THE (NON-)CIRCULARITY OF DECOMPOSITIONAL SEMANTICS

DAVID ZORNEK

Semantic decomposition is the analysis of linguistic meaning by breaking up the meaning of a linguistic item (e.g. words, sentences) into smaller pieces, ultimately grounding out in some sort of semantic/logical atom that is taken as primitive to the theory. Although decomposition is not without its opponents, notably Fodor and Lepore [12] [14] [13] [15], it remains one of the most popular (if not *the* most popular) approaches to semantics. One of the main complaints that has been issued against decomposition is that it is ultimately incapable of non-circularly specifying the meanings of linguistic items; a more-or-less standard argument for this will be given below. Nevertheless, as Fodor and Lepore have themselves observed [15], there is an overwhelming consensus among cognitive scientists that decomposition is a true representation of how we represent meanings.[1] Given this consensus, it is too hasty a move to abandon decomposition on the grounds that no solution to the circularity problem has yet been given. In fact, it may even be too hasty to regard this sort of circularity a problem at all; it is not that decomposition is inherently circular, but that a circularity arises when we take a formal decompositional theory as a total theory of semantics, rather than an analysis of some particular semantic problem, some aspect of meaning, or an answer to a particular set of questions or concerns. Instead of throwing out decompositional theories entirely, we might instead search for some way of situating our decompositional theories as part of a broader semantic theory that grounds our atoms in a way that alleviates worries of circularity.

First, I will give an overview of a few decompositional theories, in order to familiarize the reader. Then, I will argue that, if taken by themselves, all such theories give a picture of meaning that is either hollow or circular. I will propose that we can provide content and alleviate circularity by mapping atoms into some set of cognitive structures and will identify some necessary conditions that will aid us in identifying *which* set of cognitive structures will be able to serve our ends.

[1] Fodor and Lepore express puzzlement at this consensus, on the grounds that "as far as [they] can tell, there is practically no evidence to support it." However, they imply that they will only accept as evidence an argument that it is impossible to represent meaning in any manner other than decompositionally: "For example, there is no scientific evidence that you *can't* have a word that expresses the concept BREAK$_{\text{TR}}$ unless you have the concept CAUSE. But there ought to be if CAUSE is a constituent of BREAK$_{\text{TR}}$" (emphasis added). But this places too high a standard of justification for the empirical claims that cognitive scientists make; they require only that we *do* represent meaning decompositionally, not that we do so *necessarily*.

1. Decompositional Semantic Theories

Since it would be impossible to give a survey of all possible decompositional theories, I have chosen four samples that will give some indication of the variedness of theories in this class, as well as an idea of the ways in which decompositional theories can vary. We can classify theories according to what sort of linguistic item is being decomposed, according to what sorts of atoms linguistic items are decomposed into, or according to whether they decompose subjective representations of meanings or some objective notion of absolute meaning. [11] decomposes sentence meaning, while the other three decompose lexical meaning. The three kinds of atoms shown below are word senses [11], relations between words [10], and types. The final two theories both decompose lexical meaning into types, but [18] is an account of how an individual mind represents meaning, while [25] is an account of how we should decompose some idealized notion of objective meaning (such as when we talk about *the* meaning of a word as decided by a community of language users, or the meaning of a word in some computational linguistic model)

1.1. **Sentential decomposition into word senses.** Although Fodor, in recent years, has argued against lexical decomposition, an earlier paper with Jerrold Katz—*The Structure of a Semantic Theory*—is regarded as one of the classic endorsements of decomposition. The central problem of a semantic theory, according to Fodor and Katz, is the *projection problem.* Speakers of any natural language have only encountered finitely many sentences of that language, yet they are able to understand and use a potentially infinite number of sentences. This, of course, means that speakers are somehow able to understand and use sentences they have never encountered, which means that a semantic theory must give some solution to the problem of how meanings are generated by language-users for novel sentences. They write:

