Fachbereich Sprachwissenschaft

Universität Konstanz

Arbeitspapier 104

Regine Eckardt

On the underlying mechanics of certain types of meaning change

On the underlying mechanics of certain types of meaning change

Regine Eckardt
University of Konstanz / SFB 471
regine.eckardt@uni-konstanz.de

Introduction

How can truth value based semantics in the tradition of Montague (1974) be extended by a diachronic dimension in an explanatorily adequate way? The present paper offers a first answer to that question. Yet, it is not exclusively focused on language history.

In order to allow for a meaningful link between synchronic language stages, the architecture of these stages has to be adjusted in the first place. More concretely, we will have to reconsider the notion of word meaning in logical semantics and implement a way to capture their internal conceptual structure. This refined picture of synchronic semantics will then be suitable to explain a certain type of variation in word meaning, and moreover offer the basis for an account of metaphoric language use.

The changes in synchronic semantic theory that I will propose do not, in and off themselves, look very spectacular. Nevertheless, they should eventually allow us — apart from being of use in diachronic semantics — to answer some of the core criticisms that have been raised against truth value based "logical" semantics by those who suggest that logical semantics should be replaced by "conceptual" semantics (e.g. Jackendoff (1983), Langacker (1990), Gärdenfors (t.a.), Lakoff (1994), Lakoff & Johnson (1999)). The mistrust in truth value based semantics may have been nurtured by the rigorous positions defended by some proponents of truth value based semantics, notably Kripke and Putnam (but not only these).

However, it is not clear how losses and gains would add up, once we gave up a reliable and interpersonal notion of truth, semantic composition, quantification and logical implication relations, in favour of a theory which is exclusively concerned with concepts-in-the-head. Hopefully, the account I propose will allow to integrate some favourable aspects of both positions. My arguments will, however, be focussed on the concerns of diachrony, complemented if necessary by observations about language acquisition.

The first part of the paper will offer an account of how word meanings are established in a given language in the first place, based on a modified variant of the causal theory of reference. In a brief interludum, I will discuss how the "meanings" computed in part one can be grasped by single speakers. This ties in with the question whether and to what degree the resulting account deserves the label “cognitive semantics”. After this intermezzo, part three will turn synchronic language stages into a moving diachronic picture. In a final part, I will briefly relate the resulting theory to prototype theory, the interpretation of metaphor, and the question of scientific progress.

1. The creation of meaning

1.1. Baptising, classical version; and three drawbacks

If we want to investigate the question how speakers can *change* the meaning of words in a language, it might be a good starting point to ask how speakers can *establish* the meaning of a new word in the language. Within the framework of truth value based semantics, the causal theory of reference (Kripke (1972), Putnam (1975), with a wealth of subsequent literature) is at present still one widely accepted account for this question. Its paradigm case is the introduction of a new proper name into a language. Kripke (1972) argues that the meaning of a name is established in an initial act of baptising, in which the name is "attached" to an individual R. This leads to rigid designation, or, in terms of intensional semantics, to constant functions from worlds to an individual like in (1). Let *name* be a proper name in the language under investigation.

(1) [[*name*]] is a function $f: D_S \rightarrow D_e$
where $f(w) = R$
for all possible worlds w.

The causal theory of reference nicely captures some appealing basic intuitions: Firstly, we feel that those who baptise an individual by a name should be in full command of the meaning of that name. At the time of being baptised, however, the respective individual often does not yet exhibit those distinctive properties which might later on be linked to the name (and, according to description theories, constitute the "meaning" of the name). Thus, the meaning of a proper name can only depend on its referent. Secondly, the definition reflects the linguistic convention that proper names stand for one and only one thing.

The account was subsequently extended to natural kind terms[1] which are treated like proper names for natural kinds. According to Putnam (1975), the meaning of a natural kind term is again established in an initial act of baptising with reference to some sample *R* of the natural kind in question. This time, however, we want to name the entire kind, not only sample *R*. This is captured by the definition in (2). Let *nkt* be a natural kind term, let R be the sample that was pointed at in the baptising. *Subsidentity* is meant to be a trans-world relation which holds between two lumps of matter exactly if they are of the same substance.

(2) [[*nkt*]] = f: $D_S \rightarrow D_{(e,t)}$ where
$f(w)$ = { *b* | *b* is a lump of matter in *w*, and *b subsidentical to R* }

Importantly, the causal theory of reference in its classical form is committed to the position of philosophical *realism*: According to this position, the extension of the *substidentity* relation is determined by reality (rather than common knowledge, scientists, etc.). The external world will determine *which* lumps of matter are of the same kind, *which* animals are of the same kind, etc. This leads to the prediction that the meaning of a word like "gold" does not depend on how much we know about gold (both, as individual speakers, and as speaking community). The position of realism has earned its high reputation in philosophy by the fact that it can account for scientific progress in a simple and elegant way. The reader is referred to Devitt & Sterelny (1987), (1998) for a more thorough discussion of the causal theory of reference.

From a linguist's perspective, however, the account is less satisfying. Let me just name three drawbacks.[2]

a. For proper names, there is a linguistic convention that one name should refer to one person/object, even if speakers cannot distinguish the referent. This convention can be observed whenever speakers realise that they have erroneously used one name *N* for two different individuals *a* and *b* (which might be twins). Such a discovery requires immediate *correction*: Speakers will ask which one of *a* and *b* is the actual bearer of the name *N* or else will decide that there were two homophonous names N_1 and N_2 in play.
There is no comparable observable convention for an alleged class of natural kind terms in language.

---

1 I will assume, in accordance with most of the literature, that words like "gold", "water", "jade", "tiger", "zebra" are examples of natural kind terms.

2 Point (a) and (b) are adopted from Zemach (1976); (c) to my knowledge has not yet been stated in this clear-cut form. Yet it may be implicit in other work (or explicit in work which I missed).

Firstly, there is no *grammatically* distinct class of such nouns. For example, the words "gold" and "earth" behave exactly parallel in all linguistic respects; yet, one qualifies as a natural kind term, the other doesn't.

Secondly, there is no evidence in favour of a semantic convention to the end that a certain class of words should refer to one and only one "natural kind". This can be seen from the *lack* of immediate corrections in those cases where speakers discover that they systematically applied the same word to two or more distinct natural kinds (i.e. not just in stray cases; see section 4.2). They will never react in a way similar to the erroneous use of a proper name. They will *not* ask: "which of the substances is the true *N*?", and they will *not* decide to adopt homophonous natural kind terms N_1, N_2, for the natural kinds in play.

Therefore, "natural kind terms" can only be those words which accidentially happen to apply to natural kinds. There is no empirical evidence that would justify the claim implicit in the causal theory of reference, namely that certain words obey the linguistic (*lexical*) convention "apply to a natural kind". No word does, as little as any word obeys the *lexical* convention "applies to substance with ph below 7".

b. If "being a natural kind term" only depends on the eventual nature of the word's referents, then what will happen if nature does not fall into "natural kinds"? Let us, for the sake of exemplification, adopt the position that animal species are a case of natural kinds.[3] Presently, our increasing knowledge about genetic structure of animals and command of techniques to alter this structure might lead, in the not-so-far future, to a continuous space of gen-designed animals which all are to a certain degree inclined to interbreed, or uninclined to do so, and do so with fertile children.[4] If such a scenario became true, then it would be hard to find a sensible notion of "being of the same kind as animal X" and the idea of species as natural kinds would break down.

It is somewhat more far-fetched but not impossible to imagine similar developments in *all* domains where we diagnose natural kinds today. There might be no natural kind terms at all. A theory of meaning which relies so much on a

3 For an opposite view, see Wilkerson (1995).

4 The family of canides offers several examples of "soft" species boundaries. Generally, it has been found that the species of wolves/dogs (*canis lupus*), coyotes (*canis latrans*) and gold jackals (*canis aureus*) can interbreed and have fertile offspring; they just are not inclined to do so under natural circumstances (Senglaub (1978)). More specifically, Grzimek (1987) reports that "...coyotes and dogs interbreed without human interference...", "... the bastards reproduce quickly..." but "... wolves do not interbreed with coyotes, because the two species don't like each other. If a wolf can get hold of a coyote, it will kill and eat it ..." (p. 106; my transl). Wolves and dogs, however, are accepted as one single species in the biological sense (Senglaub (1978); thus the question whether *canis lupus* and *canis latrans* are distinct species is open. Note however that Grizmek has to be understood as a tendency rather than a universal; Fox (1975), Kennelly (1978) describe cases of wolf x coyote hybrids in captivity.
Such complications are only partially tractable with the official criterion to distinguish animal species: The technical term is that animals of one species must be capable to form an "ecological unit" (Mayr (1942), (1966), (1988)).

certain kind of reality might turn out to describe an empty set of words. Note once more that this theoretical discovery would not be matched by any empirical discovery “that certain words did not refer at all”.

c. The causal theory of reference will predict meaning change where, according to all empirically observable linguistic behaviour, no meaning change takes place. Let me give an example.

The Chinese word *yu* was introduced by ostension, on the basis of lumps of the chemical substance nephrite.[5] By geological accident, until about 1200 all instances of *yu* that were met by Chinese speakers were of nephrite. Only in the 13th century the first pieces of the substance *jadeite* reached China from Burma, and trade with jadeite at a larger scale only started in the 18th century. The first pieces of jadeite, however, were immediately accepted as *yu*. Not only did they look like previous *yu* but they were even considered *yu* of superior quality. Jadeite, so to speak, filled the prototypical core of the previously known *yu* category.[6]

The question never arose whether lumps of jadeite qualified as *yu*. Chinese speakers in 1900 could not say something like "In former times, they would not have called this lump here (pointing at jadeite) *yu* but we today do." Nor could they say: "In former times, they would call this here *yu* but today, we don't." Present-day speakers *do* call both, jadeite and nephrite, by the name *yu*. (*jade* is actually mentioned in Putnam (1975) as an example for a natural kind term which is not a synonym to a chemical substance term.) According to the internal view of the language community, *no meaning change* has taken place.

The causal theory of reference, however, will diagnose that those speakers who first accepted jadeite as *yu* made a mistake which led to a *change in meaning* of the word *yu*.[7] This diagnose is in evident mismatch to the empirical findings.

Apart from these criticisms, the causal theory of reference is a bit inconvenient in a further respect: It only applies to a very small fraction of even the nominal lexicon. If the account was satisfying otherwise, this aspect might have been acceptable. Yet, it will turn out that the modifications which bring us closer to a theory of meaning of actual words of natural languages will also extend the range of application. Thus, there seem to be several good reasons to ask for a more flexible, less externalist account.

---

5 For the sake of simplicity, I will ignore the fact that other substances were classified as *yu* of inferiour quality which were lateron excluded from the extension of *yu* altogether when the systematic correlation between "low quality" and "different chemical substance" was revealed. If the reader is worried by this fact, a shorter version of the story can be told for the European word *jade* which circumvents these difficulties.

6 See Turner, J. (1996): *The dictionary of art*, vol. 5 and 6 (“China”)

7 This becomes clearly visible in Haas-Spohn's (1994) explicit account of the causal theory of reference.

1.2. A modified account

The idea is certainly appealing that the class of objects that is denoted by a word ω should somehow "cluster around" those things which were *first* called "ω". We will leave it open whether these first uses were deliberate baptises in the literal sense or whether the word ω was established in a more implicit manner (like for instance loan-words). Can we retain this appealing aspect of the classical theory while avoiding its drawbacks?

In the definition in (2), "realism" is reflected in the relation of *substidentity.* Only reality (or, perhaps, God) knows which things are identical in that sense. Certainly, speakers don't know it. I suggest to replace it by some trans-world *similarity relation* ρ. Loosely speaking then, the word ω applies correctly to the first sample *R* which was referred to in baptising, plus any other object *K* which resembles (= ρ) *R* in the appropriate sense. This will require some more comments.

In contrast to the (classically) unique concept of identity, there are more than one respect in which things can be similar. Different choices of respect-in-which-things-are-compared will lead to different classes of things-which-are-similar-to-*R*. A classical place where this observation is discussed (but by no means the only one) is Goodman (1972). If we want to establish a new word ω by pointing at a sample object *R*, we therefore need to specify the respect in which other referents of ω should resemble *R*. It is in that sense that I was talking about "some" similarity relation ρ in the previous paragraph.

How do we determine which is the intended similarity relation? In addressing this question, it is of interest to note that even the classical causal theory of reference is based on analogous choices. Sterelny (1983) observes that, in order to accomplish a baptising, one needs to specify the taxonomic level at which *subsidentity* is to be determined. After all, he argues, it is not predetermined whether we see a female pig *R* as a sample of the class of sows, the class of pigs, or the class of mammals. All taxonomic levels are equally well supported by contemporary science, and in is not clear whether reality attributes primacy to one level over the others.[8] Sterelny calls this the *qua question* and suggests that it is answered by the *context* in which a baptise takes place.

