analysis of German equatives based on the standard marker *wie*.

In the similarity project a generalized account of German equatives was developed including scalar as well as non-scalar comparison, see (3). The core of this account is (i) the assumption that the standard marker *wie* denotes similarity, like the demonstrative *so*, but without a deictic component; (ii) the idea that equatives express similarity and that this is not inherent in the construction – there is no meaning of equatives as such – but instead due to the meaning of the standard marker *wie*.

- (3) a. Anna hat (so) ein Auto wie Berta. 'Anna has a car like Berta's.'
 b. Anna hat (so) getanzt wie Berta. 'Anna danced like Berta.'
 c. Anna ist so groß wie Berta. 'Anna is as tall as Berta.'

The syntactic analysis is such that *wie*-clauses in equatives are (elliptical) free relative clauses and that equatives are correlative constructions with *wie*-clauses resolving a pro-form *so*. The semantics is compositionally built upon *wie* denoting similarity of either individuals or events: In (3a) Anna's car is similar to Berta's car, in (3b) Anna's way of dancing is similar to Berta's way of dancing, and in (3c) Anna's height is similar to Berta's height. Resolution of the pro-form *so* is implemented via term unification in order to avoid vacuous shifting operations. The analysis has been presented in various talks and is summarized in Umbach, Hinterwimmer & Gust (2021), section 5.1. A full paper on equative comparison is in preparation by Umbach & Özge.

Asymmetry

While in non-scalar cases the generalized analysis of equatives is straightforward, it is more involved in scalar cases because the similarity relation is symmetric while scalar equative facilitate, in some contexts, an asymmetric reading – (3c) may be interpreted such that Anna's height exceeds Berta's height. Moreover, symmetry mostly fails under negation – *Anna ist nicht so groß wie Berta*. ('Anna is not as tall as Berta.') mostly mean that Anna is smaller than Berta.

Asymmetric readings of scalar equatives are accounted for in the formal framework by allowing for similarity classes of extended granularity, including upward segments of a scale, see section 4. But beyond the formal solution the issue of asymmetric readings was addressed in an experimental study (see the term paper by Coulmy 2017). The study starts from the hypothesis that scalar equatives can be denied if the compared item exceeds the standard thereby indicating a symmetric reading (which would be ruled out in standard degree semantics). The availability of such a reading is predicted to hinge on the distance between the standard and the compared item. The stimuli presented in the study were such that asymmetric readings were plausible. For example, Simon is asked whether he is as fast in running 100m as Martin. He answers:

- (4) Simon: (A) Ja, ich bin so schnell. ('Yes, I am.')
(B) Ja, ich bin sogar schneller. ('Yes, I am even faster.')
(C) Nein, ich bin schneller. ('No, I am faster.')

Stimuli are varied with respect to the distance between standard and compared item. In the example Simon's 100m speed is either some milliseconds or two seconds less than Martin's. The results of the study are such that type (B) answers are preferred regardless of whether the distance is small or large (recall that the contexts support asymmetric readings), but that acceptability of type (C) answers increases in large distances. This is evidence that scalar equatives need not have an asymmetric interpretation and the degree semantic account should be reconsidered.

The observation that negated scalar equatives are preferably understood asymmetrically is independently explained by a convexity constraint: on a symmetric reading, negated scalar equatives would yield a disjunctive reading according to which Anna is either smaller or taller than Berta. This entails that the result of negation would no longer be convex. Given that convexity plays a primary role in cognitive economy it is hardly surprising to find such effects in semantics (see also Solt & Waldon 2019 on numerals under negation).

Two strategies of equative comparison

The generalized account of equative comparison proved to be well suited for languages like German in which non-scalar and scalar equatives are uniformly built on a standard marker denoting similarity. This is the case, for example, in Polish, Spanish and Danish. In English, standard markers differ: Non-scalar equatives compare nominal or verbal parameters and use *like* while scalar ones compare adjectival parameters and use *as*, see the translations in (3). However, what came as a real surprise in the course of the project was the finding that Turkish equatives exhibit two different standard markers

distinguishing non-scalar and scalar cases (*gibi* 'similar' vs. *kadar* 'amount', 'as much as'), but unlike English they can both be used across categories. Thus we find equatives based on adjectival as well as nominal and verbal parameters expressing scalar comparison when combined with *kadar* and non-scalar comparison when combined with *gibi*, (5)-(7).

(5) Anna Berta kadar / gibi zeki.