> This problem requires for its solution rules which project the infinite set of sentences in a way which mirrors the way that speakers understand novel sentences. In encountering a novel sentence the speaker is not encountering novel elements, but only a novel combination of familiar elements. Since the set of sentences is infinite and each sentence is a different concatenation of morphemes, the fact that a speaker can understand any sentence must mean that the way he understands sentences which he has never previously encountered is compositional: on the basis of his knowledge of the grammatical properties and the meanings of the morphemes of the language, the rules which the speaker knows enable him to determine the meaning of a novel sentence in terms of the manner in which the parts of the sentence are composed to form the whole. Correspondingly, we can expect that a system of rules which solves

> the projection problem must reflect the compositional character of the speaker's linguistic knowledge.

So sentential meaning is composed, by means of projection rules, out of some basic elements. But what are these basic elements? The immediate elements out of which a sentence is composed are morphemes; a sentence, according to the above, is a "concatenation of morphemes." But morphemes are not to be regarded as the most basic elements of a semantic theory. Consider the following two sentences involving the morpheme "bill":

(1) The bank teller handed me a large bill.
(2) This duck has a large bill.

It is clear that "bill" is understood differently in each case. One of the fundamental components of any semantic theory, then, is a dictionary which assigns to each morpheme a variety of senses; it is the task of projection rules to select the sense most appropriate to the sentence. A sample (partial) dictionary entry for "bachelor" is given below.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Parameters such as "noun" or "verb" are called *grammatical markers*, while the parameters enclosed in parenthesis above are known as *semantic markers*; parameters inclosed in brackets above are *distinguishers.* Grammatical markers express the syntactic relations between lexical items, while semantic markers express systematic semantic relations between lexical items, e.g. sex-antonymy. Distinguishers express the non-systematic semantic information which is idiosyncratic to a word sense. These distinctions are not *too* important for our purposes here, so I will not dwell on them.

What is important is that the other major component of a semantic theory is a clear specification of how projection rules are generated from the grammar of a language and the dictionary. A semantic theory takes sentences and the dictionary as inputs. From the syntactic structure of the sentence and the grammatical and semantic markers of a dictionary entry, projection rules are generated which enable speakers of the language to interpret each morpheme in the sentence as a path along the dictionary entry terminating in a distinguisher. More precisely, language speakers will interpret morphemes as word senses, which are not displayed in the above model of a dictionary entry. An alternative model, while collapses semantic markers and distinguishers into senses is offered as well; the two models are regarded as equivalent, with the difference being purely notational. That senses can be associated with semantic markers and distinguishers is not to imply that the senses are not primitive; semantic markers and distinguishers are not primitives out of which senses are composed, but are to be understood as different kinds of information associated with the senses, which we can make use of when inferring a sense from the grammatical and semantic relations entailed in a sentence. Fodor and Katz do not give a specification of how projection rules are to be generated; it lies outside the scope of their project. As stated in the first sentence of their paper, “This paper does not attempt to present a semantic theory of a natural language, but rather to characterize the form of such a theory.”

Different semantic phenomena are spelled out in terms of the paths from morphemes to senses. To indicate a few: When a projection rule is generated by the theory which does not terminate in a unique sense, then the sentence from which the projection rule is generated is said to be ambiguous. (Note that the case where a projection rule does not terminate in a sense at all is impossible, given a fully specified lexicon, since a projection rule, by definition, picks out a path terminating in some sense.) When two lexical items share a common path to some sense, they are said to have synonymous senses; when all paths of two lexical items are identical, they are synonyms.

To give a one-sentence summary of the features of [11] that are especially relevant to our purposes here: Sentential meaning is decomposed, by means of projection rules generated according to some semantic theory, into word senses provided in a dictionary.