I will adopt his proposal and assume that the utterance context of a baptise will render one similarity relation ρ salient which will then serve as the respect in which the sample is to be generalised.[9] Thus, we come to the following modified theory of baptising:

---

[8] This argument is based on the assumption that contemporary science is, or at least might be, approximating reality. Without such an assumption, both the status of claims about reality and the role of science become dubious. I want to thank Anna Pilatova for making this clear to me.

[9] Yet, our range of possible answers is much larger than Sterelny's. Where he considers:"Should Fido stand for the class of mammals, for dogs, for male dogs, ... ?", we will also allow for "should Fido stand for the class of pets, eatable animals, trainable animals, house guard devices, ...".

(3) If word ω has been introduced in baptising context *c* on the basis of sample *R* then
[[ω]] = a function *f* from worlds to sets of objects such that
f(*w*) = { *x* | *x* is an object that is ρ-similar in *w* to *R* }
The similarity relation ρ was rendered salient by context *c*. It will sometimes be called "the similarity relation that belongs to the meaning of ω".[10]

This definition is intended as the starting point for an epistemic notion of meaning. It is no longer *reality* which cuts out the classes of objects denoted by a word. Extension and intension depend on speaker's *interests*, as well as their *knowledge* at the time of introduction of ω. Similarity is a "realistic" notion only in the sense that the human way to think about reality is part of reality itself. Similarities reflect how speakers will structure the sensual input caused by an independently existing external reality.

At this point, two questions naturally arise: The question about the content of similarity relations, and the question about their formal nature. These questions will be addressed in the next section.

1.3. Ways of being similar

1.3.1. Empirical support

It has often been observed that humans have an intuitive notion of similarity (of objects, sensory inputs), and that these similarity judgements are, at least in certain cases, at the core of category formation and, indirectly, at making judgements about the extension of words.

> "We cannot easily imagine a more familiar or fundamental notion than [similarity], or a notion more ubiquitous in its applications. On this score it is like the notions of logic: like identity, negation, alternation, and the rest. And yet, strangely, there is something logically repugnant about it. For we are baffled if we try to relate the general notion of similarity significantly to logical terms."
>
> Quine (1969):117)

---

10 Note that this manner of speaking suggests that each reading of word ω is uniquely based on one distinguished similarity relation ρ, and perhaps even stronger that this similarity can be uniquely determined from the intension of ω. Actually, this might be too strong an assumption. Yet, no problems should arise by this sloppyness. In order to account for cases where the meaning of ω could equally well be based on several different similarity relations $\rho_1, \ldots, \rho_n$, we can replace "the similarity relation of ω" by "one of the similarities" or "all the similarities", depending on context. I will avoid this further complication throughout the text.
Note moreover that I will be talking about "the meaning" of ω where we should say more precisely: "The meaning of word ω in a certain reading".

The influence of such similarity judgements becomes most prominent in those cases where they can not be replaced or justified by explicit criteria for categorization. Explicit criteria would certainly be more satisfying — yet, no human has yet been known who would structure her world exclusively by explicit criteria according to which that world neatly falls into categories.[11] A great number of categories is formed on the basis of intuitions beyond justification:

> "How am I able to obey a rule? (...) If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what I do.' (Remember that we sometimes demand definitions for the sake not of their content, but of their form. Our requirement is an architectural one; the definition a kind of ornamental coping that supports nothing.")
>
> Wittgenstein, Philosophical Investigations I, §217

There also is common agreement that even little children are in command of at least some similarity judgements. Without this ability, the sensory input they are confronted with would be an unstructured, ungraspable continuous noise.

> "Without some such prior spacing of qualities, we could never acquire a habit; all stimuli would be equally alike and equally different. These spacings of qualities, on the part of men and other animals, can be explored and mapped in the laboratory by experiments in conditioning and extinction. Needed as they are for learning, these distinctive spacings cannot themselves be learned; some must be innate."
>
> Quine, (1969, p. 123)

A growing stock of experience and knowledge allows the child to evade the "original sim" (Keil 1989, chap.10) and get access to more sophisticated ways of being similar.

> "...working out a system of perceptual dimensions, a system of *kinds* of similarities, may be one of the major intellectual achievements of early childhood."
>
> Smith (1989:146)

---

[11] It is not even clear whether such an enterprise can succeed at all. While the nature of mathematical objects like "possible geometry", "possible group", etc. can be fixed by relating some core operations and constants *to each other* axiomatically, an analogous exercise for the entire body of concepts has so far, at best, resulted in thesauri or other encyclopedic collections. The absurdity of a task like "learning Japanese exclusively by studying (nonillustrated!) Japanese encyclopaediae" suggests that such a system of cross-references is hardly sufficient to convey meaning.

It is also commonly assumed that there is a recursive process of deriving higher similarity relations from the more basic ones (which are, conveniently, accessible through language).

> " ... many of the quality dimensions of human conceptual spaces are not directly generated from sensory inputs."
>
> Gärdenfors, (t.a.), chap.1:24.

This already implies that things can be similar in more than one way. Goodman, in "Seven strictures against similarity", writes:

> "...we must recognize that similarity is relative and variable, as undependable as indispensable. (...) We have to say (...) in what respect two things are similar."
>
> Goodman (1972:444)

Adult categorization behaviour, in turn, is commonly studied on the basis of similarity judgements. Statistical evaluation of such elicited similarity judgements allows to derive the number of independent quality dimensions the subjects base their judgements on. These techniques can be tested and adjusted by applying them for similarities where we already have a well-developed model of the quality spaces which are involved; the most frequently quoted example are judgements about colour similarity, which lead to the dimensions of hue, brightness and saturation. Gärdenfors (t.a.) offers a rich range of examples of similarity relations on some perceptual domain, and also introduces the techniques which lead from pointwise similarity judgements to a fully measured domain.

There were some attempts to capture similarity or similarities in terms of a binary, or ternary, or polyadic relation plus an appropriate axiomatization. Goodman (1951: chaps. IV-VI), (1972: chap. IX) offers one of the more sophisticated approaches of this kind. Another brief formal proposal is offered in Lewis (1983:347-8). Yet, judgements about similarity are not turned into a coherent and sufficient axiomatization as easily as, for instance, judgements about linear ordering. Let me illustrate this with an example. Tversky (1977) observed that subjects' responses to two questions in pairs like the following can differ:

> Is China similar to North Vietnam?
> Is North Vietnam similar to China?

He took this as evidence that similarity relations do not obey an axiom of symmetry. However, he failed to take into account that confrontation with different core samples will

lead to different salient respects in which the similarity judgements are to be made. In order to judge similarity to North Vietnam (in the early 1970s), subjects will probably consider different properties as crucial than in order to judge similarity to China. (The notion of similarity relation that I will adopt in the next section will thus be symmetric; in spite of Tversky (1977)).

Let me stress that our enterprise in formal semantics does *not* consist in explaining or deriving qualities on the basis of similarity judgements. My more modest proposal is this:

> Replace one kind of empirically observable, theoretically basic ability of speakers — the ability to determine real and counterfactual word extensions
> by another empirically observable, theoretically basic ability of speakers — the ability to make similarity judgements for arbitrary objects relative to given core exemplars, and a given similarity respect.

The advantage of such a shift of the "basic" level of semantic theory lies in the fact that the new level of empirical import that I propose will allow us to capture more semantic phenomena than just what some words mean.

1.3.2. Formal issues

What is the formal nature of similarity relations? Following recent proposals by Gärdenfors (various places) and Zeevat (1998), I will assume that similarity should be spelled out as "close in distance", relative to a given metric d on the domain of objects D one wants to classify.[12] A metric d on D is a function d: $D^2 \rightarrow |R^+$ such that

(4) $d(x,y) = d(y,x)$
$d(x,x) = 0$
$d(x,y) \leq d(x,z) + d(z,y)$

A welknown metric is the distance between points in an Euclidean space. Gärdenfors discusses the metrical conceptual spaces of colour perception, sound perception and a simple gestalt model. Another example of a metric is the trivial metric that can be defined by a partition $\{A_1, \ldots, A_n\}$ on domain D where

[12] For the advantages of this view over a theory of concepts within fuzzy logic, see the extensive discussion in van der Does and Lambalgen (1998) and the references therein.

(5) $d(x,y) = 0$ iff x and y are in the same partition set A_k
$d(x,y) = v$ for some positive real v else

The trivial metrics evidently fail to capture the distinction between core and periphery of a category; however they offer a simple way to generalize our approach to cases where categorization is actually based on simple yes/no criteria (like male/female, being adult in the legal sense, etc.)[13]

There are various ways in which distances can be turned into similarity. For one, we could decide that two objects are similar $sim(x,y)$ iff their distance is below a contextually fixed threshold value ε.[14]

(6) **similarity as closeness**
$sim(x,y)$:<=> $d(x,y) \leq \varepsilon$ for some given ε.

Using this notion of similarity, one predicts that the extension of a category will lie circle-like around a fixed set of core exemplars. (Evidently, it is crucial that this set be fixed.) (6) was acknowledged as an appealing picture for category formation by Quine:

> "One might be tempted to picture a kind, suitable to a comparative similarity relation, as any set which is 'qualitatively spherical' in this sense: it takes exactly the things that differ less than so-and-so much from some central norm."
>
> Quine (1969:119)

Yet, this idea does not do justice to the intuition that, due to the closeness or remoteness of known negative samples, a category might "stretch out further" in one direction than in another.

Zeevat (1998) proposes a way of category-formation based on positive and negative samples. Negative samples obviously are a first step of remedy against circle-shaped categories. However, Zeevat's definitions primarily lead to a notion of categorization which inherently has a temporal dimension.[15]

---

13 Note that Gärdenfors makes the further assumption that d(x,y) = 0 iff x=y. Many actual examples of distance functions, and especially the empirically supported ones which Gärdenfors discusses obey this further restrcition, but it disallows to account for such categorical distinctions in a simple fashion.

14 This proposal can be understood as the binary version of those continuous similarity measures used in cognitive psychology, where similarity is defined as the negative exponential of distance (see e.g. Hahn and Chater (1997))

15 This is not explicitly noted or discussed in Zeevat (1998).

(7) **inductive similarity (Zeevat Categorization)**

- Let (D,d) be a domain D with a distance function d over D.
- A *sample* of D is a pair of finite sets (D^+, D^-) such that $x \ D \ y \ D^+ \ z \ D^-(d(x,y) < d(x,z))$
  (existence of a prototype in D^+)
- The dissimilarity k of a sample (D^+, D^-) is defined as
  $k := \max_{x,y \ D+} d(x,y)$.
- A new object fits into the sample iff its addition does not increase the dissimilarity of the sample.

These basic definitions suggest that categories are formed by inductively increasing the sample by objects which do not increase dissimilarity. The result of this process, however, is not well-defined. It depends crucially on the order in which objects are looked at, as a simple example in the two-dimensional euclidean plane will reveal: Depending on whether we augment the initial sample $D^+ = \{a,b\}$ by point c or d, the resulting category will look like in figure 1 or figure 2. (The pictures are not entirely faithful - the categories actually will look like "blown-up triangles".)

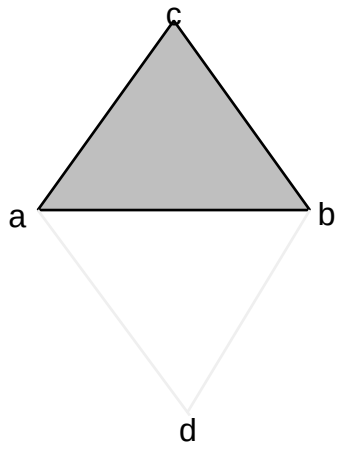

figure 1

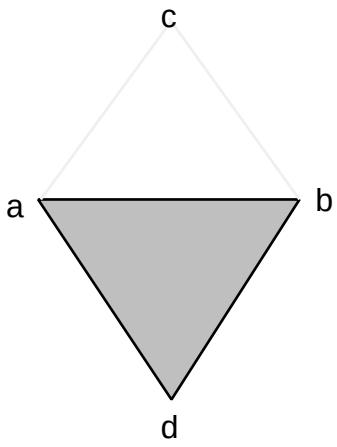

figure 2

It would be an interesting task in its own right to apply this order-dependency of Zeevat's definition in order to account for the way in which (especially artifact) categories depend on the accidential order of new discoveries or inventions. I will not follow this line here.

