Word meaning and the preformationist fallacy
Jan Koster

1. The inadequacy of naming and preformationism

Words usually have a meaning. It sounds like a self-evident truth but, upon a closer look, it appears to be remarkably difficult to make this simple statement precise or to give it a formulation leading to a comfortable consensus. Traditionally, one of the most common ideas has been that words are naming things and that meanings are the things named by our words. In its most naïve form, naming was seen as following the Adamite example, which goes back to Genesis 2:19-20 (King James version):

19 And out of the ground the LORD God formed every beast of the field, and every fowl of the air; and brought them unto Adam to see what he would call them: and whatsoever Adam called every living creature, that was the name thereof.
20 And Adam gave names to all cattle, and to the fowl of the air, and to every beast of the field; but for Adam there was not found an help meet for him.

In most of the Western tradition since classical antiquity, direct naming was replaced by mediated naming, according to which the relation between words and the world was bridged by an extra layer of items referred to as ideas, concepts or meanings (seen as universals or not). This extra step left the naming paradigm intact.

The same is true for the famous contribution of Ferdinand de Saussure in the early 20th century (Saussure 1959 [1916]). Saussure’s approach is a small step forward in that he makes naming more flexible. In fact, he makes much of his idea that a language’s vocabulary is not a nomenclature. Nevertheless, upon closer scrutiny, Saussure’s idea of word meaning happens to be a variant of the naming paradigm. The well-known relation between signifiant (the outer form of a sign) and signifié (the inner, conceptual side of a sign), ultimately, is a naming relation, in which the former can be seen as the name of the latter. Where Saussure differs from some older traditions is that he does not see concepts as fixed once and for all but as culture-dependent. The color terms of the various languages, for instance, break up the color spectrum in different ways. Some languages may have one word for “blue” and “green,” while others use two words. Another way of saying this is that a color term has a “value” in a system, i.e., it thanks its identity to its differences with other color terms dividing up the same “semantic field.”

Saussure’s view has led to a great amount of anecdotal evidence as to how different languages divide up reality. I think it is fair to say that this has shown that there is no strict uniformity across cultures, even with respect to the most basic vocabularies, like those naming body parts. English differs, for instance, from Dutch in that it has only one word for the body part connecting the head and the trunk, neck, whereas Dutch has two words: nek (mainly the back part) and hals (the rest, particularly the front part).

The way Saussure’s contribution is normally interpreted is unsatisfactory in a number of ways. First of all, making conceptual content culture-dependent is hardly original. As a matter of fact, it was the general trend in the 19th century, which was in turn the outcome of an assault on classical universalism that had gone on in Europe for centuries. In the late Middle Ages, the attack was led by the nominalists, to be taken over by their heirs in early modern times, the British empiricists. As is clear in the writings of John Locke, the philosophical anti-universalism was further fed by the European expansion and the ever
greater familiarity with the world views and concepts of other peoples.¹ But the real explosion of anti-universalist multiculturalism came with the transition to the Romantic movement, at the turn of the 18th century, with figures like Herder and, to a lesser extent, von Humboldt.

Like most French *philosophes*, particularly Herder was influenced by British empiricism, although Enlightenment universalism did not disappear overnight.² Nevertheless, the movement towards historicizing epistemology in general, and language and meaning in particular, was unstoppable by this time, even among thinkers that were upholding universalist Enlightenment values in many other respects, like Wilhelm von Humboldt (cited and emphasis added by Wierzbicka 1992: 5):

To be sure, a midpoint, around which all languages revolve, can be sought and really found, and this midpoint should always be kept in mind in the comparative study of languages, both in the grammar and lexicon. For in both there is a number of things which can be determined completely *a priori*, and which can be separated from the conditions of a particular language. On the other hand, there is a *far greater number* of concepts, and also grammatical peculiarities, which are so inextricably woven into the individuality of their language that they can neither be kept suspended between all languages on the mere thread of inner perception nor can they be carried over into another language without alteration. (1903-36, v.4: 21-23)

Herder and von Humboldt set the tone for a large part of 19th-century thinking about the relation between language and thought and Saussure, but also Boas, Sapir and Whorf later on, are heirs to this tradition.

Clearly, von Humboldt’s historicism was kept in check by a notion of universalism, but cultural relativism was on the rise. In many respects, then, Saussure was a child of a century, which, in the hands of Darwin, succeeded in making even biology a science driven by historical contingency. What is particularly problematic about Saussure’s theory of the sign is that the forms conceptualizations can take are said to be “arbitrary.” Thus, although it is correct that different languages break up conceptual fields in different ways, thinking that arbitrariness reigns in this respect, is an error. Color terms are a case in point. Although the spectrum is divided up in different ways, there will be no languages whatsoever with one word for “yellow” and “black” instead of, say, one word for “blue” and “green.” Obviously, possible choices are constrained in all kinds of ways and in many cases choices seem to be made within a rather narrow range.³ I think the latter might particularly apply to those parts of the lexicon that are relatively basic, but even here it seems a great exaggeration in the other direction to maintain that selection of conceptualizations is of the same limited kind as parameter setting in syntax.

Apart from the unsatisfactory relativism entailed by Saussure’s notion of arbitrariness, I do not believe at all that what he calls *signifié* can be properly characterized as word meaning. What Saussure really does is to point out that there are differences in the common sense theories of reality as entertained by various cultures. These theories involve meaning, to be sure, but saying that the atoms of such theories (the varying *signifiés*) are meanings turns his view into a relativized version of the naming paradigm. Meanings are seen as single entities as before (but see Saussure (1959 [1916]: 108). Instead of naming universally-fixed entities, Saussure’s words (his *signifiants*) name culturally-fixed entities (his *signifiés*). Both views are mistaken, as I will argue, because the set of meanings corresponding to most words is not determinate and finite but indeterminate and infinite.