A. B. kadar / gibi intelligent.Cop3sg

kadar: 'Anna is as intelligent as Berta.' (scalar, same degree of intelligence)

gibi: 'Anna is intelligent like Berta.' (non-scalar, similar in the way of being intelligent)

(6) Anna'nın elbisesi Berta'nın-ki kadar / gibi.

A.Gen dress.Poss3sg B.-Gen-Rel kadar.Cop.3sg / gibi.Cop.3sg

kadar: 'Anna's dress is as _____ as Berta's.' (scalar, e.g., same length or price)

gibi: 'Anna's dress is like Berta's.' (non-scalar, e.g., design & color & fabric)

(7) Anna Berta kadar /gibi koşuyor.

A. B. kadar / gibi run.3sg.Prog

kadar: 'Anna runs as _____ as Berta.' (scalar, e.g. duration or frequency)

gibi: 'Anna runs like Berta.' (non-scalar, e.g. style and equipment)

There are at the moment two types of analyses available for the semantics of equative comparison, which take opposite perspectives. Degree-semantic analyses (e.g., Bierwisch 1987, Kennedy 1999) are tailored for scalar adjectival equatives and fail to handle non-scalar cases. Kind-based (Anderson & Morzycki 2015) and similarity-based accounts (Umbach & Gust 2014) take non-scalar equatives as basic and require extra efforts to account for scalar cases. While the two perspectives are commonly considered as competing theories, the Turkish data are evidence that this view can no longer be maintained. It has to be acknowledged that, within the same language, two different strategies of performing equative comparison are manifest – a degree-based as well as a similarity-based strategy – while the choice between strategies depends upon the standard marker. These findings suggest that an appropriate semantic analysis of equatives has to handle the two strategies in parallel without reducing one to the other, and to interpret equatives compositionally based on the meaning of the standard marker. The similarity framework makes a substantial contribution to this end (see Umbach & Özge 2019 and Umbach & Özge in prep.).

3. Reconstruction of the notion of manner by similarity

The expression *wie* is used in comparison constructions, but also as a manner wh-word, in manner questions and also in subordinated clauses. It is argued in the project that the notion of manner can be reconstructed by similarity – manners can be seen as classes of events that are similar, for example, in the way they are performed. From this perspective, *wie* in comparison constructions and in manner use is the same word, with the same meaning, viz. similarity.

The reconstruction of manner by similarity turns out to be of help in the interpretation of complement clauses headed by *wie*. These clauses may have two readings, a regular manner reading and a reading which is close to a declarative clause headed by *dass* ('that'). The latter is restricted to events in progress (in German) and thus called *eventive*. The similarity account reveals that the two interpretations are based on the same wh-word *wie* and moreover explains the subtle difference between the eventive interpretation and a genuine declarative clause (Umbach, Hinterwimmer and Gust 2021).

The two readings of *wie* complement clauses are shown in (8). They can be distinguished by appropriate clarification questions which is a manner clarification question in (8a) and a clarification question addressing the overall event in (8b).

- (8) a. Anna sah, wie Berta ihre Tasche packte. manner reading
 'Anna saw how Berta packed her bag.'

clarification: Und WIE hat sie ihre Tasche gepackt? 'How did she do that?'

- b. Anna sah, wie Berta ihre Tasche packte. eventive reading
 'Anna saw Berta packing her bag.'

clarification: Wie ist das gekommen? 'How come?'

In the analysis, manners are considered as similarity classes. A manner of bag packing, for example, can be seen as a subclass of bag packing events which are similar in being performed sloppily and hastily. When methods are included, a manner of bag packing can be seen as a subclass adhering to a certain sequence of subevents, e.g., running shoes in, then a T-shirt, then a book, and on top a sweater.

The difference between the two readings is traced back to different base-positions of the *wh*-word: In the regular manner reading *wie* has a base position within the VP, whereas in the eventive reading it is base-generated above VP. When base-generated within the VP, the manner modifier combines with the event type (basically by intersection). But when base-generated in the left periphery of the clause the event type is no longer available for composition, and the modifier combines instead with the event token such that it is a stage of performing the event, i.e. an initial sequences in a set of similar sequences of subevents.

Extending the analysis to other languages facilitating a declarative-like use of manner complements it turned out that, in some languages, complements are restricted to events in progress (as in German), for example Polish and Dutch. In other languages, however, for example in English, declarative-like uses may also express plain facts. This is the reason why in Umbach, Hinterwimmer and Ebert (submitted) the analysis has been generalized.