1.2. **Lexical decomposition into contextual relations.** [10] gives an explicitly non-formal account of lexical decomposition. Cruse’s approach is “descriptive, rather than formalistic,” and as a result some theoretical rigor is sacrificed in favor of descriptive richness, by his own admission. The lack of theoretical rigor leads to a small degree of imprecision of language regarding what are to be taken as atoms. At times, he writes as though the atoms of word meaning are just other words: “An extremely useful model of the meaning of a word, which can be extracted from the contextual relations, is one in which it is viewed as being made up, at least in part, of the meanings of other words.” I do not think we should put too much weight on this way of speaking, however. For one, taking words as the fundamental constituents of word meaning is obviously and viciously circular; this cannot be how Cruse intends to be read. Moreover, he goes on to say that, when one word-meaning participates in the meaning of another, the former is a “†**semantic trait**” of the latter. Semantic traits are given a “†**status**” (discussed below), which is a degree of necessity of contextual relations between words; to say that the meaning of one word is made up of the meanings of other words turns out to be saying that the words exist in a certain type of semantic relation to one another.

At other times, he seems to take word senses as semantic atoms: “Lexical units are those form-meaning complexes with (relatively) stable and discrete semantic properties . . . The meaning aspect of a lexical unit will be termed a †**sense**.” Still more compelling as evidence that he regardes senses as atoms: “It has been argued up to now that although word-meaning is in a sense infinitely variable, nonetheless discrete units—‘atoms’ or ‘quanta’ of sense—can be identified which at least in some respects are stable across contexts, and which are the appropriate basic units for lexical semantics.” However, immediately following this, he writes, “Certain aspects of word-meaning, however, are difficult to reconcile with this view:

particularly awkward are what we shall term †**sense-spectra**." A sense-spectrum is a "sort of semantic continuum," in which senses are actually not discrete units with clearly defined boundaries. Instead, "we shall recognize sense-units along a sense-spectrum–to be called †**local senses**—by their participation in distinct lexical fields (here, to be understood merely as sets of lexical items interrelated by determinate meaning relations. . ." So ultimately, fixed senses are a product of relations between lexical items.

Other kinds of "atoms" are considered, but it turns out that, in all cases, the "atom" is ultimately reduced to the contextual relations in which a lexical item participates. The set of "†**contextual relations**" for a word is defined as "[t]he full set of normality relations which a lexical item contracts with all conceivable contexts." A normality relation is nothing more than an intuition that native language-speakers will tend to have about the oddness of a word's usage in a sentence. If no intuition of oddness is present, the relations of a word to other words in the sentence are normality relations. For example (using some of Cruse's examples):

(1) ? This is a club exclusively for married bachelors.
(2) This is a club exclusively for married men.
(3) ? It's a dog, so it must be a cat.
(4) It's a dog, therefore it's an animal.
(5) ? It's a dog, but it can bark.
(6) It's a dog, but it can't bark.

(2), (4), and (6) are sentences containing only normality relations and therefore are in the set of contextual relations for all words in them. (1), (3) and (5) provoke an intuition of oddness and therefore are not in the set of contextual relations for some word(s) in them.

The statuses (mentioned above) of a word's relations to its semantic traits account for the different ways in which a sentence might be regarded as odd. For example, the †**criterial** status can be diagnosed by recognizing an entailment relation between sentences, as in (4); "It's an animal" is entailed by "It's a dog." The †**excluded** trait is diagnosed in similar fashion, as in (3); "It's a dog" entails "It is not a cat," which is why (3) appears as odd. The †**expected** trait is diagnosed by the "*but*-test", as in (5) and (6); barking is not entailed in being a dog, but it is an expected trait of dogs, which is why (5) is odd, but (6) is not.

Cruse's book is an extensive exploration of a long list of normality relations, their statuses, and diagnostic tests for oddness. A one-sentence summary of the relevant features of [10] is given by Cruse himself: "It is taken as axiomatic in this book that every aspect of the meaning of a word is reflected in a characteristic pattern of semantic normality (and abnormality) in grammatically appropriate contexts."