---

¹ See Carey (2004) for the role of travel reports in Locke’s rejection of innate ideas.
² For the influence of British empiricism on the ideas of Herder, see Forster (2007).
³ See Berlin & Kay (1969)
A more radical attack on the naming paradigm is usually ascribed to the later Wittgenstein, who explicitly criticizes the “Augustinian view of language” at the first pages of the Philosophical Investigations (Wittgenstein 1953). He does not see meanings as things at all, neither as entities in the mind (“concepts”) nor as entities in some mind-external reality (some Platonic realm or physical reality or whatever). Meanings are found in forms of use following the rules of certain language games. What I find particularly interesting is Wittgenstein’s critique of the idea of meanings (or concepts) as superfluous “shadows” between words and the meaning manifested in the use of words (Blue Books, Wittgenstein 1958: 36-37). More generally, the idea that the interpretation of a representation is itself a representation leads to infinite regress (Blue Books: 34):

> What one wishes to say is: “Every sign is capable of interpretation; but the meaning mustn’t be able of interpretation. It is the last interpretation.”

I think Wittgenstein had it right: meaning can only be found in use and the traditional quest for “shadows” as an extra layer between words and their interpretations is a dubious move.\(^4\) If one maintains that a word stands for a concept, there is practically nothing that can be said about the concept that could just as well be said directly about the word. It should be noted that I only follow Wittgenstein here under a certain interpretation. Wittgenstein’s remarks about meaning depending on language games are often taken as an argument for relativism or post-modern perspectivism. It is questionable whether this was Wittgenstein’s intention (see Haller 1988, ch. 8). No matter the right Wittgenstein exegesis, I adopt the meaning-as-use view under the usual rationalist assumption that it is not the case that just anything goes but that possible meanings fall within a limited range defined by human nature.\(^5\)

Wittgenstein’s insights go remarkably well with the philological tradition, that has always insisted that a word only has meaning in a particular context or situation. Placed in different contexts, words show an endless flexibility of interpretation, usually called polysemy, mostly based on variants of metaphor and metonymy. Such phenomena have been observed at least since Aristotle’s Poetics.\(^6\) Take a color term like “red.” The range of things

\(^4\) See also Travis (2000). I take the term “use” as including not only overt language games but also introspection and internal processes leading to understanding. If I refer to concepts as the result of construction or interpretation, it is not my intention to reify meaning in some alternative way. What I have in mind is “last interpretation,” a state comparable to qualia in that it cannot be grasped from the “third person” perspective of science.

\(^5\) According to a common Wittgenstein interpretation, Wittgenstein refuted traditional metaphysics, both in its Platonic, ontological flavor and in its Kantian, epistemological flavor (the latter sometimes thought to include psychologism). Baker & Hacker (1984: 62), for instance, attribute to Wittgenstein “... a conception of language, thinking and understanding that is antithetical to psychologism and to Platonism.” They base that interpretation on Wittgenstein’s critique of the “Augustinian” naming paradigm. Whether this interpretation of Wittgenstein is justified or not, it is not what I propose. What I propose is to give up the Augustinian naming paradigm (and its idea of pre-existing concepts) in favor of a constructivist (“epigenetic”) approach to word meaning. Constructivism is not antithetical to metaphysics, as there must be constraints on possible constructions. To the extent that these are transcendental, constructivism is compatible with both Platonic and Kantian realism.

\(^6\) Discussion of metaphor and metonymy was common in the rhetorical tradition and is usually believed to go back to part 21 of Aristotle’s Poetics. The first clear definition of metonymy is said to be found in the Rhetorica ad Herennium (Anonymous [Cicero] 1954), sometimes ascribed to Cicero. Nerlich, Clarke and Todd (1999, 361) further cite Householder (1995: 93), who refers to Democritus (as quoted in Proclus’ commentary on Plato’s Cratylus) as one of the early sources about polysemy and metonymy. There are numerous modern studies about polysemy and metonymy, like Panther & Radden (1999). See also Pustejovsky (1993) and Moravcsik (1998).
it can refer to is literally infinite, depending on context (possible interpretations between square brackets):

(1)  
a. Can you buy me a tube of red? [paint]  
b. The East is red [socialist]  
c. He prefers red with beef [wine]  
d. He got into the red [deficit]  
e. Red defeats white, 20-14 [wrestling team, message on internet]  
f. Red is going out of fashion [cars, clothes, you name it]  
g. Red available again [ad about red peppers on internet]  
h. etc.

The polysemy shown in such examples can be demonstrated with practically any word, even with logical words like “and” and “all.” There cannot be any doubt, in any intuitive sense of the word “meaning” that “red” has different meanings in examples like those in (1). Different interpretations of “red” leads to difference in reference and truth values. Thus, sentence (1c), *He prefers red with beef*, might be true when “red” is interpreted as “wine” but false if “red” is interpreted as “paint” (as in 1a). Since different interpretations of “red” lead to different truth values, the conclusion is inescapable that different concepts are created, not just contextual variations of the same concept. In principle, each of these concepts could be assigned a different word but thanks to the beautiful creative economy of polysemy we can refer to infinitely many more things than we have words for. Polysemy, therefore, is a central fact about word meaning and human creativity, which, to look ahead a bit, makes it impossible to see word learning as the labeling of concepts.

To avoid the obvious but perhaps uncomfortable conclusion that the number of meanings of “red” is infinite, a vacuous distinction is made sometimes between semantics and pragmatics. In this view, polysemy is a matter of pragmatics while at the semantic level “red” has only one meaning, also known as the concept RED. A major problem for this view is that nobody has ever succeeded in coming even close to give an exact description of the nature of this common concept. This has been observed many times before. Stern (1931: 40), for instance, formulates matters as follows:

> Those theorists who have operated in their analysis with word and meaning only, neglecting the referent, have, in seeking for the constant element in the meanings of a word, been compelled to assume an abstraction variously termed *Begriffskern*, *Kernbedeutung*, *begrifflicher Inhalt*, etc., which was supposed to recur unchanged in the meaning of a word in all its occasional uses. Nobody has been able to define this *Begriffskern* in a satisfactory manner; it has remained a phantom of logical abstraction.