4. The Similarity Framework

The core technical result of the similarity project is a representational framework featuring a non-metric qualitative notion of similarity (see also Tversky 1977, Gärdenfors 2000), which was developed in cooperation with Dr. Helmar Gust (Cognitive Science, University of Osnabrück). The final version is presented in Gust & Umbach (2021). The framework is aimed at the role of similarity in natural language semantics, for example in comparison and in ad-hoc kind formation (see also Barsalou 1983). Starting point was the interpretation of expressions of similarity and sameness such as *so/such*, *wie/like*, *ähnlich/similar*, and *gleich/same*. Although all of these denote similarity in some sense, it would be unsatisfactory to interpret them by a single primitive similarity predicate because their communalities and differences could not be accounted for – for example, the fact that *ähnlich/similar* are gradable while *so/such* and *gleich/same* are not (see Umbach and Gust 2021). Moreover, a primitive similarity predicate would leave no room to account for the observation that certain similarity expressions, in certain contexts, can be used to form ad-hoc kinds, while others cannot, see section 2 and Umbach & Stolterfoht in prep. The elaborate notion of similarity developed in the project provides a more precise semantic interpretation of natural language similarity expressions. Beyond, it points to mechanisms of classification relevant in concept formation.

The framework developed in the project spells out the notion of similarity in some detail without leaving the well-established ground of referential semantics. In accordance with referential

semantics it is assumed that natural language expressions refer to entities, or categories of entities, in the real world. However, similarity judgements are only indirect, mediated by *generalized measure functions* mapping real world entities to points in attribute spaces representing complex features of individuals. The core idea is to have attribute spaces equipped with predicates defined on points therein. These predicates are used to determine the granularity of representations. Similarity is then defined as *indistinguishability with respect to a representation*: Individuals count as similar if their features in a particular attribute space, given a particular granularity, cannot be distinguished. It is important to realize that this system is basically a multidimensional generalization of degree semantics (e.g. Kennedy 1999) complemented by a notion of granularity. This is the reason why this framework is suited to handle a degree-based and a similarity-based strategy of comparison in parallel (see section 2). Finally, it is important to note that this framework is anchored in referential semantics just as much as degree semantics is.

The building blocks of the framework are (i) multi-dimensional attribute spaces with dimensions of various scale levels (ratio, ordinal, nominal); (ii) generalized measure functions which are a generalized version of measure functions in degree semantics mapping individuals point-wise into attribute spaces; (iii) systems of classifiers which are predicates on points in attribute spaces; they determine the granularity with respect to which points in attribute spaces cannot be distinguished. Similarity of individuals is defined via *indiscernability* (this term is borrowed from Rough Set Theory, Pawlak 1998). Below is a slightly simplified version: Given a representation $\mathcal{F} = \langle D, F, \mu, P^* \rangle$ with domain of individuals D , attribute space F , measure functions μ mapping elements of D to points in F and the convex closure P^* of a set of basic classifiers P , similarity of individuals or events is defined via indiscernability of points in attribute spaces:

- (9) a. *Indiscernible* For $x, y \in F$: $x \sim_{\mathcal{F}} y \equiv \forall q \in P^*: q(x) \leftrightarrow q(y)$
 b. *Similar* $\forall x, y \in D : \text{sim}(x, y, \mathcal{F}) \equiv \mu(x) \sim_{\mathcal{F}} \mu(y)$

Similarity defined in this way is an equivalence relation (Tversky's concern against an equivalence relation are addressed in the Gust & Umbach 2021). Asymmetric readings of scalar equatives are accounted for by right closure of classifiers. Representations are ordered by granularity such that points indiscernible in the finer representation cannot be discriminated in the coarser one, (10a). Granularity of representations provides a Kleinian style notion of *more similar*, (10b).

- (10) a. \mathcal{F}' is coarser (\geq) than \mathcal{F} $\forall x, y \in F : x \sim_{\mathcal{F}} y \rightarrow f(x) \sim_{\mathcal{F}'} f(y)$ (f aligns attribute spaces)
 b. Given a representation \mathcal{F} , *more_sim*(a, b, c, d, \mathcal{F}) iff
 $\exists \mathcal{F}' \geq \mathcal{F} : \text{sim}(a, b, \mathcal{F}') \wedge \neg \text{sim}(c, d, \mathcal{F}')$ and
 $\forall \mathcal{F}' \geq \mathcal{F} : \text{sim}(c, d, \mathcal{F}') \rightarrow \text{sim}(a, b, \mathcal{F}')$