1.3. **Lexical decomposition into types.** Both [18] and [25] decompose lexical items into types, but in somewhat different ways. [25] will be discussed at length and in great detail later, so here, I will limit myself to only a few comments here. Pustejovsky's type system is derived from Ann Copestake's Linguistic Knowledge Builder (LKB) [9] [8], which is in turn based on Bob Carpenter's *Logic of Typed Feature Structures* [6]. In these systems, types are given in an *inheritance lattice* or inheritance hierarchy in which type-assignment is transitive. Anything $\mathbf{x}$ of type β is also of every type α, where α is an ancestor of β in the lattice; we say that $\mathbf{x}$:β (to be read "$\mathbf{x}$ of type β") *inherits* typing of the ancestors of β. Composition and generation rules are given in a system of typed feature structures which appeal to the inheritance lattice as the source of atoms. In one sentence: Lexical meaning is decomposed, by means of typed feature structures and generative rules, into types provided by an inheritance lattice.

Jackendoff's type system is different from Pustejovsky's in the kinds of problems it is intended to model, but not in the basic relations between types that form the basis of meaning.[2] Jackendoff's primitives are types and tokens, but for our purposes here, we can consider tokens as types. Both are the same kind of atom, with the difference being relational. Given some atom β, we call it a token of some other atom α if α is an ancestor of β in the inheritance lattice; at the same time,

[2] Here, I should point out that Jackendoff's stated view is that word meaning is decomposed into concepts, which is ultimately the view that I will be arguing for. However, his usage of the word "concept" is broader than the usage of "concept" by cognitive scientists, which is the way I will be using the word. For Jackendoff, a concept is ultimately any mental representation, and it includes subjective representations of categories (which is, more or less, what cognitive scientists mean when they talk about concepts), subjective representations of propositions, mental representations of individual objects, and perhaps even the cognitive structural rules which Jackendoff believes to be the source of basic grammatical structure for natural languages. Jackendoff's notion of concept is, ultimately, too broad to do the work that will be required here. Moreover, his notion of concept is entirely subjective. Although, given his goal of understanding subjective representations of word meaning, talking about subjective conceptual representations is justified, we cannot extend this way of thinking to other semantic theories, such as Pustejovsky's, which have the goal of providing some objective notion of *the* meaning of a word, as decided by a whole community of native speakers of a language. Jackendoff is aware of this limitation. He writes, "the information conveyed by language must be about the projected word," where the "projected world" is taken to be synonomous with "experienced world" or "phenomenal world," and where the projected world may or may not be representative of the actual world. [18, pp. 28-29] Reference, on Jackendoff's view, is not to the real world, but to mental projections. [18, p. 36] Where semantics is concerned, it is the relation between type and token "concepts" which is ultimately taken as atomic, so I will regard him as taking types as atoms. To the degree that Jackendoff's types are grounded in concepts, he will not face the circularity or emptiness issues raised here. The issues themselves, however, persist; Jackendoff is able to avoid them precisely because he employs a version of the strategy I will argue for.

β might be regarded as a type relative to some atoms $\gamma_1, \gamma_2, \ldots, \gamma_n$, where the γ_i are decendants of β in the inheritance lattice.[3]

It is perhaps a slight misappropriation of terminology to talk about an inheritance lattice in the context of Jackendoff's system, since he does not mention any such structure in his theory. However, the type-token relations that he does discuss amount to basically the same thing. His discussion of semantic structure begins with a consideration of three sentence types which express different type-token relations:

(1) A dog is a reptile. (Generic categorization)
(2) Clark Kent is Superman. (Token-identity)
(3) Max is a dog. (Ordinary categorization)

Generic categorization, in a sense, expresses the set-inclusion relation: $\{x|x : \textbf{dog}\} \subseteq \{x|x : \textbf{reptile}\}$. Jackendoff abandons set-theoretic notation for (3) and extends this abandonment to (1), instead calling the relationship *IS INCLUDED IN*, but the difference appears to be, where generic categorization is concerned, purely notational. (2) states a relation of wholly different character, the relation *IS TOKEN-IDENTICAL TO.* Both *IS INCLUDED IN* and *IS TOKEN-IDENTICAL TO* are forms of a more fundamental relation *IS AN INSTANCE OF.* The relation *IS AN INSTANCE OF* is simply the inheritance relation.