I agree that Wittgenstein’s “shadows” are even better characterized as “phantoms,” but nevertheless a problem remains: what do the various uses of “red” in (1) have in common? If we reject the traditional view that it is the concept RED, what other options do we have? I think that, borrowing from the history of biology, we can choose between preformationist and epigenetic modes of representation (see Koster 1994 for discussion). Older theories of ontogeny held that, say, elephants, grow from sperm with little elephants in it. Absurd as this might sound nowadays, it was as late as during the 19th century that preformationism was completely replaced by the modern, epigenetic view. According to the latter view, the

---

7 For a historical perspective, see Westfall (1977: 100) and Hankins (1985: 141).
ontogenetic beginning of an animal is not at all animal-like but a fertilized egg cell initiating constructive processes regulated by genes that only gradually lead to elephants. In a time of DNA and blooming evolutionary-developmental biology (“Evo-Devo”) the epigenetic view has become a truism. Any code, including DNA, is intrinsically meaningless, but only is “something” thanks to an interpretive environment, like the one provided by a living cell or the physical environment created by the development of the emerging embryo.

Our world of computers and information-carriers is full of examples essentially demonstrating the same point. Thus, two copies of a music CD represent the same music but not in a preformationist way, i.e., the CDs do not contain music but code that is in no way music-like. Music is only produced thanks to interpreting machinery that adds information of its own in crucial ways. As the users of vinyl records know (or remember), not even the speed of the music is represented on the LP itself. I would like to use these facts to make a general distinction between preformationist representation, which is iconic in various ways, and epigenetic representation, which is radically non-iconic, like the digital forms of representation familiar from everyday life. Epigenetic representation is not only radically non-iconic, it also is radically incomplete (information-wise) with respect to the thing represented. It only represents thanks to the crucial information-adding contributions of an interpreting environment.

It seems to me that the traditional idea that the common core of the meaningful interpretations of a word is itself also meaningful (namely, a concept) is an instance of what I would like to call “the preformationist fallacy.” Take the following typical statement (from Fodor 1980):

I take it from this that literally isn't such a thing as the notion of learning a conceptual system richer than the one that one already has; we simply have no idea of what it would be like to get from a conceptually impoverished to a conceptually richer system by anything like a process of learning.

Whatever else one might hold against the notion of “learning,” this certainly is a non-argument. Language learning (including concept formation) is often compared to biological growth. Well, all known forms of growth in nature involve development to a “system richer than the one that one already has,” like the development from elephant DNA to full-fledged elephants. I suspect that the preformationist fallacy is behind both the particular form of Fodor’s idea of innate concepts and his ideas about the language of thought (Fodor 1975). I will return to the latter in the next section. Like Fodor, I prefer rationalist theories over empiricist ones, but that does not force us to commit ourselves to preformationism. On the contrary, I believe that in matters of meaning and knowledge, preformationism will eventually meet the same fate as preformationism in biology.

What, then, does an alternative to preformationism for meaning and knowledge look like? First of all, we can make the same distinctions as for other forms of epigenetic representation. In other words, also for humans and other beings with cognition it makes sense to make a distinction between data structures and interpreting or processing environments working on these data structures. From this point of view, it is only a metaphor to say that books in libraries contain “knowledge.” The knowledge stored on media of any kind is in fact not knowledge at all but only counts as such in crucial conjunction with external interpreters, particularly --in this case-- the interpretive powers of the human interpreter. In accordance with what we see in all other cases of epigenetic representation, the information in books is not knowledge-like at all and is greatly impoverished in comparison to the information structures, and ultimately to the knowledge, that the mind constructs on the basis of what is coded in books.
What is often overlooked is that similar considerations apply to what happens inside the brain. Thus, also inside the brain, a distinction must be made between coded data representations on the one hand and other brain parts and computational mechanisms interpreting these data structures on the other hand. It is just as meaningless to say that the brain contains meanings and knowledge as it is meaningless to say that libraries contain meanings and knowledge. Meaning is only created in interpreting data structures (a form of use), where it makes no principled difference whether these data structures are, to whatever degree, situated inside or outside the skull. Taken apart and in isolation from one another, neither data representations nor interpretive mechanisms can be seen as instances of meaning or knowledge. It is, of course, normal usage to say that someone “has” knowledge, but this usage should not mislead us into believing that knowledge dwells in the brain in abstraction from the interpretive acts and processes of the living interpreter. Ignoring this elementary fact would lead to absurdities as assuming that if one dies, one’s knowledge survives in the brain. For similar reasons, a metaphor is used when it is said that knowledge is innate. We are born, to be sure, with rich data structures and interpretive powers growing according to our genetic program, but it would be foolish to say that knowledge is in place somehow even before the interpretive powers can be used.

A consequence of my view is that there is reason for skepticism about expressions like “the mind/brain” and the usual distinction between internalism and externalism in cognitive science (see Koster 2009a for more elaborate discussion). In spite of the ubiquitous availability of computers helping us, most of our personal data processing is still done inside the skull, but for many data structures the distinction is meaningless. Thus, I do not see an epistemological difference between playing a piece of music from brain-internal memory or from a brain-external score.