Zeevat uses negative samples only to ensure that the initial samples form a continuous space, so to say. The idea that categories should have no "holes" has been stated in a more explicit fashion by Gärdenfors (Gärdenfors (1993), (1996b), (t.a.))
Gärdenfors proposes a notion of concept which is also based on a geometrical structure on the domain. Offering a great number of actual case studies from psychology and cognitive sciences which support the assumption of such notions as distance and in-betweenness, he argues in favour of the following main points:

- **Distance**: Categorization takes place on a domain *D* with some distance function *d*. This reflects the empirical evidence that our domains of perception come equipped with (simple or derived) notions of distance.

- **Dimensions**: Different notions of distance in different conceptual domains (dimensions) can be integrated into a notion of distance on a multidimensional space. The integration can happen in certain mathematically determined ways. But, humans might also master notions of distance on the multidimensional space which build on the lower metrics in a less perspicious way.

- **No holes**: The following possible restrictions on concepts are discussed:
  (a) Only convex spaces in *D*, with respect to *d* can be concepts.

(b) Only star-shaped spaces *D*, with respect to *d* can be concepts.

- **Categories cluster around prototypes**: Voroni Tesselation allows to derive a partition of an object domain which is based on distances and prototypes:
  Given a domain *D* with a notion of distance *d*, and a set of (isolated) prototypes $P = \{p_1, p_2, \ldots p_k\}$, there is exactly one way to cut up *D* in such a way that
  x is in category of p_i iff $d(x,p_i) = min_i(d(x,p_j) \mid p_j \quad \{ p_1, p_2, \ldots p_k \})$
  We define $[p_i] := \{ x \mid d(x,p_i) = min_i(d(x,p_j) \mid p_j \quad \{ p_1, p_2, \ldots p_k\}) \}$

Gärdenfors' initial goal was to distinguish those properties which are used in inductive generalizations from those that do not give rise to inductive generalizations. He suggests that the difference between "prejudice-prone" sets and mere set-theoretically existent ones can be captured by the same geometrical notions that can be used to model similarity judgements, to explain concept formation, and to reflect the prototype-centered internal structure of many natural categories.

The use of Voroni tesselations presumes a more global viewpoint of the categorizing agent, and hence allows for more than circle-shaped categories around core exemplars. Let me repeat the core definition for further reference:

(8) **Similarity by shared prototype**

Given a domain *D* with a notion of distance *d*, and a set of (isolated) prototypes $P = \{p_1, p_2, \ldots p_k\}$, there is exactly one way to cut up *D* in such a way that

x is in category of p_i iff $d(x,p_i) = min_i(d(x,p_j) \mid p_j \quad \{ p_1, p_2, \ldots p_k \})$

Define $[p_i] := \{ x \mid d(x,p_i) = min_i(d(x,p_j) \mid p_j \quad \{ p_1, p_2, \ldots p_k\}) \}$

We can indirectly link any two elements *x*,*y* in *D* by

sim(*x*,*y*) :<=> there is a p_i such that *x* $[p_i]$ and *y* $[p_i]$

Under this notion, similarity is primarily a relation on *D* x $\{ p_1, p_2, \ldots \}$ and crucially depends on *D*, *d*, and the choice $\{ p_1, p_2, \ldots \}$.

Note that non-spherical categories (Voroni tesselation) require the previous choice of prospective prototypes. These are presumably chosen on the basis of similarity as closeness. The account for computing intensions that was proposed in (3) is open with respect to the question whether the similarity to *R* is determined by distance alone, or by using foils (i.e. core elements of "competing" categories). That choice will probably depend on the degree to which domain *D* is explored or *terra incognita*. Possible similarity respects can therefore be of either of the two forms given in (9):

(9) **similarity as closeness**:
Based on distance *d*, domain *D*, threshold ε.
$sim(x,y) <=> d(x,y) \leq \varepsilon$

similarity by shared prototype:
Based on distance *d*, domain *D*, set of prototypes *P*.
$sim(x,y)$ iff *x* and *y* are in the same $[p_i]$ for some $p_i \quad P$.

Judgements of local closeness make sense even on very small domains *D*, while global closeness develops from local closeness when more and more objects of comparison are encountered. Explorations lead from local to global similarity judgements; and better-developed cases of global closeness will often come along with the claim that the domain *D* is also *comprehensive* . Once more, the epistemic background of the categorizing agent or community is singled out as a driving factor in the choice of similarity; we will refer to this aspect of knowledge as "domain knowledge of the speaker / community". I will however treat domain knowledge as part of, and not as distinct from, the general knowledge of the speaker communities under consideration — one reason being that a clearcut distinction between "general" and "domain" knowledge is difficult if not artificial. This should not hide the fact that knowledge about range and nature of objects to be classified is of prime influence in determining some reasonable similarity relation. The interests of the speakers will then lead them to focus certain specific domains.[16]

(10) Knowing or specifying the full domain *D* often means, knowing the **relevant dimensions of variation**.
Example: If we consider some core dogs *p* in a domain *D* which also contains non-dogs, interbreeding (yes/no) will be a dimension of variation. If we restrict our interest to a domain *D'* which only contains other dogs, the resulting similarity will have to be based on shape and colour criteria.

(11) Specifying the full domain means, knowing the **maximally possible distances**:
If we think about "colours similar to the red of a ripe tomato" within the full colour spectrum, rose red, burgundy and English red will still count as similar; if we think about "colours similar to the red of a ripe tomato" within the shades of red only, we will only accept reds of almost full brightness around the carmesine shade.

16 I thank Remko Scha for clarifying discussions about the impact of "knowing D" in the choice of the similarity relation. The dog example is his.

While domain knowledge is based on the *known* objects in the real world, we make crucial use of the assumption that similarity judgements can also be made for counterfactual objects. (Our definition of word intensions on the basis of samples and similarities rests on this assumption.). It might be of interest to characterize those cases where several plausible similarity relations on a known domain can compete. Before doing that, let me briefly relate categorisation by similarity to the mathematically more respectable partitioning of a domain by an equivalence relation. Given a global notion of similarity, ideally the following classes form a partition of D:

(12) $[\ p_i\ ] := \{\ x \mid d(p_i,x) < d(p_j,x)$ for all $j \neq i\ \}$

Yet, according to our definitions it might occur that some object $a \quad D$ has equal distance to two prototypes: $d(a,p_i) = d(a,p_j)$. Following definition (12), this object a will fall into neither $[\ p_i\ ]$ nor $[\ p_j\ ]$. There is some truth in this mathematical dilemma, though, as human categorisation seems to count on there being "gaps" between the classes around core exemplars.

If humans tentatively categorize domain D on the basis of d and $\{\ p_1,\ p_2,\ \ldots\ \}$, advance in knowledge can reveal that the “gap expectation” fails to hold true. We might discover hybrid objects a halfway between two prototypes. Such discoveries will lead to uncertainty about the proper classification of a, the adoption of further dimensions which allow a justified classification of a as "being more like p_i" or "more like p_j", or a complete re-investigation of the domain D in question. In such cases, it will also depend on the "artificialness" of the hybrid object (—like, being malevolently confronted with something exactly between cup and bowl—) whether the further classification criteria are adopted permanently, or just occasionally.

If the gap expectation is fulfilled, however, then the classes $[\ p_i\ ]$ above give us a partition of D, with the equivalence relation $x \approx y$ iff the closest prototype to x is the same as the closest one to y.

We can now talk about the cases where two categorizations can compete.

(13) Let D be a domain of (real or counterfactual) objects
Let D_o be the known subpart of D (i.e. known to some speaker community, at some time).
Let moreover P', P'' be two different choices of prototypes in D_o, and d', d'' distance measures on D, such that the categorizations $(D,\ d',\ P')$ and $(D,\ d'',P'')$ are distinct.
We say that (D,d',P') and (D,d'',P'') *compete* on D_o iff both give rise to partitions of D_o, and the partitions

$\{ \{ x \quad D_o | \, d'(x,p_i') < d'(x,p_k') \text{ for all } k \neq i \} \mid p_i' \quad P' \}$ and

$\{ \{ x \quad D_o | \, d''(x,p_i'') < d''(x,p_k'') \text{ for all } k \neq i \} \mid p_i'' \quad P'' \}$ coincide.

We are often confronted with competing categorizations. While it is the business of science to find arguments in favour of one or the other of these competing classifications, I will assume for the purposes of natural language semantics that some specific one of these, the "simplest" one or perhaps only the one which finds most support in the group of those who are experts in the relevant domain, is the one that underlies "the" similarity relation used for that purpose.

This offers us a potential link to integrate scientific progress. If we can increase D_o so as to contain one of the crucial objects h on which competing classifications do not coincide, the nature of *h* can often be turned into an argument in favour or against one or the other classification. But also if the political climate in the expert group changes, or if experts change their view of what is judged to be the most elegant or simple way to look at things, they might turn down one of several competing classifications in favour of another one. All these processes have in common that not much changes for the common speaker.

Evidently, we could complement the notion of "competing" by "almost competing" in order to characterize those constellations where two ways to categorize D *do* differ on known objects, but only on objects of marginal status.

These definitions and observations not only serve to make my claim, namely that a context of word introduction (or later even: word use) highlights one specific similarity relation, more plausible. They also should offer the basis to resolve the tension between scientific progress and semantic change. (We will come back to this question.)

Summarizing, I assume that, in a context of word introduction or, sometimes, word use, there is

- a salient transworld similarity relation *sim*
- based on a notion of distance *d*
- which is of either of the two forms given in (9)
- which is therefore symmetric
- which is unique (modulo a social component of expert taste & politics)

1.3.3. Some linguistic examples

In addition to experimental work in psychology and cognitive science, the synchronic semanticist can offer a range of similarity respects which are operant in determining the extension and intension of given words. Cross-linguistic studies in categorization have revealed that speakers' judgements about similarity of two objects on the one hand, and their inclination to call these two objects by the same name on the other hand can differ considerably (see Malt et. al. (1999)). If we subtract instances of evident polysemy, the remaining examples show that a word's extension (and thus meaning) is often determined by more intricate, historically grown notions of "being similar" than those run-off-the-mill similarities that prevail in the literature. The following list of similarity relations that are operant in determining the meaning of words is at best exemplaric; its purpose for the main course of my argument is solely to increase the reader's imagination as to in what ways things can be similar.

Shape is probably the most prominent respect in which an initial sample can be generalised. Checking our lexicon, we find that shape is operant most purely in the meaning of words like "triangle" or "circle" (in their everyday sense where also a somewhat shaky triangle counts as a triangle). These are clearly based on the idea of "looking like a prototype triangle" and "looking like a prototype circle".

More concepts rely on shape. In order to decide that the person who entered my office yesterday and the person who comes in today are one and the same individual, I will primarily rely on the fact that the two look quite similar (although the person might have been at the hairdresser's in the meantime).

Shape also is primary respect used by children in categorisation and language acquisition (see Keil (1989), (1994) for an overview over the shape/knowledge discussion in the literature on first language acquisition).[17]

In artificial language learning situations as those arranged in psycholinguistic experiments, both adults and children will use shape as the default similarity criterion in absence of any other clues (see Landau et al. (1998)).

The semanticist and lexicographer might note that all words which refer to visible things have a reading that can be paraphrased as "looking like a core instance of …" Thus, "apple" can apply to a plastic fruit that looks like an apple and a functionally useless rubber hammer is nevertheless a "hammer". Unlike "circle" and "triangle", these words have another prominent reading, resting on functional, biological, physical, and other criteria. This might obscure the fact that the shape based reading is another respectable reading of the word.[18] Yet, if we organise meaning in terms of samples and

[17] Luzie (2;4) consistently includes wolves and foxes into the extension of "dog", thus proving quite a good biological eye.

[18] This is sometimes overlooked by those who believe that words usually have only one reading. To name a case in question: I think that the discussion in Bloom (1996) in part suffers from that supposition. He asks subjects to tell him the extensions of function-based words like "boat" and

generalisations, and if we acknowledge the fact that shape is something like a default generalization, the shape/ρ polysemy (for any other similarity ρ) is no longer a surprise.

Shape readings can finally become operant when old words are "recycled" in creating new names for objects or kinds. The mountain *Matterhorn* got its name on basis of the word "horn" in its shape reading ("looks like a horn"), not on the basis of its biological reading ("part of an animal, usually somewhere on the head").