5. Expressions of similarity beyond *so* and *wie*

There are numerous words expressing similarity, or indistinguishability, beyond *so* and *wie*, for example *ähnlich* and *gleich* ('similar', 'same'). It seems reasonable to assume that the common core of their meaning is similarity, as in the case of *so* and *wie*. Still, there are significant differences, as shown in Umbach & Gust (2021): while *so* is a demonstrative, *ähnlich* and *gleich* are two-place predicates, and while similarity as denoted by *so* is reflexive, this is not the case for *ähnlich* and, surprisingly, also not for *gleich*. The most challenging difference, however, is gradability: while *ähnlich* is gradable, *so* and *gleich* are not, (11)

(11) (speaker points to Berta's haircut):

- a. *Anna hat mehr so einen Haarschnitt als Claire.
'Anna has more such a haircut than Claire has.'
- b. Annas Haarschnitt ist dem von Berta ähnlicher als der von Claire.
'Anna's haircut is more similar to Berta's haircut than Claire's is.'
- c. *Annas Haarschnitt ist mehr der gleiche wie Bertas als der von Claire.
'Anna's haircut is more the same as Berta's haircut than Claire's is.'

In Umbach & Gust (2021) two questions are addressed. The first is about different constraints on the similarity relation. It is assumed that nouns are lexically associated with canonical sets of features (not to be mistaken for criteria of identity, see Gupta 1980 and Barker 2010). It is shown then that (i) *so* makes use of a subset of these while allowing for additional features; (ii) *ähnlich* makes use of features made salient by the antecedent and (iii) *gleich*, on its type reading, requires all canonical features to coincide while its token reading expresses referential identity and thus features are irrelevant.

The second question is about gradability. Gradability of *ähnlich* is implemented with the help of the Kleinian style *more-similar* predicate presented in section 4, which makes use of less granular (coarser) representations, see (10). In the case of *so* gradation is blocked by the fact that representations other than the actual one are inaccessible because *so* is a demonstrative and thus contexts, including representations, cannot be shifted. In the case of *gleich*, maximal discriminative capacity is required – being *gleich* means being maximally similar. Type identity thus entails indistinguishability in any representation spanned by canonical features. Token identity even entails indistinguishability in any representation whatsoever. To conclude, *so* as well as *gleich* are not gradable because comparatives would require less granular representations, which are ruled out.

In addition to the study of *so*, *ähnlich* and *gleich*, there are two papers on issues of sameness of types and tokens. The first is about the notorious contrast in German between *derselbe* and *der gleiche*. The two expressions appear synonymous in every day language. Nevertheless speakers, on being asked, insist that there is a rule distinguishing the two expressions such that *dasselbe* means token identity while *das gleiche* means type identity – two persons can eat *die gleiche* pizza but not *dieselbe* pizza. It is argued in Umbach (2019) that complementing this rule with the distinction between relative and absolute identity (Geach 1973) explains why *dasselbe* and *das gleiche* appear synonymous – the resulting differences are very subtle. This suggests that the popular pizza rule is not fully wrong but too coarse grained.

The second paper is about the two taxonomic nouns *Art* and *Typ* in German (Umbach to appear). It is hypothesized that they differ in their way of classification: while *Art* refers to kinds indicating that individuals share essential properties canonically associated with lexical meanings, *Typ* refers to classes in arbitrary systems saying that tokens match a model or prototype. The hypothesis is successfully tested against lexicographic data and corpus data and, moreover, confirmed by experimental results.

6. Bridging theme: Comparison via *eher*

The second part of the similarity project includes a section on semantic microvariation connecting to the project *Degree Attenuators* (Stephanie Solt, SO 1157/1-2) by studying cases of semantic microvariation between English and German. Instances were found in a number of the topics touched upon in the two projects. First, there is Solt's work on *much* words cross linguistically (see Solt & Wilson 2021). Concerning equatives, the difference in uniformity of standard markers paved the way to the insight that there are different strategies, which eventually became evident in Turkish, see section 2.

Furthermore, the semantics of German *dermaßen* as an equative marker in contrast to Japanese *hodo* was examined, focusing on polarity sensitivity and presuppositions (Tanaka, Mizutani & Solt 2019).