Jackendoff gives us a list of fundamental types (or "ontological categories") which are regarded as being on even footing with each other, linguistically and conceptually. Via the inheritance relation, all word meanings are given in terms of these types: **things**, **place**, **direction**, **action**, **event**, **manner**, and **amount**. This sort of system should remind the reader of Aristotle's *Categories*, which might be regarded as the earliest exemplar of a type-inheritance theory of lexical semantics.

2. What do the atoms mean?

In all of the above theories (and this is a variant of the definition of a decompositional theory), the meaning of a linguistic unit is inherited from some basic atom. But this is only possible if atoms have meaning, so it makes sense to inquire after the meanings of atoms. Fodor and Katz [11] have told us that sentences get their meaning from word senses, Cruse [10] has said words get their meanings from the contextual relations in which they participate, and Jackendoff [18] and Pustejovsky [25] have said that words get their meanings from types. We now face the question: What is the meaning of a sense/contextual relation/type? Do they have meaning at all?

[3] Some tokens cannot be regarded as types—those which are representations of particular objects, expressed by proper nouns—but discussion of these will be outside the scope of this project, which is concerned only with what Aristotle has called *secondary substances*, i.e. general concepts and collective nouns.

These questions are somewhat loaded. We cannot make full sense of them unless we know what sort of thing meaning is, which is knowledge we will have to do without at present. One response might be that it doesn't makes sense to ask whether a sense *has* meaning, because senses *are* meaning. This kind of response does not address the real concern here, which is that we take ourselves to be talking *about* something when we use words, and except when what we are actually talking about is senses, to say that senses *are* the meaning of our words doesn't capture what words are *about*. When I say, "The sky is clear today," I've said nothing at all about senses. I've said something about the sky. Whatever sort of thing meaning is, it must have something to do with *aboutness*. We might say that "The sky is clear today" is about the sky because the word "sky" refers to the sky. This is on the right track, but some words have meaning without referring to anything, e.g. unicorn, goblin, etc., so reference cannot be a necessary condition of meaning. As a preliminary guiding intuition, we might say: When a word has reference, its reference plays an important role in meaning. In cases where a word has no reference, some story will need to be told about the meanings of words, but I think that what I will argue is compatible with any story that might be offered.

With certainty, we can say that either atoms have meaning in their own right or they do not. David Lewis [19] has given a very compelling argument that atoms do not have meaning in their own right. A decompositional theory provides a set of symbols or "markers," which are essentially a lexicon for the theory. The compositional rules provide a syntax according to which we combine the markers into some sort of structure representing the semantics of whatever linguistic unit is in the domain of the theory. On this picture, a decompositional theory translates natural language into the language of "Semantic Markerese." "But," Lewis writes,

> we can know the Markerese translation of an English sentence without knowing the first thing about the meaning of the English sentence: namely, the conditions under which it would be true. Semantics with no treatment of truth conditions is not semantics. Translation into Markerese is at best a substitute for real semantics, relying either on our tacit competence (at some future date) as speakers of Markerese or on our ability to do real semantics at least for the one language Markerese.

Proponents of decompositional theories are able to get by without providing content to their atoms precisely because they and their audience are both native speakers of some natural language. Notice that the symbolic representations of atoms (markers) are the same as symbolic representations of English words. Fodor's and Katz's senses are represented as phrases in English, Cruse's contextual relations are relations between English words, and Pustejovsky's and Jackendoff's types are represented as English words. Since we are all able to understand these words and phrases, even without a "real semantics" for English (in fact, we might think of "real semantics" as an analysis or spelling-out of how we already understand meanings of a natural language), decompositional theorists rely on fluency to supply content to atoms where none is given within the theory. One of the main goals of this project is precisely to indicate one way in which we can do "real semantics" for Markerese, at the same time situating our decompositional theories within the broader context of "real semantics" for natural language.