Function is often used as *the* alternative respect in which things can resemble each other (this bipartition is exemplified once more by the experimental design in Landau et al. (1998)). Note, however, that unspecified function will often be too broad for our purposes. Artefacts are usually built to fulfil a *specific* function. Usually we will take these more specific functions-for... to stand for single similarity respects under the header of "function". A word like "tin-opener" exemplifies this kind of similarity. Tin-openers are created for the purpose of opening tins (while leaving the contents unspoilt), and everything which resembles core tin-openers in that respect will qualify as a tin-opener, no matter what it looks like. An object which resembles (my) core tin-openers in the functional respect of "being an instrument with which one could severely hurt enemies" will *not* qualify as a tin-opener (wrong function). An object which looks like a core tin-opener but does not work like one will only be a "tin-opener" in the shape sense of the word.

Yet, things can be even more subtle. Words for traditional tools like "hammer" do not depend on function alone. At least I would hesitate to call *anything* which can draw nails into wood a "hammer", no matter how it works. In fact, we have learned that core hammers are also of use for many other purposes (like, opening nuts, flattening bulges on my car, ...) which other devices might not fulfil. We have to face the fact that a certain amount of *deliberate definition* can play a role in determining whether something *x* resembles something *R* sufficiently to call them by the same name. In the case of artefacts, the inventor or creator has, to a certain extent, the right to decide in what class the new artefact should fit. This is especially so for innovations or other borderline cases. The fact that the trackball of my laptop is *not* called a "mouse" although it exactly fulfils the function of the mouse offers an example. The decision will often depend on marketing strategies— calling your new trackball a "mouse", you suggest that you have merely improved the old "mouse" somewhat. Calling it a "trackball", you stress the fact that you have invented a device that may fulfil the old function, but is radically different from the old tool in all other respects. (Guess what would be the better marketing strategy.) Such deliberate instances of classification are also discussed in Bloom (1996), who moreover

---

acknowledges, surprised, that people are willing to accept something as a boat that *looks* like one, although it might be functionally useless.

includes a discussion of pieces of art. Bloom concludes that the creator of an object, and especially the artist, is extremely free in relating his creations to existing samples. [19]

Definitions in terms of *necessary and sufficient criteria* can be modeled by making use of the trivial similarities which were briefly mentioned in the previous section in (5).

Names for certain dishes, or generally foods, often are assigned depending on *mode of preparation* rather than similarities in taste alone. The right kind of ingredients (artificial vs. naturally grown) can also be crucial.

It might also be worth stressing that explicit yes/no criteria and implicit similarities can combine. If we assume, for example, that "red wine" is every alcoholic beverage that has been produced on the basis of red grapes, following certain procedures more or less closely, then "Chianti" is that kind of beverage which is moreover explicitely characterized as "grapes grown within a fixed, welldefined territory. We find many such names where *provenience* is a restricting factor.

1.4. Context as a further parameter

In section 1.2 I proposed that the context (of baptising, and sometimes of use of a word ω) will suggest some similarity relation. This similarity will then determine the class of objects that are represented by the samples. Now that I have given a more specific idea of what kinds of similarity will play a role, we can also specify in some more detail the contextual factors that determine this choice.

If someone introduces, or uses, some word ω in a context *c*, then *c* will determine the following aspects (the last aspect being optional):

(i) When does *c* take place? — *Time* of *c*.
(ii) Who were speaker and addressee in *c*? — Speaker and hearer, and in a broader sense: *language community*
(iii) What was the overall aim in the conversation? — *Interests* of speaker and hearer in *c*.
(iv) What was the conversation about? — *Referent* of ω in *c*.

[19] I do not agree with the conclusion of Bloom that no other criteria apart from the creator's intension are necessary in order to classify artefacts. Bloom's allegedly criteria-free account hides the common criteria for classification at the point where the community decides: Is an intension of a creator still rational - "yes, this might still qualify as a boat" - or already lunatic - "anyone who dares to call this thing a boat must be crazy".

Indirectly, time and speaker community of *c* will determine the *epistemic state* of the speaker community at time *t* (= "what they know"). Knowledge and interests are two major independent factors that determine the choice of some similarity relation, as we have already discussed at some length in the previous sections. (Domain knowledge will be treated as part of general knowledge rather than forming an own parameter; further adjustments are possible at that place.)

Two contexts c_1 and c_2 in which speakers with the same interests talk about some lump of matter *R* can differ in the knowledge available to these speakers, thus leading to different similarity respects ρ_1 and ρ_2. This theoretical assumption is confirmed by two kinds of empirical evidence, coming from language acquisition and the history of science.

Keil has conducted a great number of experiments which attest the development of children's categorization criteria in accord with their growing knowledge about the world. He reports, for instance, that small children will accept a racoon that has been changed in shape by painting, shaving etc. as a skunk, thereby proving that they categorize animals mainly according to shape. Older children, on the other hand, insisted that the animal before and after the treatment remained a racoon, paying more attention to criteria like origin, kinds of parents of animal etc. Moreover, such shifts typically concern entire domains. They are based on general knowledge (in that example: about animals) rather than specific knowledge (about racoons). While there is an ongoing debate in language acquisition research whether the infant is born with literally no knowledge, or whether even at level zero is equipped with a minimal folk physics and psychology, the impact of growing knowledge on categorization behaviour is undeniable.

In the history of science, one might want to take the development of biological taxonomy as an example of knowledge-dependent categorization. Animal categorization was traditionally based on distincitve shape, and only secondarily concerned with cross-breeding or, more recently, genetic closeness. The discovery of "twin species" was the final blow against this criterion: Two groups of animals which look exactly alike, but which are not able to have common offspring. Since then, the ability to interbreed is considered a necessary criterion for specieshood (though not a sufficient one; see Mayr (1942), Senglaub (1978)). A general survey of taxonomic criteria in biology is offered in Rheinberger (1983).

The opposite case consists in a pair of contexts c_1 and c_2 which take place at the same time, and within the same speaker community, but with different interests. In one of the situations, for example, speakers might be interested in the biological structure of a bird's

feather while in another situation, they have just discovered that the feather can be used nicely to scratch ink onto paper. Different interests will result in different ways to generalise the sample.

This assumption finds empirical support from recent literature in psycholinguistics, specifically the paper by Landau et al. (1998). Subjects were shown fantasy objects and told a name for "that kind of thing". Without further instructions, subjects tended to apply the name to further objects of similar shape. However, when the same phantasy objects were introduced in another context c_2 by pointing out some specific task one could fulfil with this object, speakers were more likely to apply the name to further things which also suited that specific task (independent of shape) and rejected objects of similar shape which were unsuitable (due to unsuitable material, for example). Landau et al. were primarily interested in *differences* in categorisation behaviour between adult speakers and children. They report that children up to age 3 tend to stick to the shape-based classification, while adults switched to function based categorisation more easily. As I am interested in long term developments of *adult* language rather than children, this does not affect my main point.

A more sophisticated case of different interests which lead to different categorizations is offered in Kitcher (1984). He proposes that different notions of “species” are supported by different taxonomic interests, namely those of evolutionary vs. synchronic biology.

Let me finally comment on the difference between speakers and speaker community. Evidently, we do not want to make word meanings dependent on the knowledge of arbitrary individual speakers. The idea of some linguistic "division of labour" was already introduced in Putnam (1975), and has been accepted without much objection. In that sense, single speakers in some given context *c* may not even themselves be fully aware of the kind of similarity relation in play.[20] We might speculate that, at least in the case of baptising, there could be some sociologically directed correllation between who are considered to be experts in certain questions, and who are authorised to introduce certain words into the language. Only the biologist may establish the name of a new species, and only the expert in chemistry or pharmacy may name a new substance.

In the third part of the paper, however, we will elaborate the idea that also non-baptising contexts *c* where a word ω is used may be relevant for the further meaning of that word ω. All speakers are experts in language: They can counterfactually consider „what would be the qua answer if the meaning of word ω was *established* by our present context of use?“. In such contexts at latest, we have to face the possibility that speakers

20 Note that once more the theme of partial semantic knowledge turns up (see footnote 4).

and hearers might be laymen in the relevant subject. But they still are part of a speaker community and in that sense participate in expert knowledge.

In section 3.2, the functional dependencies between contexts, times, interests, speakers and knowledge will receive a formal treatment and get integrated into a diachronic dimension. Before doing that, I want to clarify the position of my proposal as a link between a certain type of conceptual semantics and truth value based semantics. Part three is self-contained, however, and the reader who wants to pass on to the tradition of meaning from this point is free to do so.

2. The grasp of meaning

2.1. How can intensions be in the head?

I have proposed a way in which intensions are determined by more basic ingredients, namely referents and ways how to generalize them. In doing this, I claim, we can integrate insights about word meaning from cognitive semantics into traditional truth value semantics.

It has always been clear that no speaker can literally hold "intensions", functions from an infinite set of possible worlds into some possibly also infinite extension, in her head as an explicit list.[21] The position adopted in much semantic literature is thus that intensions are to model a certain *ability* of speakers, namely the ability to decide for an arbitrary object whether it should go into the extension of some given word (under the present circumstances) or not. For example, Chierchia & McConnell-Ginet (1990:74) write: "This shows how our procedure (i.e. a simple version of truth value based semantics, R.E.) can be regarded as an abstract representation of our capacity of pairing sentences with the situations that they describe (...)".[22]

Is it enough for semantical purposes to model just and only this (idealized) ability of speakers? In spite of the impressive amount of theory that is based on this simplified assumption, some would enthusiastically object. Truth value semantics has remained, programmatically, agnostic with respect to the question *how* this ability is organised in

[21] The relation between semantic idealization and psychological reality is discussed in Partee (1979). With respect to the finiteness of the brain, she writes: “For one thing, you don’t need to represent all of the possible worlds distinctly in order to know a function that has them as a domain. We know the function for adding arbitrary real numbers without being able to represent all the real numbers distinctly.” (Partee 1979:3)

[22] The case can be compared to addition of integers. Mathematics is usually studied without reference to *how* the addition table is implemented in our head. And mathematics that would account for the fact that humans can not, for practical reasons, add two numbers once they become very very very large, would not be "better" mathematics.

our heads, what bits of knowledge evidence that I, for example, master the meaning of the word “dog”, and which kinds of functions from possible world into extensions are such that humans can grasp them at all.

One possible, and widespread, answer is that I am in command of a “dog-concept” which is linked to the word “dog”. Yet, giving a name to the thing in my head does not, in and off itself, tell me what it is.[23] Crucially, it does not tell me how to find out whether two speakers have the same “dog-concept” or different ones. We might spell out “concept” as “mental image” or “stereotype description”. However, the mental images or stereotype descriptions of “dog” of any two speakers can differ while they still mean the same thing by “dog”. The only safe test for concept identity is the extension test: The implementations of the meaning of "dog" of two speakers coincide only if they produce the same extensions and (as far as that can be tested) intensions. In doing this, speakers may express uncertainty or they may rely on experts, but whenever one of them insists that something is a dog where the other one is sure that it isn't, we'll have to conclude that they don't mean the same thing by "dog". We are back at the intensions of truth value based semantics.

Scholars in cognitive semantics usually avoid to acknowledge this fact. For example, Jackendoff writes in (1983:238) that "we have observed on many occasions that truth, purportedly a relationship between language and reality, has little relevance to the nature of linguistic and cognitive judgements, if it can be defined at all." Consequently, he never, in that book, addresses the question of how we can make sure that two different speakers actually have acquired or master the same concept.

The account proposed here keeps intensions as word meanings, and retains the knowledge about a word’s combinatoric behaviour in semantic composition as part of its meaning. Its main proposition is that these meanings are based on more primitive ingredients: Core referents and similarity relations. It can realistically be claimed that speakers master both.

We, as little as others, can explain “meanings” in terms of some natural science. Previous truth value semantics assumed word meanings as given, we have to assume that speakers can make similarity judgements, and take these as primitives in our account. Yet, breaking up “meanings” in this way and understanding the space of possible similarity relations will allow us to tackle semantic questions which remained outside focus in traditional truth value based accounts:

---

23 The proposal by Zwarts and Verkuyl (1994) for an algebraic reconstruction of Jackendoff’s concepts, to my eye, highlights this point. While their aim is, to show that conceptual structures are not beyond formal treatment, they thereby reveal that the resulting formalism looks very much like higher order logic on a sorted ontology, plus constraints on concept combination (i.e. conjunction).

– the acquisition of word meanings through ostensive presentation of core exemplars and an increasing understanding of the similarities in question
– the grasp of *some* meaning by individual speakers
– a notion of “closeness” and “remoteness” of two meanings
– an account for changes in meaning

The resulting meaning is truth-value based in the traditional sense. However, it is computed on the basis of human cognitive abilities. Humans, not reality, provide the notion of similarity. And clearly, the samples as well as all further objects in the extension are filtered through our perception. Once more, the question is not whether some *x* and *R* are ρ-similar in any absolute sense (whatever that might mean) but whether *we* would perceive them as similar.[24] This perspective coincides with the position of Jackendoff (1998) where he stresses the compatibility of cognitive semantics and intensional semantics. While my focus on diachronic matters has led me to refine the intensional structure of meaning in other ways than those followed by Jackendoff, our construal of model-theoretic semantics is the same.