Back to word meaning, we can say the following. When we use a word like “red” we do not retrieve a meaning from our brain but we apply whatever information we dispose of to construct a meaning. Assuming that there are meanings or concepts before we actually construct one is falling victim to the preformationist fallacy. Like in ontogenesis, I suppose that the outcome of the construction process, is much richer than the original selection from the data structure associated with “red.” As in all cases of epigenesis, much information is added during the process of interpretation and construction itself, including responses to the information provided by situation and context. The rationalist aspect of this process is not pre-existence but the fact that possible constructions must be heavily constrained by principles of necessity, biological or otherwise. As far as I can see, the logic of the problem is exactly like what we see during biological growth: development is from something relatively simple to something rich and complex, all within a framework of necessity provided, for instance, by physics. What I am advocating, then, is not adopting empiricism but liberating rationalism and cognitive science from preformationism, as happened with biology in the 19th century.

As for the nature of the information associated with words, it is impossible at this point to be very explicit because there is no reason to expect that the information from data structures is directly accessible for introspection. What we usually do when we reflect on the meaning of a word via introspection is to already interpret the stored information, with results that go way beyond what is stored with individual words. For instance, it is hard to think about word meaning without tacitly considering contexts of use and generating potential perspectives of various kinds. We are just fooled by introspection --a form of active construction of meaning-- into thinking that there are ready-made meanings stored in our brain.

Traditionally, for related reasons, reflection on word meaning is accounted for by verbal means, leading to definitions, paraphrases, feature sets or meaning postulates. All of
this is explication of word meaning by other words, often with the describing words disguised as “concepts” (usually written with capital letters) or “markers.” There is absolutely no reason to assume that the data structures associated with words are limited to information that can be expressed by other words. Our ways of being acquainted with the world are much richer than suggested by this traditional practice. Sensory impressions and imagery (including prototypes), for instance, must play a very important role as well. As everyone who has ever tried will confirm, verbal explications of even the most common nouns, like “cheese,” are doomed to failure. What comes to mind when a word like “cheese” is used is not necessarily other words, but sensory impressions and images.

Very often, imagery (including prototypes) is put aside in discussion about meaning because it is, rightly, pointed out that images are not meaning, as they fail to contribute to normal compositionality. Such objections are beside the point since nothing else stored with individual words can be properly called meaning either. Meaning, once more, is what is constructed on the basis of data structures, not what is stored in the databanks associated with words. Seen that way, there is no reason to artificially restrict permissible input data (for meaning construction) to what can be expressed verbally.

This much being said, we can be sure that the information stored with individual words is varied and rich and differs enormously from person to person. A word is the “address” of a databank that is growing throughout one’s life. This common sense observation has been discussed for ages. Reichling (1935: 321-22), for instance, building on earlier sources, discusses the use of the word “cow” by the farmhand, the expert-biologist and the child. This is in no way different from Putnam’s well-known discussion of elm and beech and what he called “the linguistic division of labor” (Putnam 1975). The same can be demonstrated with almost any word, including the word “red.” Thus, the databank one has for “red” can be almost empty, when one asks a question like (2a). It can be informed by common sense knowledge (2b), or by information derived from scientific theories (2c):

(2) a. What does “red” mean?
   b. This red is very vivid
   c. Red falls between 630 and 740 nanometer

These examples show once more that it is futile to postulate a common meaning for a word like “red.” Red, like all other words, has a databank, which is, at least partially, filled with different data structures for different people. These databanks overlap and what the various uses of “red” have in common is that they select and apply information from the same data sources.

It is sometimes said that common sense understanding has a privileged status in word meaning because, being based on innate concepts, it is acquired in ways different from scientific knowledge. This would make a division in our vocabulary between “water” as the universally familiar liquid and, say, as H₂O (cf. Chomsky 2000: 43). However, apart from the truism that common sense knowledge precedes scientific knowledge, this idea is beside the point. First of all, as we have seen, what is stored for words is not concepts at all but data structures. Furthermore, as shown by (2a), a word can be used with a very limited databank, so limited that it is not even informed by common sense knowledge. Even with children, it can be demonstrated that word-related databanks grow over time, even within the domain that is usually referred to as common sense understanding. There is just a continuum from information based on the most elementary common sense to the most sophisticated scientific theories. There simply is no cut-off point of where we leave common sense behind and enter the halls of science.
But the whole discussion is misguided and informed by preformationist prejudice. What is stored for words, I must repeat, is not meanings but information structures (that can be innate to various degrees). Meanings, however, are at a different level in that they are constructed from data structures, no matter their origin. In other words, both the interpretive capacities and certain data structures might be innate, but for the interpretation of a word (the construction of “meaning”) it does not make an ontological or epistemological difference if input data are taken from innate resources, from family or school, from books or from the annals of science. In all cases the task facing the language user is the same: how to construct meaning from available data.

If this view is correct, it is also demonstrated once more that “internalism” is applicable perhaps to our interpreting faculties but not to the other necessary component of language, the data structures. It is irrelevant whether the input data for the interpretation of words come from my brain or from some source external to the brain. Thus, I do not temporarily step out of the English language, when in the middle of producing or understanding an English sentence, I inform myself by looking up a word in a dictionary.

Since the databanks associated with words are so different from person to person, communication is always a “more or less” phenomenon, depending on overlap in databanks and the background beliefs that enter into the constructive (interpretive) process. Apart from that and given that there is no upper bound to the amount of meanings that can be constructed for a given word, one might wonder why people often converge so rapidly on the interpretation of a word or converge at all. Rapid convergence of interpretation is indeed one of the wonders of language. Convergence must be facilitated by inborn constraints, largely unknown so far, on how we see the world and interpret it. Against this backdrop, context and situation further limit the “search space” almost conclusively. Thus, if we talk about food and wine we do not first try the interpretation “paint” or “socialist” when somebody says that he prefers red with his beef. In many other cases, possible interpretations are constrained by convention. Suppose that Sarah Palin, in an attempt to emulate the wisdom of Ronald Reagan, makes the following mistake:

(3) Washington is not the problem but the solution!