If we follow Lewis in thinking that atoms do not have meaning in their own right, then one way of doing "real semantics" for decompositional theories is to augment them with some other structure that provides meaning to the atoms. Absent some structure that provides meaning to atoms, a decompositional theory is, literally, meaningless, and it is difficult to see how semantics, the science whose task it is to identify and explain the behavior and source of meaning can make use of such a theory. Nevertheless, linguists and philosophers alike (many of them at least) seem to agree that decompositional theories play a crucial role in semantics as a whole. This is not to say that a decompositional theory is completely useless until and unless we can find ourselves in possession of a convincing account of the meaning of its atoms. Any decompositional theory comprises only part of a total science of semantics, which studies this or that aspect of meaning or problem involving meaning. Just as I have identified *aboutness* as a feature of meanings without being able to give a full account of what meaning is, decompositional theories can tell us a great deal about meaning without being able to give a full account of the meanings of its atoms. In particular, decompositional theories can tell us a great deal about the structural relations between meanings without supplying *aboutness*.

In fact, it turns out that any decompositional theory is necessarily unable to non-circularly provide content to its atoms. Within the theory, the meanings of atoms can only be decomposed into other atoms; a decompositional semantic theory is only equipped to give meaning in this way. But if the meanings of atoms are given decompositionally, then either they aren't *actually* atoms of the theory, or we will ultimately bottom out with the meanings of some atoms given in terms of themselves. Either way, given that we agree with Lewis that atoms do not have meaning in their own right, we cannot form a coherent view of how a decompositional theory can provide meaning to its own atoms. Therefore, the meanings of atoms must be provided from without. As has been said, the current external source of content for atoms is their identification with English words, which retrieve content from fluency of native speakers. But if this is the actual source of content for the theory, then we have an obvious circularity: Linguistic items supply meaning to atoms, which supply meaning to linguistic items. I do not think there is any need to spell the circularity out in greater detail than this, so instead we need some proposal for what kind of structure might be able to non-circularly give content to the atoms.

My proposal is that we can find some cognitive or mental structure to interpret atoms in order to provide a non-circular external source of meaning. First, it seems to me to be intuitively plausible that understanding and meaning go hand-in-hand.

Joe-on-the-street, who lacks theoretical commitments which might confuse matters by giving rise to undue complications, might tell us that precisely what it means to understand a language is to grasp the meanings well-formed sentences of that language. But we can say a bit more. Our sense of *aboutness* is undoubtedly a mental phenomenon, and therefore we should look inward for an explanation of this sense. Moreover, by identifying atoms with some cognitive structure, we might be able to allow ourselves to account for *aboutness* while simultaneously screening ourselves off from reference in a way that will leave room for meaning without reference where necessary. This type of picture should be able to satisfy referentialists such as Lewis, by allowing our cognitive structures to mediate between words and reference while also being palatable to subjectivists such as Jackendoff, who wish to give a theory of internal semantics in which linguistic units have no meaning beyond what is present in the mind.

Of course, this is not a full argument in favor of identifying atoms with some cognitive structure. However, by identifying a set of conditions that a cognitive structure must meet in order to meet our semantic ends, finding a class of cognitive structures meeting these conditions, and giving a detailed demonstration of how a decompositional theory can be grounded in and/or augmented by the members of this class, I will take myself to have delivered such an argument.

3. What kind of cognitive structure?

By "cognitive structure," I mean something roughly along the lines of what Jackendoff calls "concepts." A couple of necessary conditions have already been implied above. (i) First, in order to avoid circularity, we need a cognitive structure that does not itself appeal to linguistic meaning or language. Second, one of the major things that is missing from the durrent decompositional accounts is that they fail to explain *aboutness*. This seems to be Lewis' major complaint as well, since the theory he ultimately proposes is a version of referentialism, and *aboutness* is an intuition about reference.[4] In order to explain *aboutness*, (ii) we need a cognitive structure that has a clear connection to the real-world states of affairs that linguistic items refer in cases where they do refer.

The third condition I will offer is harder to come by. Since we want a cognitive structure that actually reflects the semantic structure implied by decompositional theories, we must inquire after what kind of structure these theories imply. The remainder of this section constitutes such an inquiry.