Concerning the details, my proposal evidently owes to recent work of Gärdenfors (see references); while he focusses on the cognitive abilities in categorization in general, I have offered a first implementation of his ideas into a semantic theory.

2.2. Degrees of ignorance

How does the knowledge of a single speaker, or the knowledge offered by a traditional lexicon relate to the account of word meaning developed here? A nice aspect about the approach is that we can easily locate what has elsewhere been called "stereotype knowledge" or "default knowledge": This is primarily knowledge about the core *samples* for a certain word. According to our view, this does not amount to the idea that the set of samples already is the full extension of the word in question, or that the descriptions amount to a definition of the word in terms of necessary and sufficient conditions. Clearly, speakers will understand that these samples will have to be generalised.

Entries in a traditional lexicon might also provide some indication with respect to the similarity relation according to which these samples are to be extended to a full class. Certain aspects of this generalisation, however, might be implicit knowledge and rarely found in a circumscription of a word's meaning. One of the advantages of breaking down speakers’ knowledge about the meaning of a word into knwoedlge about core referents,

[24] In order to prevent misunderstandings, let me hasten to add that we are in command of quite sophisticated modes of perception — using further technology, we can for instance distinguish an apple that emmitts radiation from one which doesn't, although the "look the same". Still we are in principle unable to access the *Ding an sich.*

and knowledge about ways how to generalize them, is that implicit and explicit knowledge is separated rather than merged into one untractable “ability to determine actual and counterfactual extensions of the word”. Our picture corresponds to the division into “what you have to tell someone about the meaning of a new word” and “what you can leave to them” in a clearer way than the traditional intension.

Moreover, we can mirror the fact that speakers might not be in full command of the meaning of a word, that they only have *partial* knowledge. There are varying degrees of ignorance.

Take the word "fox", for example. I know enough about foxes to be able to describe, and maybe even find, a typical exemplar of a fox. I also know enough about biology and English to know that the similarity relation in question should be "similar in the sense of being of the same species". However, I must leave it to experts to explicate the criteria that apply in the case of foxes. In other words: I can not, for an arbitrary animal, determine whether it is a fox or not. Nevertheless, this kind of knowledge seems sufficient to agree that I know the meaning of the English word "fox".

However, I also know words for which I know typical exemplars, but have no idea as to how these should be generalised. To reveal my ignorance of English: I know what a typical instance of "cattle" looks like, but I am uncertain whether these should be generalised to the full species, or whether the boundaries are drawn on the basis of functional aspects (like, being owned and exploited by a farmer). In this case, it is questionable whether the speaker (= me) still knows enough to know the meaning of the respective word (= cattle). Yet, that speaker (once again, me) can fight her way through a good deal of conversation without this lack of knowledge becoming apparent.

The picture also explains why even speakers who don’t know how to generalize the core exemplars of a category will be able to use the respective word ω felicitously. If we assume that referents of ω are more frequently normal ω’s than not, then I, as a hearer, can follow any conversation about ω’s by assuming that the referents in question will look like my stereotype. Most times, I will be right. But, if I am wrong, I am not surprised either — being aware that my lexical knowledge about ω was underspecified in the sense of a missing respect of generalization.[25]

Many interesting questions arise at that point. How many speakers need to have knowledge of the first kind about a word such that we would count it as an item of the language under discussion? Will a word shift its meaning(s) more easily at a time when only few speakers of the language can master it in the full sense? (Which is most likely.)

---

25 This is, of course, at the core of *any* theory of lexical knowledge, be it Putnam’s stereotypes, Fillmore’s frame theory or whatever. The new point in my story is that stereotypes fall out as part of the overall picture, rather than - like e.g. in the writings of Putnam - running alongside with “official” semantics as a kind of “auxiliary semantics for real speakers”.

Do vague category boundaries depend on the fact that different speakers have different knowledge (or rather believes) about the meaning of words?
Questions like these have, however, to be left for future research.

3. The tradition of meaning

In section one, I proposed a modified version of the causal theory of reference. It offers a conceptually meaningful way to construct (truth value based) extensions and intensions by generalising initial *samples* according to a salient *similarity relation.*

Section two was devoted to the task of locating the account at the borderline between truth value based and cognitive semantics. Meanings, although being computed on the basis of cognitive operations, are still couched in a truth value based framework. Thereby we combine some of the explanative potential of cognitive semantics with the structural potential of truth value based semantics.

The present section is devoted to the core issue of the paper, a diachronic theory of meaning. In order to be able to talk about word histories, we will first introduce a further temporal parameter into the representation of words in order to distinguish different language stages (3.1.) . Next, we will introduce some notation in order to be able to talk about the uses of a word ω at stage *t* (3.2.). In 3.3., I will elaborate the idea that the uses of word ω at time *t* determine the spectrum of possible readings at time *t*+1. Section 3.4. will finally review the possibilities for meaning changes which are implicit in the approach and will offer examples of actual word histories which instantiate these patterns.

3.1. A temporal parameter

We need to distinguish the meaning of a word ω at time *t* from the meaning of the same word ω at time *t'*. We will make the idealising assumption that time proceeds in discrete steps. Eventually, we want to determine the meaning of word ω at time *t*+1 from its meaning at time *t*, the exemplars that were usually referred to by ω at time *t*, plus the circumstances under which word ω was predominantly used at time *t* (where „predominantly" can mean both „predominating in quantity", or „by socially dominating speaker groups"). However, we are not as far yet, and will use this section in order to clarify the formal nature of the temporal index.

Our subject of investigation are words-in-time and their meanings. I will base my considerations on synchronic semantic models **M** in the Montagovian tradition where meanings are objects in models of type theory (Montague/Thomason (1974)). These models will have to be augmented by several ingredients which will be introduced in turn.

All models **M** will contain a set T of points of time which carry a discrete linear order. If t T, I will use t+1 to refer to the point in time which immediately follows t. The trans-temporal meaning of word ω is a function from T into an appropriate synchronic meaning of ω.

(14) If ω is an expression in *L*, then the diachronic meaning of ω relative to some model **M** is $[[\omega]]^{\mathbf{M}}$ = a function, mapping T → [D_s → D_α] where α is a logical type matching the syntactic category of ω
$[[\omega]]^{\mathbf{M},t}$ denotes the synchronic meaning of ω at time *t*. We have
$[[\omega]]^{\mathbf{M},t} = [[\omega]]^{\mathbf{M}} (t)$.

In order to make matters easier to read, I will from now on focus on the case of a *proper noun* ω which would synchronically be interpreted in $D_{(s,(e,t))}$.[26]

(15) If ω is a proper noun, then the diachronic meaning of ω relative to some model **M**, $[[\omega]]^{\mathbf{M}}$, is defined as follows:
$[[\omega]]^{\mathbf{M}}$ = a function, mapping T → [D_s → $D_{(e,t)}$]
$[[\omega]]^{\mathbf{M},t}$ is interpreted as $[[\omega]]^{\mathbf{M}}$ applied to time *t*.

Note that $[[\omega]]^{\mathbf{M},t}$ is not the same as „the things we'd call ω at time *t*" in the sense of temporal logic. If we take the noun *meat*, then $[[meat]]^{1000aC}$ is the meaning of *meat* in the sense of 1000 a.C. (roughly, "food" in the modE sense). The extension of modern *meat* at t=1000 a.C. in the sense of temporal logic, however, means „everything in 1000 a.C. which would qualify as a *meat* in the modern sense of the word".

3.2. Yesterday's uses of a word

We will now introduce the formal ingredients which allow us to capture yesterday's contexts of use of a word ω, or at least all those aspects of such contexts which might become relevant in the further (semantic) life of the word. Let therefore

(16) C = { c_1, c_2, c_3, …} be a set of contexts
SC = { s, s', s'', ... } be a set of speaker communities
INT = { int_1, int_2, int_3, ... } be a set of interests

Moreover, assume the following family of functions:

---

26 I'd like to stress that this does *not* mean that the theory is in principle restricted to the case of proper nouns. I adopt it purely for expository purposes. The range of the full account will be discussed at the end of the paper.

(17)

time: C → T
int: C → INT
sc: C → SC
ep: T x SC → P(W)
word: C → word ω that was used in *c*
qua: C → ρ where for each *c*, *qua*(*c*) is a similarity relation on the domain that is appropriate for *word*(*c*)
ref: C → D_α where for each *c* C, *ref*(*c*) is the referent of *word*(*c*) in *c*.

The first three functions, *time, int,* and *sc* will capture the fact that each context of use takes place at some specific time, with specific interests in the objects talked about, and as part of a specific speaker community. The function *ep* will give us for each time and speaker community the epistemic state of that community at the time. This will be used presently in order to capture the idea that different knowledge states can result in different categorizations of the same domains.[27]

As we want to trace more than the history of only one word, we will have to specify the word ω that was used in context c (or for which context c is of interest). With these specifications available, we can now proceed to the functions *ref* and *qua* which finally give us the referent of ω in context *c* and the qua answer that is rendered salient by context *c*.[28] The latter function will require some comment.

In view of what was said about the *qua answer* in section one, it might be surprising to find that *qua* should not only be defined for baptising contexts but arbitrary contexts. Can any context provide a qua answer? Yes, and no. I assume that the linguistic abilities of speakers allow them to consider the following counterfactual question:

> "I am in a context *c* of *use* of a word ω. What would happen if this context *c* was actually a *baptising* context (i.e. a context of introduction for word ω)?"

Sometimes the interests and knowledge available in context *c* will lead to the similarity relation ρ which belongs to the meaning of word ω anyway. Sometimes, however, the interests of the interlocuters, or the advanced knowledge of the community might be such that *c* supports a *different* qua answer than the one that would actually lead to the traditional meaning of ω. Of course we don't want to say that the meaning of word ω

---

27 In fact, we could omitt the parameter "speaker community" altogether, as we will never look at cases where, given the same interest and knowledge, different speaker communities will come to different word meanings. I keep SC for the sake of concreteness, as it might strike the reader as too abstract to see entire language communities vanish into epistemic states.

28 In case it turns out that contexts should sensibly comprise the utterance of more than only one word, one might instead turn *ref* and *qua* into binary functions from contexts plus words into objects and similarities, respectively.

changes immediately under such circumstances. Yet, if something like that will happen frequently, it can enrich the spectrum of possible meanings of word ω.

Should *qua* be a total function on C? Realistically thinking, it is rather optimistic to assume that literally *all* contexts of use of a word will *really* suggest some respect of generalization in a substantial way. There might be many occasions where some word is used in a fairly neutral way without any specific interests in the referent. We might account for this fact by assuming that *qua* is a partial function. Alternatively, we can also assume that in such contexts of use, $qua(c)$ is the similarity relation that belongs to the meaning of $word(c)$. I will make this assumption, as it will lateron automatically allow us to explain the tradition of the actual, "true" meaning of a word ω from time t to $t+1$.
In order to account for possible nonspecific uses of word ω, I will also assume that *ref* is a partial function.

We can formulate some restrictions to the functions above. Restriction (R1) ensures that if the same word is used in two contexts c_1 and c_2 at the same time, by the same community, with the same interests and with reference to the same object, then the qua answer rendered by these contexts should also be the same:

(R1) If two contexts c_1 and c_2 with $word(c_1) = word(c_2)$ are such that
$time(c_1) = time(c_2)$, $sc(c_1) = sc(c_2)$, $int(c_1) = int(c_2)$, $ref(c_1) = ref(c_2)$
then $qua(c_1) = qua(c_2)$[29]

The second restriction will express that speaker community and time are only relevant insofar as they give us an epistemic state:

(R2) If two contexts c_1 and c_2 with $word(c_1) = word(c_2)$ are such that
$ep(time(c_1), sc(c_1)) = ep(time(c_2), sc(c_2))$
and $int(c_1) = int(c_2)$, $ref(c_1) = ref(c_2)$
then $qua(c_1) = qua(c_2)$

Let us now put the formalism to use in order to talk about the use of word ω at time t:

(18) $C(\omega,t) := \{ c \in C \mid time(c)=t \ \& \ word(c)=\omega \}$

[29] Could we assume that *qua* directly operates on tuples <w,t,sc,int,r> of a word, a time, a speaker community, interests and referents? This would lead to difficulties because we want to be able to count occurrences of word uses. The above tuples will not allow to distinguish co-temporal uses of a word, under the same interests, in the same community, and with the same referent. If, for example, the moderator of a TV show used a word in a new sense, these tuples would not distinguish whether the show was watched by three or three million people. Evidently, we should be able to represent such frequencies.