One might think that proper names, like “Washington,” are the flagships of the naming paradigm. However, nothing is easier than demonstrating polysemy with proper names. “Washington” can be used to refer to a certain city, but, more often than not, it is used for something related to that city. Since there is no upper bound to the amount of things that can be related to “Washington,” the amount of meanings (concepts) corresponding to the word is infinite. Nevertheless most of us instantaneously take “Washington” in (3) as referring to the American political establishment in Washington. This interpretation is conventional and the use of words is massively governed by such conventions. Altogether, the use of words depends on three kinds of conventions: the Saussurian conventions of arbitrary sound-meaning coupling and dividing up semantic fields and, more importantly, the conventions of usage as just demonstrated.

Summarizing so far, we can say that traditional theories of lexical semantics have been led astray by the naming paradigm and by the preformationist fallacy. The most naïve, Adamite, naming theories were rejected since the dawn of philosophical reflection. However, the naming paradigm survived in the practice to place “concepts” between interpretations and referents in the world on the one hand and our words on the other hand. Concepts were seen as reified and pre-existing, as inhabitants of a Platonic heaven or as mental entities of some
kind. I think a forceful step in dismantling this more sophisticated form of the naming paradigm was taken by Wittgenstein and the ordinary language philosophy influenced by him (sometimes within an unnecessary empiricist framework). This tradition opened the way to considering meaning and knowledge without shadows and phantoms standing between words and their interpretations.

From a tradition independent of Wittgenstein (Reichling 1935 and sources cited there), it was concluded that the static conception of word meaning must be replaced by a dynamic view. In my modernized version of this view, a distinction was made between data structures on the one hand and our capacities of meaning construction (and interpretation) on the other hand. Concepts only exist as outputs of the construction/interpretation process applied to words. Referring to Jespersen’s similar insights, Reichling (1935: 357) alludes to Wilhelm von Humboldt when he says that the word (particularly its meaning) is an “energeia” and not an “ergon.” It has become common to see von Humboldt’s notions as applying to recursive syntax, but I agree with Reichling that they also apply to the construction of word meaning (concept formation). As in syntax, when we construct meanings, we make “infinite use of finite means.” In the case of concept formation, the finite means include the context-sensitive selections we make from the data structures associated with words. Following a largely unknown but finite set of principles of interpretation, we apply the available information to ever changing situations, generating appropriate concepts. There is no upper limit to the amount of concepts compatible with a given word and therefore it is pointless to see a word as something naming a concept.

The idea that meanings (concepts) are in place prior to acts of construction/interpretation in context was rejected as “the preformationist fallacy.” The pervasive idea of “preformationism” in lexical semantics should be replaced, it was argued, by an “epigenetic” approach, not unlike the epigenetic turn developmental biology took in the 19th century. All of this can be done within an overall rationalist framework, i.e., by maintaining the idea that possible semantic constructions fall within a narrow range. As we will see in the next section, preformationism is still underlying much discussion of word meaning, up until the present day.

2. Overcoming preformationism

Preoccupation with formal languages in logic and mathematics has often led to misleading ideas about natural language. I will not get into this rich topic here but will only briefly mention one aspect, namely what is sometimes referred to as “referentialism.” According to referentialism, the words of a language have a relation to mind-external entities expressed by an interpretation function that assigns values to the expressions of the language in some domain D. Developed and used for the study of formal objects and systems, as in logic and mathematics, such forms of semantics have proven their value but it has nothing to do with how natural language works, as was explicitly recognized by the most important pioneer of this type of semantics, no one less than Alfred Tarski himself (Moravcsik 1998: 14). Applied to natural language, referentialism comes down to the most naïve version of the naming paradigm. For natural languages it is an antediluvian, even pre-Adamite view. I therefore entirely agree with Chomsky (2000: passim, 2007: 7 and 2009: 27) that natural language has no semantics in the intended technical sense.

It is more difficult to evaluate how Chomsky’s own internalism fares with respect to the naming paradigm. As will be clear from the preceding section, what I reject is not just referentialism but any version of the naming paradigm. Saying that words refer to concepts is
just as bad as saying that words have a fixed relation to mind-external entities. A word is not the name of a concept (or a limited number of concepts) but an address in a databank with information that can be used to construct an unlimited amount of contextually appropriate interpretations. In other words, concepts are not properties of words but the outcome of interpretations of words, where, generally speaking, the interpretation process adds information beyond what is stored in association with the word itself.

Frankly, therefore, I do not quite understand what Chomsky says about UG in passages like the following (Chomsky 2007: 6):

In addition to Merge applicable without bounds, UG must at least provide atomic elements, lexical items LI, each a structured array of properties (features) to which Merge and other operations apply to form expressions.

Recall that UG is a technical term characterizing the initial state of the language faculty or the natural endowment of a child prior to language acquisition. As I understand this passage, it can only mean that a child disposes of some kind of proto-words, the lexical items LI, that are not part of, say, English but of UG. If so, one wonders how these proto-words are related to the words of English or Chinese. If we ignore the difference between proto-words and concepts for a while, this appears to be a process involving labeling (Chomsky 2000: 65-66):

However surprising the conclusion may be that nature has provided us with an innate stock of concepts, and that the child’s task is to discover their labels, the empirical facts appear to leave open few other possibilities [emphasis added --JK].

The inescapable empirical evidence is seen in the light of the familiar poverty-of-the-stimulus argument (Chomsky 2000: 61):

Acquisition of lexical items poses what is sometimes called “Plato’s problem” in a very sharp form.

[...] At peak periods of language acquisition, children are acquiring (“learning”) many words a day, perhaps a dozen or more, meaning that they are acquiring words on very few exposures, even just one. This would appear to indicate that the concepts are already available, with much or all of their intricacy and structure predetermined, and that the child’s task is to assign labels to concepts, as might be done with limited evidence given sufficiently rich innate structure [emphasis added --JK].