Finally, there is the paper in Umbach and Solt (to appear) on German *ehler*. It is argued that temporal *ehler* is indexical (unlike *früher* / 'earlier'), comparing closeness to a perspectival center, and that the non-temporal readings (epistemic, preference, metalinguistic) inherit their basic structure from the temporal one. The analysis of the non-temporal readings is embedded in a Kratzer-style ordering semantics. *Eher* has also received attention from a diachronic perspective, where it has been compared to the English *rather*. It is shown in the paper, however, that from a semantic point of view *ehler* and *rather* differ substantially.

Other activities

Video Strukturelle Grammatik, semantische Universalien und Arbitrarität -- ein Gespräch mit Manfred Bierwisch

Manfred Bierwisch is professor emeritus at the Humboldt University of Berlin, a member of the Berlin-Brandenburg Academy of Sciences, and one of the most eminent linguists in Germany. In cooperation with Annette Leßmöllmann (Karlsruhe Institute of Technology) a video interview with Manfred Bierwisch was conducted talking about his linguistic research which had a significant influence on the development of linguistics – not only in Germany, but also internationally. In the interview, Manfred Bierwisch describes how and why he moved from a structuralist understanding of linguistics influenced by the Prague School, via Chomsky's notion of a generative grammar to a theory of semantics in which the meaning of lexical units is determined only in connection with conceptual knowledge. At the same time, the audience learns how the first working group on structural grammar came into being at the end of the 1950s, why structuralism went from being a point of contention within linguistics to a political issue, why the group was finally disbanded, and how the second working group came into being and survived the fall of the Berlin Wall.

The interview was edited and subtitled in English and is available at www.gespraech-manfred-bierwisch.de/.

Cooperation

Dr. Helmar Gust (Cognitive Science, University of Osnabrück) – formal aspects of the similarity framework;

Dr. Stephanie Solt (ZAS Berlin) – the meaning of degree expressions, semantic microvariation, modality;

Prof. Louise McNally (UPF Barcelona) – cognitive aspects of semantics, degree semantics;

Prof. Britta Stolterfoht (University of Tübingen) – manner modifiers, features of comparison, experimental methods;

Prof. Umut Özge (Middle East Technical University, Ankara) – similarity demonstratives and comparison constructions in Turkish;

Dr. Yvonne Treis (CNRS, LLACAN) – expressions of similarity in Cushitic languages, typological aspects of equatives

Dr. Stefan Hinterwimmer (University of Wuppertal) – semantic representation

Workshops

Workshop *Comparison Constructions* (im Zusammenhang mit Seminar "Vergleichskonstruktionen"), 21. Dezember 2015, Universität Köln

ESSLLI 2016 Workshop *Referential Semantics one step further: incorporating insights from conceptual and distributional approaches to meaning*, organized with Louise McNally (UPF Barcelona), 22-26 August 2016, Bolzano, Italy.

Working session on *Exclamatives*, with Cleo Condoravdi (Stanford / Göttingen) and Stephanie Solt. 26. July 2017, ZAS Berlin.

Workshop *Microvariation in Semantics*, organized with Stephanie Solt, 6. Sept. 2017, ZAS Berlin

Workshop *Records, frames and attribute spaces*, in cooperation with CLASP Gothenburg, 13.-14. March 2018, ZAS Berlin

Annual Event Semantics Meeting, organized with S. Hinterwimmer, 24.-25 Nov. 2017 Universität Köln

Working sessions on planning/evaluation of experiments investigating *Manner modification via demonstrative 'so'*, with Britta Stolterfoht (Tübingen) 27. June 2018 and 25.-26. Jan. 2019. at ZAS.

Working sessions on *Strategies of equative comparison* 27.-31. May 2019 with Umut Özge (METU Ankara), and 15.-22 Jan. 2020 with Umut Özge and Yvonne Treis (CRNS, LLACAN).

Annual Event Semantics Meeting, organized with Berit Gehrke, 1.-2. Nov. 2019, HU Berlin.

Workshop *Non-interrogative subordinate wh-clauses*, organized with Łukasz Jędrzejowski (University of Cologne), February 7-8, 2020, University of Cologne

(The final workshop of the project had to be canceled due to the Corona pandemic.)

Supervision of theses / teaching

Andres Soria Ruiz, PhD (with Isidora Stojanovic, Institut Jean Nicod, Paris),

Regina Zielecke, PhD (together Klaus von heusinger, Universität Köln),

Felix Frühauf, BA, HU Berlin

Julia Otterpohl, BA, Universität Köln,

Victoria Mental BA, Universität Köln,

Seminar *Formen von Äquativen*, SoSe 2017, Universität Köln, resulting in a number of experimental term papers, including Coulmy 2017 (quoted above).