More interestingly, all c in C(w,t) will lead to a referent and a similarity relation. We can collect all contexts of use at time t which render the same qua answer:

(19) $C(\omega,t,\rho_i) = \{ c \in C \mid word(c)=\omega \ \& \ time(c)=t \ \& \ qua(c)=\rho_i \}$

Now we are ready for the final step, namely turning these static descriptions of language stages at some time t into a moving diachronic picture.

3.3. Today's meaning of word ω is determined by yesterday's uses

Throughout this section I will use the time variable t to stand for "yesterday" and t+1 for "today". Let furthermore ω be the word which we want to describe in time (and, to keep matters simpler, we will assume that ω was introduced into the language before time t). We can now consider the set C(t) of all contexts c at time t in which ω was used.

(20) $C(\omega,t) := \{ c \in C \mid time(c)=t \ \& \ word(c)=\omega \}$

This set of contexts can be subdivided into equivalence classes by collecting all those contexts c which have the same value $sim(c)$. Let for all c in C(ω,t)

(21) $[\ c\ ]_{sim} := \{ c' \mid c' \in C(\omega,t) \ \& \ sim(c')=sim(c) \}$

Next, we can collect all the objects that were referred to in such contexts of use. For each equivalence class $[\ c\ ]_{sim}$, $ref([\ c\ ]_{sim})$ is defined as follows (where w_o is the actual world):

(22) $ref([\ c\ ]_{sim}) := \{ x \in w_o \mid$ $c' \in [\ c\ ]_{sim}$ such that $ref(c')=x \}$

These actual referents of ω in the contexts collected in $[\ c\ ]_{sim}$ will be the pool of samples to be generalized according to the common similarity relation of all contexts in $[\ c\ ]_{sim}$. Realistically, one can assume that the actual samples only comprise a subset of $ref([\ c\ ]_{sim})$, namely the most common referents in such contexts of use.

(23) $samples(c) :=$ $\{ x \in ref([\ c\ ]_{sim}) \mid$ "x is one of the more common exemplars in $ref([\ c\ ]_{sim})$"$\}$

I will now define the range of potential readings of word ω at time t+1:

(24) f is a potential reading of word ω at time t+1 iff
$f: w \rightarrow \{ x \mid c \ \text{y} (c \ \text{C}(\omega, t) \ \& \ y \ \ samples(c) \ \& \ sim(c) = \rho$
$\& \ x \ \rho$-similar to y in $w \}$

Which ones of these potential readings will become actual readings of word ω? Historical accidents will eventually draw the distinction, but we may speculate about the driving factors. One major criterion will certainly be the mere size of the set $[c]_{sim}$, the number of contexts which all support the same respect in which samples are to be generalized. If a large number of contexts support the same reading for ω, this reading is more likely to be actualized than if only one or two sparse contexts did.
Remember that we assumed that all "neutral" contexts support the similarity relation which belongs to the reading of ω at time t. In the likely case that a considerable number of contexts of use actually is more or less neutral, the actual meaning of ω at t has, for cardinal reasons alone, good chances to be passed to t+1.

It is not entirely clear how written uses of a word should count, in a literate society. As long as such written sources are still available, and as long as there still are *some* speakers to understand ω in the intended reading of that written source, such a meaning can survive even if the number of actual uses in conversation at time t is low.

The less the number of speakers who reliably know the meaning of word ω at time t, that is, who could name samples and master the appropriate way of generalization, the better are chances for other potential readings to get actualized. Once again, those readings which are supported by a greater number of contexts of use at time t will more likely get established than those with low frequency. Speakers who are not sure about the "true" meaning of ω will guess what the meaning might be on the basis of the contexts of use they know. Such speakers will act out the counterfactual question:

> "I am in a context c of *use* of a word ω. What would happen if this context c was actually a *baptising* context (i.e. a context of introduction for word ω)?"

Frequency, however, is not the only decisive factor. For example the social rank of the speaker in c is certainly another important factor. The prominent and fashionable leading speakers in a society are more likely to establish their view of the world in the lexicon of a given language than less prominent and influential groups of speakers. In times of modern mass media, even one single context of use may become influential.[30] Our

[30] Such examples reveal that we should eventually consider the number of recipients/hearers of an utterance rather than utterances themselves. If Bill Clinton says "hi" in front of a TV camera, millions of Americans might listen.

modelling so far does not reflect the speaker of a context, yet this aspect can be added in the evident way.

Of course, more than only one potential reading of ω can be realized at time t+1. New readings can become useful in addition to the old one(s); new readings which typically differ from the old reading only in subtle ways and have overlapping extensions with the old meaning. The cases of meaning multiplication that are at the core of our model are not as spectacular as metaphors and metonymies typically are. The account focusses on the unobtrusive cases of meaning change. Which leads us to the next section.

3.4. Room for change

There are two main places for variation inherent in the account I propose. The first source of variation, and the one which has gained most attention so far, is the choice of similarity relation belonging to word ω. The second place where variation can take place is the choice of samples which are to be generalized. In the previous section, we briefly introduced the idea that speakers will only use the most common referents of ω at time t to compute the meaning of ω at t+1. Formally, this was captured by the step from $ref([c]_{sim})$ to $sample(c)$. The shape of the most common referents of ω may change over time, thus leading to subtle changes in meaning. I will discuss the word history of the German word "Karussell" as a case of meaning change through changing samples, and the cases of German "Kreide", "Horn" and "Pferd" which illustrate meaning change through changing *qua* answer.

"Karussell"

The word "Karussell", which was taken over from French "carrousel" in the second half of the 17th century, originally denoted a form of tournament game where the task was to hit a ring with the lance while riding past it on a horse. In the 18th century, a variant of the original game became popular where the players had to hit the rings from a pivot mounting, sitting on wooden horses instead of real ones. While this variant was still, in a sense, the old game, it was open to a much wider variety of competitors.

This new "typical referent" of the word "Karussell" differed from the original game insofar as, while the participants' interest *still* consisted in "having fun", the kind of "fun" taken out of the game changed: The original "Karussells" were sportive games , the new version was already fun because one could travel in rounds on hobby horses.

It was reported to me that in Holland, at least until 30 years ago, caroussels had a bell attached at the roof. Those who hit that bell could win a free ride on the caroussel. This version still maintained an element of skillful competition.

In Germany today, the rings or bells to be hit have completely vanished from the "Karussell", and the horses now come in the company of little cars, swans, bikes, closed carriages and all other sorts of vehicles which would have been unsuitable for a "Karussell" in the sense of 1650.

"Kreide"

The word "Kreide" (= chalk) has two meanings in contemporary German. In one reading *Kreide$_1$*, it means "calcium carbonate" and is used as a name for a chemical substance. In a second reading *Kreide$_2$* it denotes a writing tool, typically the small handy piece of *Kreide$_1$* which can be used to write on blackboards. The two readings differ. Modern industry has created *Kreide$_2$* not only coming in different colours, but also made from other materials which resemble *Kreide$_1$* to a lesser or greater degree. The pieces of *Kreide$_2$* which deviate most from *Kreide$_1$* are "Malkreide" or "Wachskreide" (crayons) on the basis of wax (or, again, its modern substitutes) for kids to draw colourful paintings. It is easy to see how both readings are generalizations over the same core referents, small pieces of a soft white stone which was good for drawing and writing.

Actually, *Kreide$_1$* nowadays becomes fashionable in a second tool sense, namely as a white powder which, mixed with water, can be used to paint walls. Time will show if even a further reading *Kreide$_3$* will develop on this basis which has an extension distinct from both, *Kreide$_1$* and *Kreide$_2$*.

The case of "Kreide" is paralleled by many other words which have both a tool sense and a natural substance sense. Stern (1934:380ff.) discusses the example of English "horn" in its readings "part of an animal" and "musical instrument". While Stern offers basically the same intuitive explanation for this polysemy as I do, his notion of meaning is too internalistic. Stern assumes that the meaning of a word, roughly, consists of the knowledge the speaker links to the word. This not only leads to the puzzle how two different speakers can ever grasp the same meaning; it also forces Stern to diagnose meaning changes even for proper names while their bearers age and are linked up to more and more anecdotes and adventures.

"Pferd"

Many examples concern the fact that some natural substance or object becomes useful as a tool. However, there is also a nice example of a fairly technical "tool" term which eventually developed into a name for an animal species: The German word "Pferd" (= horse). The Latin root[31] of the word, "paraveredus", literally meant "side horse" and was a technical term in the Roman mailing system. While the main post lines were well

[31] Traditional etymologies stress the fact that *para-veredus* is composed of a greek prefix and a latin root. However, the language community in command clearly were the Latin-speaking Romans.

equipped with spare horses for faster delivery of mail and goods, on side lines the carriers had the right to require horses, the "paraveredi", for a change from every local person. We know from contemporary sources (for an overview see Hartman (1868)) that the demand for "paraveredi" constituted an enormous economical pressure at certain times and was all the less acceptable as the transport system as such was not open to the public.

It seems that the local language community got the impression that any reasonably strong horse qualified as a "paraveredus" (or, later: NHG *pfarifrit*, MHG *pherfrit*, *phärit*, *pfert*). At an intermediate stage, *Pferd* was still restricted to "any horse apart from the war-horse" (probably because Roman carriers would first visit the farmer in their search for a *paraveredus*). In the modern German sense of *Pferd*, the word denotes all members of the species *equus equus*. The word has thus lost all functional aspects and turned into a natural kind term.

For the sake of completeness, let me mention two more patterns of meaning variation supported by our model.

Firstly, it is evident that changing epistemic states, more optimistically called "the acquisition of knowledge", can support different qua-answers in face of the same samples and under the same interests. The refinements in biological, physical and chemical classification offer well-known examples. My account forces me to adopt the position that the meaning of words like "gold", "water" and even "dog" changed many times in history, as a consequence of scientific progress. I will comment on this position in section 4.2.

Secondly, we can develop a variant of the baptizing theory of section 1.2 to model the creation and use of *metaphor*. In order to establish a metaphorical sense for some given word ϖ in utterance context *c*

(a) it must be clear which individual or object (or state of affairs, or event) is to be the referent *R* of word ϖ in that context *c*
(b) it must be clear that *R* would not be in the extension of ϖ in its literal sense
(c) speaker and hearer must know the typical referents *S* of ϖ in its literal sense, and must know for which properties these referents are taken to be the most salient examples.

The metaphorical sense of ϖ is derived by generalizing all samples *R* and *S* on basis of a similarity relation ρ which, on the one hand, classifies *R* and *S* as similar, and on the other hand "has to do" with the properties for which *S* are especially typical. Metaphorical meanings, according to this view, are introduced exactly like ordinary meanings of words, except that the choice of similarity relation is not left to context alone, but is more

or less willingly driven by the speaker by putting forward an odd set of samples R and S. The procedure will be elaborated and exemplified in section 4.3.

The diachronic semantic account thus offers room for all kinds of patterns of change where old and new meaning have overlapping extensions, common samples, at some time. This allows us to capture cases of generalization, specialization, inductive generalization, and metaphor.[32] The account as I see it offers no link to metonymic processes.

4. Further issues

4.1. Relation to prototype semantics

One main ingredient of the account I propose is the derivation of a word's extension and intension starting from some core exemplars. The idea of radiant categories bears evident reference to prototype theory; the relation between the two theories has been discussed in more detail by Gärdenfors (1997b), (t.a.) from whom we borrow the idea of measure based similarities. Lambalgen and van der Does (1998) offer another comprehensive overview on formal approaches to prototype theory and acknowledge the work of Gärdenfors as one of the more promising attempts to make the classical insights of Rosch (e.g. Rosch (1975)) fruitful for formal semantics.

More specifically, aspects of prototype theory are operant in our approach at the place where *ref(* $[c]_{sim}$) is reduced to *sample*(*c*): Not all, but only the typical referents of a word ω in contexts *c* which all share the same similarity ρ are used as the basis for generalization at time *t*+1. The nature of typical referents may change over time, even if the word's extension remains constant. Formally speaking, we do *not* require (25) and (26) to hold true.

(25) *sample*(*c*) = *sample*(*c'*) where
time(*c*) +1 = *time*(*c'*) and *sim*(*c*)=*sim*(*c'*) (identity of samples)

(26) *sample*(*c*) *sample*(*c'*) where
time(*c*) +1 = *time*(*c'*) and *sim*(*c*)=*sim*(*c'*) (conservativity)

[32] I don't claim that *all* instances of generalization, etc. should sensibly be couched in the account. A word might have two readings with overlapping extensions for reasons totally different from those which can be described in terms of my account. I maintain that, in order to understand the creative aspects of language properly, we should be concerned with the *mechanisms* of change instead of the *patterns* of change.