To begin with, I would like to mention a point of agreement: the explosive growth of the child’s vocabulary is indeed astonishing and shows, given the correctness of the poverty-of-the-stimulus argument, that something is innate. Beyond that, much remains unclear. It seems to me that the complete argument is based on two questionable assumptions: 1) that a version of the naming paradigm is correct (“labeling”), and 2) that innateness means pre-existence of concepts and, therefore, preformationism.

So, suppose that the child has acquired the simple English common sense-related word “water.” According to the second quotation, all the child has done is labeling something that was already in place, namely the innate concept corresponding to “water.” Labeling is a clear instance of the naming paradigm and must therefore be rejected according to what was concluded above. To repeat, the English word “water” is known to correspond not to a single concept but to an unlimited amount of concepts, depending on the results given by our interpretive capacities in context. There is simply no one-one mapping between words and concepts and the proposed labeling procedure must therefore be an error.
What about the poverty-of-the-stimulus argument? Recall that in the previous section it was concluded that what is associated with words is not concepts --a traditional, preformationist prejudice-- but data structures that can come from various sources and that vary from person to person, from almost nothing, via common sense data to expert “knowledge.” Suppose that a child asks the following question:

(4) Daddy, what is water?

Clearly, the child would have acquired the word “water.” But what would he or she label with this word in (4)? Not a common sense concept because the very point of the question is that the child’s “folder” for “water” is still relatively empty, although there might be already some clues associated with the word, making its databank not entirely empty. This example suffices to show that the acquisition of a word is not necessarily labeling a concept because it shows that word learning can precede substantial association with conceptual reality.

Suppose now that the father in (4) answers the question by pointing to the liquid that is coming from the tap. The kid sees the familiar stuff and now instantaneously associates this with the word “water.” Is this labeling a concept? Of course not. What the child adds to its database for “water” is a rich, largely innate assembly of data structures. As before, the exact nature of those cannot be determined with current understanding and technology and in this case our own introspection will be even less helpful than in analyzing the conceptual world of adults. We can be sure that the data structures are rich and varied and include information related to sensory impression, perception and our capacities for imagery. None of that is in itself meaning-like or concept-like, as meaning only arises as soon as the data structures in question are put to use in a process of construction or interpretation. As before, this process can be assumed to add much information of its own, leading to concept formations way beyond what was originally stored in the databank for “water.” As before, the relation between data structures for “water” and real concepts (only arising by actual use in actual contexts) must presumably be seen along “epigenetic” lines rather than in terms of preformationism.

Note that in this alternative account, as much is supposed to be innate as required by the original poverty-of-the-stimulus argument. What is criticized here is not nativism or rationalism, but preformationist versions of it. The assumption of innateness is further justified by the fact that human cognition does not come out of the blue but has a long evolutionary history. Also animals have capacities to represent and interpret the world in diverse ways and with internal data processing, suggesting a continuity with humans to various degrees. However, to see meaning (or concepts) in these capacities is as metaphorical as seeing knowledge in libraries and hard disks or decisions in automata. Meaning is at a higher level than data structures and much processing. It is the application of data structures to an unlimited number of situations by a human interpreter. Application involves an amount of agentive function assignment, like the use of material tools, which we share in a very limited way with other primates. Agentive function assignment, no matter how rudimentary, seems to require external signs, as I will propose in a minute.

But first, I would like to point out that the quotes from Chomsky (2000) above are not quite consistent with the quotes from Chomsky (2007). In the later quote (2007: 6), the words of real languages, like English, are not directly connected to concepts, but there is an intermediate layer of what I called proto-words: the lexical items LI belonging to UG and therefore to the state that precedes the acquisition of the words of English. This extra layer of proto-words is extremely important for the so-called biolinguistics program in its current form.
to make sense. The latter requires that the core operation of syntax, Merge, has a linguistic function before, and independent of, a crucial role being played by invented cultural objects like the words of English. If Merge is integrated into language by cultural invention, there is no such a thing as I-language comparable to an organ but rather integration of unrelated, non-linguistic biological capacities via an external, culturally transmitted route.

Note that the assumption of proto-words further works things into the morass of the naming paradigm. Proto-words are not concepts or meanings but symbols, otherwise they could not be manipulated by Merge. This leads from one naming relation to two (→ indicating the naming relation):

(5) English word → proto-word → concept

This preserves what was questionable about the singular naming relation, namely the reification of the notion “concept,” i.e., seeing concepts as pre-existing things that can be named in the first place rather than as the infinite number of interpretations words can get in context. But actually, (5) seems to make things worse, as we will have a hard time avoiding Occam’s razor: English words and their inner representations can do everything proto-words do and better. Postulating a second layer of words suggests the utmost superfluity and brings a considerable burden of proof, a burden of proof that has not been met so far.

The postulation of proto-words is a transition to ideas about a language of thought, which seems to be Chomsky’s preferred way of framing things in recent writings (see, for instance, Chomsky 2009: 29). In modern times, the language of thought idea is associated with Fodor (1975), but it is a remarkably old idea, with in fact more than superficial similarity between the late-medieval version of William of Ockham and the one popularized by Jerry Fodor (see Panaccio 2003 and 2007). The distinction between logos endiathetos (‘inner logos’) and logos prophorikos (‘uttered logos’) was widely discussed in classical antiquity and the oldest attestation is supposed to be in the writings of Philo of Alexandria (20 BCE - 50 CE), who in turn is said to be building on Stoic sources (Kamesar 2004). Claude Panaccio, who dedicated a monograph to the topic, challenges this conclusion on the ground that the idea was common among all philosophical schools of Hellenistic antiquity (Panaccio 1999). Be this as it may, Chomsky and Fodor are in good, classical company so far.