Reviewing

Maria Klimek-Cieschinger, Un.Osnabrück (Dissertation), 2015; María Inés Crespo (Dissertation), Un. van Amsterdam, 2015; Berry Claus, (Habilitation), HU Berlin, 2017.

Journal of Semantics, Linguistics & Philosophy, Natural Language and Linguistic Theory, Semantics & Pragmatics, Glossa, Zeitschrift für Sprachwissenschaft, Nous, (a.o.), and numerous conferences.

(A list of references mentioned above is provided in the appendix. A full publication list is available at www.carla-umbach.de/publications.html)

DFG project 'Similarity I' (UM100/1, 2012–2015) and 'Similarity II' (UM100/3, 2017–2020)

The linguistic expression of similarity

Part II: Final report

References

- Anderson, C. & Morzycki, M. (2013) Degrees as Kinds. To appear in *Natural Language and Linguistic Theory*.
- Barker, C. (2010) Nominals don't provide criteria of identity. In M. Rathert, A. Alexiadou (eds.) *The semantics of nominalizations across languages and frameworks*. de Gruyter Berlin, 9-24.
- Barsalou, L. W. (1983) Ad hoc categories. *Memory & Cognition* 11 (3), 211-22.
- Bierwisch, M. (1987) Semantik der Graduierung. In M. Bierwisch, E. Lang, E. (eds) *Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven*. Akademie Verlag, Berlin, 91-286.
- Carlson, G. N. (1980) Reference to Kinds in English. Garland, New York & London.
- Carlson, G. (2010). Generics and Concepts. In F. J. Pelletier (ed.) *Kinds, Things and Stuff*. Oxford, Oxford University Press, 16-36.
- Coulmy, Christa (2017) Zwerge und Riesen. Eine Studie zur Obergrenze von *at least*-Lesarten bei Gradäquativen. Termpaper in seminar *Äquative*, supervised by C.Umbach, Universität Köln, SoSe 2017.
- Gärdenfors, P. (2000) *Conceptual Spaces*. MIT Press.
- Geach, P. (1973) Ontological relativity and relative identity. In M.K. Munitz (ed.), *Logic and Ontology*, New York: New York University Press.
- Goodman, N. (1972) Seven strictures on similarity. In Goodman, N. (ed.) *Problems and Projects*. The Bobbs Merrill Company, Indianapolis and New York, 437–447.
- Greenberg, Yael (2003) *Manifestations of Genericity*. New York: Routledge.
- Gupta, A. (1980) *The logic of common nouns: an investigation of quantified modal logic*. New Haven, CT: Yale University Press.
- Kaplan, D. (1989). *Demonstratives*. In J. Almog, J. Perry & H. Wittstein (eds.), *Themes from Kaplan*, Oxford University Press, 481–563.
- Kennedy, C. (1999) *Projecting the adjective*. Garland Press, New York.
- Maienborn, Claudia and M. Schäfer (2011) Adverbials and Adverbs. In *Semantics. An International Handbook of Natural Language Meaning.*, vol. 2, 1390–1420. Berlin: Mouton de Gruyter.
- Pawlak, Z. (1998) Granularity of knowledge, indiscernibility and rough sets. *Proceedings of the IEEE International Conference on Fuzzy Systems*, 106-110.
- Prasada, S. & Dillingham, E. M. (2006) Principled and statistical connections in common sense conception. *Cognition* 99, 73-112.
- Quine W. V. (1969) *Ontological Relativity and Other Essays*. New York, Columbia, University Press.

- Solt, Stephanie & Brandon Waldon (2019) Numerals under negation: empirical findings. *Glossa*, Article 113.
- Solt, Stephanie and E. Cameron Wilson (2021) M-modifiers and varieties of degree modification. Talk at workshop on *Theoretical and Experimental Approaches to Modification* Universität Tübingen, Jan. 2021.
- Tanaka, Eri, Kenta Mizutani & Stephanie Solt. 2019. Existential semantics in equatives in Japanese and German. *Proceedings of the 22nd Amsterdam Colloquium*, 377-386.
- Tversky, A. (1977) Features of similarity. *Psychological Review* 84: 327-352.
- Umbach, Carla, Stefan Hinterwimmer & Cornelia Ebert (submitted) Depictive manner complements.
- Umbach, Carla & Britta Stolterfoht (in prep.) Ad-hoc kind formation by similarity.
- Umbach, Carla & Umut Özge (in prep.) Scalar and non-scalar comparison across categories: The case of Turkish equatives.