However, in order to maintain a certain smoothness in meaning develop-ment (and the development of the world), something like (27) might often be true.

(27) $sample(c) \cap sample(c') \neq \emptyset$
where $time(c) + 1 = time(c')$ and $sim(c)=sim(c')$ (non saltat)

These assumptions are empirically motivated by examples like "Karussell" which show that changes in the nature of samples can eventually result in changes in a word's extension and intension.

Our approach allows for the case that a word's referents R_1 which occur in contexts which support similarity ρ_1 look considerably different from referents R_2 which occur in ρ_2-contexts. The two corresponding readings of the word will therefore be generalizations of different sets of samples, thus allowing for more flexibility.

A simpler version of the theory will result if we assume that all potential readings are generalized on the basis of one and the same set of samples for word ω at time *t.* Even in such a simpler version, one will want to admitt only the normal referents of ω as samples, thus once more drawing a distinction between typical core and more exceptional periphery. This simpler variant can easily be revealed as a more formal version of the prototype based account of diachronic semantics proposed in Geeraerts (1997).

Geeraerts observes that many patterns in meaning variation and change can best be explained in terms of prototype based organization of categories at all levels. Most importantly for our case, he shows that old words can acquire new readings by forming different categories around a constant set of core prototypic exemplars. Evidently, this is exactly what the simplified variant formalizes.

However, if we adopt that simpler variant of our account which would correspond to the picture that Geeraerts draws at the level of extensions and intensions, we will predict that all new readings arise on the basis of the same set of core referents. In reality, however, the referents which give rise to different new readings may also differ in characteristic ways. If, for example, the German “Kreide” was to develop a reading “powdery base material for wall paint”, the core samples of this generalization will be sufficiently different from the core “Kreide” piece-to-write-with to disallow pieces of coloured wax in its extension.

One difficulty in providing an example where such a pair of new readings developed at one and the same time lies in the fact that the respective readings are usually so old that it is almost impossible to be certain about the exact time and order in which they evolved. For the time being, I can only offer examples of one old and two or more

new readings where the new readings probably developed at *different* times. Formally, it would thus be possible to claim that the core exemplars at the time of birth of reading 1 simply looked different from the core exemplars at the time of birth of reading 2. Yet, such a claim would be implausible in all cases, and will become impossible altogether when two new readings evolve at the same time.

lat: tavola	=	1. board (of wood)
		2. picture
		3. blackboard
		4. table, written list
		5. table (piece of furniture)
lat.: pipa	=	1. reed
		2. pipe, tube
		3. pipe (instrument)
		4. pipe (smoking-)

Consider the example “pipa”. The core examples that gave rise to the instrument reading were presumably pieces of reed with the respective holes and mouthpiece which made them suitable to play music. Reed prepared in this way, however, will be completely unsuitable to support the function that underlies reading 2. “pipe, tube”. The presumable size of these pieces, plus the fact that they had holes drilled in, would have made them utterly unsuitable for all purposes where you’d use a “pipe, tube”.

4.2. Scientific progress

In section one, we dismissed the idea of rigid designation for natural kind (and other) terms in favour of a notion of meaning where extension and intension are computed on basis of core samples and similarity relations. More importantly, we perceive those similarities as relations which can be provided and grasped by human speakers. It is a natural consequence of these premisses that words like "gold" or "water" will have changed their meaning many times in history. This contrasts with the predictions of the classical account.

According to Putnam's position, the meaning of words like "water" and "gold" has remained constant all over time. All literature about gold or water offers different theories about the referents of one and the same word; most importantly, scientists of all times were actually talking the same language (at least with respect to the name of the thing they wanted to describe). If our account were right, in what sense do natural

scientists of today still investigate the same "gold" as those who studied Greek "chrysos" 2000 years ago?

Even if the intension of the word “gold” may have changed during the centuries, much has remained constant.

(a) The core samples of “gold” in ancient Greek are still acceptable core pieces of "gold" today.
(b) The interest that natural science took in these pieces has also remained unchanged. The interest of substance classification has not been replaced by, say, functional or religious interests.
(c) Even the extension of the word “gold” in our real world has remained constant. The criteria that were applied by Aristotle to distinguish true “gold” from false are still accepted as correct necessary and sufficient criteria in today's world.

In terms of intensions, however, the Greek “chrysos” and the English “gold” of 1999 differ. Counterfactual instances of “gold” that still seemed possible to Aristotle are no longer acceptable today, and we can imagine and accept “gold” which would not have been accessible to Aristotle.

To put it more abstractly: Meaning change through scientific progress is characterized by constant interests, conservativity on core referents, and changing knowledge. More specifically, the known domain of individuals D_o can increase, and the knowledge about the objects in, and around, the extension of the word can increase. The following more or less dramatic meaning changes are possible:

(28) Old and new epistemic states ep_1 and ep_2 at t_1 and t_2 lead to two different similarity relations ρ_1 and ρ_2.

(28a) Minimal change: ρ_1 and ρ_2 coincide on the old domain of known objects $D_o(t_1)$, and even on the new domain of known objects $D_o(t_2)$. Example: “gold”

(28b) Difference in larger extension: ρ_1 and ρ_2 coincide on $D_o(t_1)$ but differ on the larger domain known at the later time $D_o(t_2)$.
(Idealized) example: Biological criteria for animal classification before and after the discovery of twin species.

(28c) Difference in both extensions: ρ_1 and ρ_2 differ already on $D_o(t_1)$. Example: Exclusion of the "Blindschleiche" (blindworm) from the category of "Schlangen" (snakes) after the discovery of its inward similarity to "Eidechsen" (lizards).

If even the judgements on $D_o(t_1)$ are shaken, we may face the more dramatic case

(29) Old and new epistemic states ep_1 and ep_2 at t_1 and t_2 lead to different similarity relations ρ_1 and ρ_2, and to an exclusion of some of the old core referents: $R_2 \neq R_1$. Difference in extensions at all times are an automatic consequence of this case.

Even the dramatic case in (29) is attested in the history of science. Biological Taxonomy has adopted a quite sophisticated naming convention which is designed so as to handle such meaning changes: According to convention, the name of a species will always cover exactly those objects which are of "the same species (according to contemporary wisdom)" as a *fixed initial sample*, used in the introduction of the name (this has been described in Bolton (1996)). While this convention is in surprising accord with the causal theory of reference, its limits are obvious. Almost all initial samples are dead. According to contemporary biological classification, hwoever, the classification of animals into species depends on mating *preferences*, not only mating *possibilities*. It will be extremely difficult to test the mating preferences of a dead specimen, even with full knowledge of all its genetic material.[33]

The more common reaction in such a case is what happened in the case of *jade*: In discovering that even the core exemplars of *jade* fell into two distinct chemical classes, speakers decided to coin new names for either sub-substance rather than artficially (but, in accord with the causal theory of reference, as we saw) maintaining the word *jade* for *nephrite* only.

Science, in linguistic terms, is the constinuing search for the ultimate qua answer for a given set of (conservative, hopefully constant) referents. If we assume, optimistically, that the search converges we can even add Putnam's natural kind terms as the points of convergence of these word meanings. However, the predecessing stages in this process of convergence are also mirrored in the theory, along with stages of those words where no visible convergence took place

Let me briefly comment on the spellout of Putnam given in Haas-Spohn (1994). She develops an account which predicts that speaker communities can already be in command of Putnam's natural kind terms, i.e. meanings which constitute the ideal points of

33 See Mayr (op.cit.) and Senglaub (op.cit.).

scientific convergence. She proposes that, for instance, the meaning of "gold" should be the function that maps contexts *c* and worlds *w* on the following set:

(30) { *x* | *x* in *w* is *of the same substance (in terms of the ultimate classification)* as the core referents of "gold" in context c }[34]

By making reference to *the ultimate classification*, speakers of all times can participate in the "true" meaning of the word even without knowing its extension. Two objections can be put forward:

(A) Definition (30) rests on the assumption that there be such a point of convergence in natural science. Our own proposal works even without such a point of convergence in science. If there is one, fine. If there isn't, we still can assign the words some meaning.

(B) Something like the word meanings I propose will still have to play a role in the philosophy of science where it comes to an evaluation of a theory. According to (30), almost all theories about gold will be automatically false due to the fact that some - real or counterfactual - objects in the ultimate extension of the "true" meaning of "gold" are not accounted for. However, we will want to distinguish those theories which use their terms in a consistent way from those which make inconsistent claims. We can only do that by looking at what is the "practical" extension of these terms at a given time, not by looking at the "ultimate" extension according to (30).

Classical causal theory of reference has gained its high reputation in part due to the smooth account of scientific continuity it offers. Yet, it will have to address the question how speakers, in spite of their ignorance about sets like (30), can use natural kind terms felicitously in formulating their theories. In order to cover such aspects as "coherence of a scientific theory", it will be necessary to consider the extensions and intensions (of "gold", "water", "elm") predicted by my account. Otherwise it would remain a mystery how humans can, using words of which they not even know the meaning, formulate good theories and even show progress towards the truth.

[34] Note that Haas-Spohn's contexts are much more coarse-grained than ours, basically giving us the core referents in the world *w* of the context, at time *t* of the context.

4.3. Metaphor

How do metaphorical uses of a word ω fit into the picture developed so far? I want to propose that such uses exploit the mechanisms of meaning introduction which were discussed in section one: Once more, a speaker attempts to establish a meaning ("reading") for ω on the basis of sample exemplars and a way of generalization. The choice of samples and similarity occurs under a slightly different perspective than in establishing a "real" meaning, which explains the special status of metaphorical meanings. Let me repeat what I take to be the basis to generate a metaphor. In order to establish a metaphorical sense for some given word ϖ in utterance context *c*

(a) it must be clear which individual or object (or state of affairs, or event) is to be the referent *R* of word ϖ in that context *c*
(b) it must be clear that *R* would not be in the extension of ϖ in its literal sense
(c) speaker and hearer must know the typical referents *S* of ϖ in its literal sense
(d) speaker and hearer must share a certain amount of cultural, expert, or in-group knowledge about these typical referents *S*. In the simplest case, they have to know for which properties these referents *S* are taken to be the most salient examples.

Basically, the metaphorical sense of ϖ is derived by generalising all samples *R* and *S* on basis of a similarity relation ρ which, on the one hand, classifies *R* and *S* as similar, and on the other hand "has to do" with the properties for which *S* are especially typical.

This proposal to account for metaphorical readings of words implements the ideas in Gärdenfors (1996a) into the present framework. We, in contrast to Gärdenfors, have the advantage of an account which can explicitely talk about meaning introduction which allows me to spell out his ideas more concisely. Another related approach can be found in Bartsch (1998); while only Bartsch (t.a.) discusses the requirement that *S* must be salient bearers of the property to be generalized.

Let us go through a simple example. Assume that some speaker in context *c** calls a person Erwin a “Löwe” (German: male lion). The intelligent listener who, obeying Grice's principles, will assume that the speaker would not bother uttering obvious falsities, is thereby invited to form a new category which covers both *core male lions S* and referent *R = Erwin.* The hearer will then compute this new metaphorical meaning of word ϖ on the basis of the same mechanisms which he uses for ordinary baptising and in word learning. The big question is, once more, the *qua* question.

Most importantly, the intended similarity relation ρ_m will have to be such that *S* and *R* are similar. Characteristically, however, the choice of samples $\{R,S\}$ is such that none of the plain, unsophisticated similarity relations will work. It will take some cultural

experience for the hearer to figure out ρ_m. In other words: Speaker's and hearer's epistemic states *ep* must both contain the relevant bits of knowledge which are necessary to figure out ρ_m.[35]

In what respects are Erwin and typical lions similar? Before jumping to evident possible answers, note that there is a large number of "wrong" anwers — all respectable *qua* answers, but strangely enough not in context *c**. Erwin and lions are both mammals. Consequently, they share all common biological structures of mammals. Erwin and male lions are both male. It might even be the case that Erwin is financially supported by his wife in the same way as, according to recent biological findings, male lions are supported with food by the female members of the group who do the hunting. Yet, the speaker using “Löwe” in a metaphorical sense will neither intend it to mean “mammal” nor “male” nor “male supported by female partner”.

In order to decide which common qualities the speaker might think of, the hearer will have to figure out which are the qualities / properties / structure / ... for which the typical instances of the source domain *are most salient* in one's culture or smaller peer group. Lions, in our culture, are prominent examples of animals who have a mane, are strong, powerful and — to the human eye — generally have a majestic, dominant body language. They are not prime examples for “being male”. Nor are they prime examples for “male who is supported by female partner”. (Male bees, drones, might qualify for that position.)