However, there is a serious problem with all theories of inner, mental language (also known as Mentalese). The problem is that nobody has succeeded in specifying the words of this language or even in giving a minimally plausible suggestion as to their nature. If the syntactically manipulated atoms of Mentalese are concepts, one runs into the problem that concepts are not things, like words, but interpretations of such things. If, on the other hand, the manipulated atoms are word-like, the entire language of thought seems to be superfluous, as everything that the language of thought does can also be done by an inner representation of good, old English. So, the idea is either unintelligible or falling victim to Occam’s razor.

There is no point in reviewing the substantial literature here that was generated by the revival of the ancient idea in Jerry Fodor’s numerous writings. But as far as I can see, there is not a single convincing argument that survives Occam’s razor (in the sense just discussed). Furthermore, Fodor’s discussions are full of the preformationism-flavored arguments that I reject (“no language without earlier language,” innate concepts, etc.). None of that is

---

8 For the biolinguistics program (going back to early generative grammar, with Lenneberg 1967), see Jenkins (2000) and Hauser, Chomsky & Fitch (2002) among others.

9 See Kaye (1999) for an excellent overview and annotated bibliography. See also Aydede (2004). Cole (1998) has a pertinent and very critical discussion of arguments given by Pinker for the existence of Mentalese.
convincing for someone who believes that preformationist rationalism should be replaced by an epigenetic style of theorizing about word meaning.

Chomsky’s version is not quite the same as Fodor’s, but the problems are the same: no plausible suggestions about the nature of the lexicon of the inner language and no way to escape Occam’s razor. What I find even harder to adopt is Chomsky’s use of the notion of an inner language to even further marginalize the external dimension of language than was done so far. What the child was supposed to do was, initially at least, not much more than labeling the concepts (or proto-words) of a language already in place. This ontogenetic hypothesis has now received a phylogenetic counterpart of similar kind. In his recent speculations about the evolution of language, Chomsky assumes that language originated as a language of thought, with Merge combining internal, not yet externalized symbols. This stage, in Chomsky’s view, existed for quite some time before the need for “externalization” arose (in an ever larger breeding group disposing of the language of thought) (Chomsky 2009: 29).

Without any historical record whatsoever, it is hard to argue about things that happened 100,000 years ago. Nevertheless, on the basis of what we know about language in the present, it seems to me that the scenario as sketched is very implausible. There are plenty of reasons to assume that having external symbols is not marginal but the essence of language, the sine qua non leading to the Rubicon that came to separate us from our ape-like ancestors. In order to construct meanings (and to combine them via Merge), we need discrete elements --symbols-- that can be manipulated and that can function as pointers to databanks corresponding to these symbols, databanks to be used for the construction of meaning. So far, even if it is not entirely impossible that this function is fulfilled in some limited way by inner symbols, there is not the slightest evidence for word-independent inner symbols. As long as the extra burden of proof is not met, Occam’s razor reigns.

Apart from the fact that there are no good reasons to assume inner UG symbols (not to be confused with inner representations of the words of English or other languages), it is easy to see that externalization redefined the human essence. Apart from more or less “automatized” social instincts, all animals, apes included, live in a relatively confined mental world, which, lacking a substantial externally coded culture, is largely determined by the data structures and processes of their own brain. Humans, in contrast, are not mental individuals in this sense but symbionts, i.e., beings with brains that interpret shared and therefore external data structures.10 Being symbionts in this sense, sets us sharply apart from animals. As briefly mentioned before, it is just an error to describe human cognition in terms of the mind/brain. That would be adequate for animal cognition, not for us symbiotic humans. Insofar as our brain contains processing facilities it is still largely individual, but for the interpretation of data structures, it does not matter whether the information comes from our own brain, from other people or from whatever records we have (like libraries or the internet). Brain is perhaps individual, mind is not. Overlooking the extended and symbiotic nature of our minds, based as it is on publicly available and shared data structures, is missing much of the human essence.11 It is externalization that made us human, not some obscure and unattested language of thought.

Somehow, the capacity to understand external symbols also touches upon the mystery of consciousness: the same capacity that enables us to access the outer expression of the thoughts of others also enables us to be aware of our own thoughts (see also Carruthers 1996). So, this is another reason to assume that externalization has been essential to the unique

---

10 See Donald (1991), (2000) and Koster (2009a) for further discussion.
character of human nature: the capacity to socially share data structures and to be aware of our own thoughts are both central in what makes us human and both seem to be crucially based on the invention of external symbols and the once evolving ability to deal with them. The latter even seems compatible with a certain degree of Darwinian gradualism, rewarding the invention and use of external symbols with reproductive advantage. Even the possession of external symbols without syntax seems an enormous advantage, which might easily have led to a form of co-evolution interlocking the further development of a cultural invention -- external symbols known as words -- with an improved brain capacity to use these symbols.\(^\text{12}\)

Last but not least, the creativity shown in the construction of word meanings is not quite of the same kind as the unconscious rule-governed generation involved in syntax. With the great apes, we share the capacity to assign new functions to arbitrary material objects, like stones or sticks. What we see in the ape behavior is the beginning of what Searle (1995) called “agentive function assignment,” also known as “application” in ordinary English. It differs from standard function assignment in biology, which is completely automatic and determined by the genetic program, as when the lungs develop for the function of breathing and the heart for the pumping of blood. Another example of non-agentive “generation” is the working of the immune system, which can therefore not plausibly be seen as analogous to the agentive creativity involved in the contextual creation of concepts (cf. Chomsky 2000: 65).

Agentive function assignment is like when a stone is used to crack nuts or -- to give a human example-- when we use an arbitrary heavy object as a paperweight. This is a creative process and, with apes at least, useful function assignments are preserved in a group by being visible, so that they can be imitated. Human concept construction is a form of agentive function assignment (no matter how elementary and semi-automatic) and can therefore only be imitated and preserved if observable, i.e., be based on external signs. In general, agentive function assignment is very hard to imagine without external support: to creatively apply something, we must be aware of it so that we can monitor its use. Words are our typical tools for concept construction and the idea of unconscious, internal tools seems to deny the very essence of what tools are, namely things involving agency to some degree. So, as soon as one gives up preformationist prejudice about concepts and realize that they have to be agentively created, the idea of inner symbols loses all attraction or plausibility. In sum, since they are “visible,” minimally for introspection, the words of English qualify as tools of thought, whereas the hypothetical subconscious words of Mentalese do not.