Having narrowed down the spectrum of interesting similarities in this way, the hearer will finally have to decide which of these respects also applies to Erwin. Assume that Erwin is also a fighter by nature, but has short hair. The hearer will conclude that the category denoted by "lion" in the metaphorical sense *intended in this utterance* is the set of brave strong fighters. It would thus be allowed (though not very original) to apply “Löwe” in the same metaphorical sense to other brave persons like Jean, Bo or Evelyn. It would not be allowed to apply “Löwe” in the *same* sense to a person with long and wild hair. Such a referent will require (or create) a new metaphorical sense of “Löwe”.

It would moreover not be allowed to call Erwin a “crocodile” to point out his fighting nature, although a crocodile is equally fierce in aggressive confrontations. Why not? Because “crocodiles” are not the most *salient* fighters, like male lions are not the most salient males who are supported by females. Crocodiles might be the prime example of other qualities which can then be taken as basis for similarity respects to form metaphorical senses of “crocodile”.

According to this view, category formation in the creation of a metaphorical sense of a word and category formation in ordinary baptising are basically the same process. In

---

35 As a result, metaphors, like jokes, can be used to create or express a "we"-feeling between speaker and hearer. These effects can be observed most easily in real life in the conversations of first term students, eagerly confirming their newgained epistemic backgrounds.

both cases, we generalise samples according to a certain respect ρ. In the case of metaphor, the respect is determined (i) by considering the properties for which the samples are prominent examples ("lions are brave and have a mane") and (ii) by choosing those which also reasonably apply to the intended referent in question (= Erwin). If no property matches both (i) and (ii), the metaphor was infelicitous.

What about the potential of metaphors to structure our view of the world? While we could certainly say that finding the correct qua answer in a context like c^* involves highlighting certain properties of the referent R (= *Erwin*), the account as it stands has little to say in a deeper way with respect to this question. In fact, my metaphor example is based on quite simple-minded properties, and it may be questionably whether calling Erwin a "Löwe" will require a serios remodelling of my "Erwin"-concept. However, I would submit that the idea can be extended to more innovative and complex cases of figurative use of language. B. Indurkhya (1992) offers a framework which allows to capture reconceptualisation on the basis of a weak notion of partial homomorphy between source and target domain. Applying Indurkhya's insights, we can assume that more complex metaphors are based on the similarity of "sharing the same underlying structure" in his sense. A full discussion of these questions is beyond the limits of this paper.

Which thereby ends.

References

Bartsch, R. (1998): *Dynamic Conceptual Semantics*. CSLI Publications, Stanford.

Bartsch, R. (t.a.): "*Generating Polysemy: Metaphor and Metonymy*" In: Eckardt, R. & von Heusinger, K. : *Meaning Change - Meaning Variation. Vol I.* Proceedings of a Workshop held at Konstanz, Feb. 1999. Arbeitspapier der FG Sprachwissenschaft, Konstanz (t.a.).

Bekoff, M. (ed.) (1978): *Coyotes - Biology, Behavior and Management.* Academic Press, New York.

Bloom, P. (1996): "Intention, history, and artefact concepts". In: *Cognition* **60** (1996), 1-29.

Bolton, C. (1996): "Proper names, Taxonomic Names, and Necessity". In: *The Philosophical Quarterly* **46**, 145-57.

Chierchia, G. & McConnell-Ginet, S. (1990): *Meaning and Grammar*. MIT Press, Boston.

Devitt, M. & Sterelny, K. (1987): *Language and Reality*. Basil Blackwell, Oxford. Second, revised edition (1998).

Dowty, D., Wall, R.E. & Peters, S. (1981): *Introduction to Montague Semantics*. Reidel Publishing Company, Dordrecht.

Dik, S.C. (1977): "Inductive Generalizations in Semantic Change". In: Hopper (ed.): *Festschrift for Winfred P. Lehmann.* Benjamins, Amsterdam.

Eckardt, R. (u.r.): "A Logic for the GEN Operator." Draft, University of Konstanz (resubmitted at the Journal of Semantics).

Fox, M.W. (ed.) (1975): *The wild canides - their systematics, behavioral ecology and evolution.* Van Nostrand Reinhold Company, New York.

Gärdenfors, P. (1991): "Frameworks for properties: Possible worlds vs. conceptual spaces", in: L. Haaoaranta, M. Kusch & I. Niiniluoto: *Language, Knowledge and Intensionality* (Acta Philosophica Fennica, vol49), 383-407.

Gärdenfors, P. (1993): "Induction and the evolution of conceptual spaces". In: Moore, E.C. (ed.): *Charles S. Peirce and the Philosophy of Science*, The University of Alabama Press, Tuscaloosa, 72-88.

Gärdenfors, P. (1996a): “Mental representation, conceptual spaces and metaphors”. In: *Synthese* **106**, 21-47.
Gärdenfors, P. (1996b): “Conceptual spaces as a framework for cognitive semantics.” In: Clark, A. (ed.): *Philosophy and Cognitive Science*, Kluwer, Dordrecht, 159-180.
Gärdenfors, P. (1997): “Does semantics need reality?” in: Rieger, A. & Peschl, M. (eds.): *New trends in cognitive science - 97 “Does Representation Need Reality?”*, Austrian Society of Cognitive Science Technical Report, 97-01, Vienna, 113-120.
Gärdenfors, P. (1998): Concept combination: A geometrical model". In: P. Blackburn (ed.): *Proceedings of ITALLC '98*, Center of Cognitive Science and Psychology, National Chung University, Chiayi, Taiwan.
— (t.a.): *Conceptual Spaces.* To appear at MIT Press, Boston,. Mass.
Geeraerts, D. (1997): *Diachronic Prototype Semantics*. Clarendon Press, Oxford.
Goodman, N. (1951): *The structure of appearance*. Harvard University Press, Harvard.
Goodman, N. (1972): *Problems and Projects.* Bobbs-Merrill Company, Indianapolis. Kap. IX: Likeness (pp.421-449).
Grzimek, B. (1987): *Enzyklopädie der Säugetiere.* Bd. 4, Kindler, München.
Haas-Spohn, U. (1994): *Versteckte Indexikalität und subjektive Bedeutung*. Ph.D.Diss, University of Tübingen. In print: Akademie Verlag, Studia Grammatica 38, Berlin (1995).
Hahn, U. & Chater, N. (1997): Concepts and Similarity”. In Lambert, K. & Shanks, D. (eds.): *Knowledge, Concepts, and Categories.* Psychology Press, East Sussex, 43 - 92.
Hartmann, E. (1868): *Entwicklungs-Geschichte der Posten von den ältesten Zeiten bis zur Gegenwart mit besonderer Beziehung auf Deutschland.* Leipzig, Verlag Franz Wagner
Indurkhya, Bipin: ”Metaphor as Change of Representation: An Interaction Theory of Cognition and Metaphor.” In: Hintikka, J. : *Metaphors.* Kluwer, Dordrecht. (1994) pp.95-151.
Jackendoff, R. (1983): *Semantics and Cognition*. MIT Press, Cambridge, Mass.
Jackendoff, R. (1998): “Why a conceptualist view of reference? — A reply to Abbott”. In: *Linguistics and Philosophy* **21**:211-219.
Keil, F. (1989): *Concepts, Kinds, and Cognitive Development.* A Bradford Book, The MIT Press, Cambridge, Mass.
Keil, F. (1994): " Knowledge and Category Formation ". In: Gleitman, L. & B. Landau (Eds.): "Lexical aquisition". *Lingua* **92** (1994), special issue.
Kennelly, J. (1978): “Coyote Reproduction”. In: Bekoff, M. (1978), 73-97.
Kitcher, P. (1984): “Species”. In: *Philosophy of Science* **51**, pp. 308-333.
Kluge, F./Seebold, E. (1995): *Etymologisches Wörterbuch der deutschen Sprache.* 23th extended edition, de Gruyter Verlag, Berlin.
Kripke, S. (1972): "Naming and Necessity". Reprinted in D. Davidson, Harman, G. (eds.): *Semantics of Natural Language.* Reidel Publishing Company, Dordrecht (1972).
Lakoff, G. (1994): *Women, Fire, and Dangerous Things*. Chicago, University of Chicago Press (6th reprint).
Lakoff, G. & Johnson, M. (1999): Philosophy in the flesh — The embodied mind and its challenge to Western thought. Basic Books, New York
Lambalgen, M. & van der Does, J. (1998): *Logic and Cognition*. Reader, European Summer School in Logic, Language and Information ESSLLI '98, University of Saarbrücken. Available via website UvA, dept. of philosophy.
Landau, B., Smith, L. & Jones, S. (1998): ”Object Shape, Object Function, and Object Name.” In: *Journal of Memory and Language* **38** (1998), 1-27.
Langacker, R. W. (1990): Concept, Image and Symbol - the cognitive basis of grammat. Mouton de Gruyter, Berlin.
Lewis, D. (1983): “New Work for a Theory of Universals”. In: *Australasian Journal of Philosophy* **61**: 343-377.
Malt, B., Sloman, S.A., Gennari, S., Shi, M. & Wang, Y. (1999): "Knowing versus Naming: Similarity and the Linguistic Categorization of Artifacts." In: *Journal of Memory and Language* **40**: 230 - 262.
Mayr, E. (1942): *Systematics and the Origin of Species*. Columbia University Press, Canada.
Mayr, E. (1966): *Animal Species and Evolution.* Harvard University Press, Cambridge, Massachusetts.
Mayr, E. (1988): *Towards a new philosophy of biology.* The Belknap Press of Harvard University Press, Cambridge, Masssachusetts.
Montague, R. (1974): *Formal Philosophy: selected papers of Richard Montague.* Edited by R. Thomason. New Haven, 1974.
Moravcsik, J. M. (1998): *Meaning, creativity, and the partial inscrutability of the human mind.* CSLI Publications, Stanford.
Partee, B. H. (1979): “Semantics — Mathematics or Psychology?”. In Bäuerle, R., Egli, U. & von Stechow, A. (eds.): *Semantics from different points of view.* Springer-Verlag, Berlin.

Pelletier, F.J. & Asher, N. (1997): "Generics and Defaults". In: J. van Benthem & A. ter Meulen (eds.): *Handbook of Logic and Language*. North Holland, Amsterdam, and MIT Press, Cambridge, Mass. (1997).

Pessin, A. and Goldberg, S. (eds.)(1996): *The Twin Earth Chronicles*. M.E.Sharpe, Armonk, New York.

Putnam, H. (1975): "The Meaning of 'Meaning'". Reprinted in: H.Putnam, *Philosophical Papers*, vol2: *Mind, Language and Reality*, Cambridge, pp.215-217.

Quine, W.O. (1969): *Ontological Relativity and Other Essays.* Columbia University Press, New York.

Rheinberger, H.-J. (1983): Aspekte des Bedeutungswandels im Begriff organismischer Ähnlichkeit vom 18. zum 19. Jahrhundert. In: *History and Philosophy of Life Science* **5** (1983), p. 237-250.

Rosch, E. (1975): "Cognitive representations of semantic categories", *Journal of Experimental Psychology: General* **104**, 192-233.

Senglaub, K. (1978): *Haushunde — Wildhunde.* Verlag J. Neumann-Naudamm, Melsungen, Basel, Wien.

Smith, L. B. (1989): "From global similarities to kinds of similarities - the construction of dimensions in development." In: Vosniadou, S. and Ortnoy, A. (eds.): *Similarity and Analogical Reasoning.* Cambridge University Press, Cambridge.

Sterelny, K. (1983): "Natural Kind Terms", in *Pacific Philosophical Quarterly* **64** (1983):110-125.

Stern, G. (1931): *Meaning and Change of Meaning.* Indiana University Press, Bloomington & London.

Tversky, A. (1977): "Features of similarity". In: *Psychological Review* **84**:327-352.

Wilkerson, T.E. (1995): *Natural Kinds.* Avebury, Aldershot.

Wittgenstein, L. (1963): *Philosophical Investigations.* Transl. by G. E. M. Anscombe. Basil Blackwell, Oxford.

Zeevat, H. (1998): "How fine-grained can you get with diagonalization?" Manuscript, University of Amsterdam, Dept. of Philosophy.

Zemach, E. (1976): "Putnam's Theory on the Reference of Substance Terms". In: *The Journal of Philosophy* **73**:116-27.

Zwarts, J. and H. Verkuyl (1994): "An Algebra of Conceptual Structure: An Investigation in Jackendoff's Conceptual Semantics". In: *Linguistics and Philosophy* **17**: 1-28.