Quite generally, exclusively internal structures, like the organs or the immune system, function in a non-agentive way, while agentive functioning seems to be reserved for things that are external and can be observed and monitored. This even creates a division among body parts, demonstrating once more that innateness is not at issue: we can give an agentive function to our hands or even to our lungs via the mouth (as in wind music), but we cannot creatively apply the kidneys or the heart.

\(^\text{12}\) Note that “cultural invention” is not antithetical to biology. On the contrary, all cultural objects exist as such thanks to human biology, usually without us being aware of it. Paintings, for instance, have an uncontroversial status as cultural objects, cannot be interpreted by animals and often exploit innate faculties, like the capacity to interpret two-dimensional images as representing three-dimensional space. Similarly, cultural objects like words have properties that non-specialists are not aware of, like a valency involving recursion. The same innate structure can have a non-agentive function and an agentive function, for instance the function of the lungs in breathing (non-agentive) and in playing wind instruments (agentive). Similarly, words exploit innate capacities of interpretation, focus them and connect them to sound and combinatorial structure, each of which also involve innate capacities. Since words are invented artifacts, they combine the biological potential involved in an agentive way and therefore as a matter of culture.
Where does Merge fit in into this pattern? Language as we know it combines the symbolic function of our words with the simple recursive mechanism of Merge. Up until the present day, the symbolic function occurs independently of Merge. Traffic signs, for instance, form a symbol system without a syntagmatic dimension and many other examples can be given of the same kind. Taken apart, symbols have minimal linguistic functionality, whereas recursive Merge has not. However, it is easy to see that combining the symbolic function and Merge has made language the fantastic instrument it is. No doubt, this led to even greater biological advantages, leading to further co-evolution, which allowed the brain to ever improve access to, and use of, the integrated system. Words are cultural inventions, which, as external elements, belong to nobody individually but to our shared human legacy. At some point our ancestors must have discovered that it is possible to make the earlier invented, linguistic words combinable by utilizing the properties of non-linguistic Merge, leading to the complex cultural objects that words are now. Words are not labels for anything but form an interface, originally between sound and meaning and, since the integration with Merge, connecting no less than three worlds: outer signs, inner data structures (and their processing) and recursive syntax.

All in all, then, there is no reason to assume that prior to the invention of external symbols, Merge ever was “in touch” with purely internal symbols to form a language of thought. Invention of external symbols, in all likelihood, came first, adding the combinatorial potential of Merge to words later on. This later integration created a “winning” combination that was possibly further rewarded by (co-)evolution. Ultimately, then, I see language, even in the narrowest sense, not as a biological phenomenon but as applied biology, based on a simple but revolutionary invention --external symbols-- that eventually brought originally unrelated resources together. The cultural part of these resources had (and has) linguistic relevance on its own, the others, no matter how biological or innate, had not (at least not prior to their integration with the help of our invention). Given the success of this cultural achievement, it was perhaps further facilitated by natural selection. In other words, the origin of language should not be sought in standard biology but in cultural innovation and the possibility of co-evolution along the nexus culture-biology. Languages of thought and innate concepts are fossils, preformationist relicts of an ancient past, not unlike the tiny little elephants that were once believed to precede the real ones.

---

13 By minimal linguistic functionality I understand the capacity to mediate between what is traditionally called sound and meaning. Unlike Hauser, Chomsky & Fitch (2000), I do not see Merge, no matter how important for language, as the “faculty of language in the narrow sense” (FLN). There is no strictly biological faculty of language. The most elementary linguistic product of cultural-biological co-evolution is our ability to use invented words symbolically and creatively. Merge, from this point of view, is a “third factor principle” in the sense of Chomsky (2007: 3).

14 Chimpanzees do not even come close to having words in this sense. It is therefore thoroughly misleading to say that some apes know 200 or 300 words. Knowing a word minimally involves (i) the ability to project its potential syntactic environment (with recursion), and (ii) the ability to use the word creatively to construct an unlimited set of meanings in context (polysemy). Obviously, apes are able to agentively assign some functions to sticks and stones but not to the information complexes associated with signs.
Bibliography

Anonymous [Cicero]

Aydede, Murat

Baker, Gordon & Peter Hacker

Berlin, Brent & Paul Kay
1969 *Basic Color Terms: Their Universality and Evolution.* Berkeley: University of California Press.

Carey, Daniel

Carruthers, Peter

Chomsky, Noam

Clark, Andy

Clark, Andy & David Chalmers

Cole, David

Donald, Merlin

Fodor, Jerry

Forster, Michael

Haller, Rudolf

Hankins, Thomas
Hauser, Marc, Noam Chomsky & Tecumseh Fitch

Householder, Fred

Humboldt, Wilhelm von

Jenkins, Lyle

Kamesar, Adam

Kay, Lawrence

Koster, Jan

Lenneberg, Eric

Moravcsik, Julius

Nerlich, Brigitte, David Clarke & Zazie Todd

Panaccio, Claude

Panther, Klaus-Uwe & Günter Radden, Eds.

Pustejovsky, James
The University of Chicago Press.

Putnam, Hilary

Reichling, Anton
Saussure, Ferdinand de

Searle, John R.

Stern, Gustaf

Travis, Charles

Westfall, Richard

Wierzbicka, Anna

Wittgenstein, Ludwig

J.Koster@rug.nl

Linguistics Department
University of Groningen
25 Nov 2009