The fundamental relation between types in Pustejovsky is explicitly the inheritance relation. Something similar is implied by Cruse's contextual relations. The contextual relations are far too numerous to go through all of them. Nor do I think that all of them imply an inheritance relation; a good portion of them do,

[4] To be completely upfront: My notion of *aboutness* is nothing more than a weakening of referentialist intuitions about meaning, in order to make room for clear-cut cases of meaning without reference.

however. Recall sentence (4) from section 1.2 above: It's a dog, therefore it's an animal. This was called a relation of criterial status: "It's an animal" was said to be *entailed* by "It's a dog." Most of us will want to say that this is a legitimate entailment, although not under the usual notion of entailment provided by first-order logic (not, unless, we introduce some number of auxiliary premises). Rather, what we have is a kind of *semantic entailment.* If we translate (4) into the language of types, we have: $\mathbf{x}{:}\mathbf{dog}{\vdash}\mathbf{x}{:}\mathbf{animal}$. But this is because, by the definition of the word "dog," $\mathbf{dog}{\sqsubseteq}\mathbf{animal}$. The criterial status, at least, which is one of the most basic statuses discussed by Cruse, implies inheritance.

Intuitively, we can get a sense that the same semantic information is exhibited in genus-differentia definition. A genus-differentia construction picks out a sub-class of come broader class, the genus, by means of the differentia. For example, if I take "rational animal" to be a suitable definition for the word "human," we have a genus of "animal" and a differentia of "rational." Now, take a definition of "animal": living being posessing faculties of consciousness and locomotion. Here, we have the genus "living being" and the differentia "posessing faculties of consciousness an locomotion." Now suppose that an object x is a referent of the word "human." Then we know that x is in the class "animal," since the definition of "human" picks out a subclass of "animal." By the definition of "animal," we know that x is also in the class "living being," since "animal" is defined to be a sub-class of "living being." We can translate this directly into the language of type-inheritance. Let $\sqsubseteq$ express the inheritance relation (to be given a precise definition below), such that $\alpha \sqsubseteq \beta$ is to be read as α inherits from β. Then the inheritance relation states that if $\alpha \sqsubseteq \beta$ then $\mathbf{x} : \alpha \vdash \mathbf{x} : \beta$. If we take types to be the atoms of semantic content, such that $\mathbf{x}{:}\beta$ means that $\mathbf{x}$ contains the semantic content of β, then $\mathbf{human}{\sqsubseteq}\mathbf{animal}{\sqsubseteq}\mathbf{living\ being}$ is implied by the above definitions. which in turn means that, $\mathbf{x}{:}\mathbf{human}{\vdash}\mathbf{x}{:}\mathbf{animal}{\vdash}\mathbf{x}{:}\mathbf{living\ being}$.

There is another kind of semantic entailment we will want to look at. Recall sentence (3) from section 1.2, which exemplifies Cruse's excluded status: It's a dog, therefore it's a cat. This sentence provokes an intuition of oddness because the semantic content of "dog" somehow excludes the semantic content of "cat." The semantic content of "dog" and "cat" entails the disjointness of these categories. In type-inheritance notation, we have $\mathbf{x}{:}\mathbf{dog}{\wedge}\mathbf{x}{:}\mathbf{cat}{\vdash} \perp$. More generally, if α and β are atoms such that $\mathbf{x} : \alpha \wedge \mathbf{x} : \beta \vdash \perp$ for any $\mathbf{x}$, we will say $\alpha\#\beta$, where # is the disjointness relation (to be given a precise definition below).

I will now set linguistics aside for a moment and develop the notion of an inheritance network, borrowing some basic definitions from Bob Carpenter [**?**]:

(Inheritance Network) An inheritance network $\mathcal{I}$ is a triple $\langle \mathcal{B}, \sqsubseteq, \# \rangle$ where:

- $\mathcal{B}$: a finite set of basic concepts
- $\sqsubseteq \subseteq \mathcal{B} \times \mathcal{B}$: the basic *inheritance* relation
- $\# \subseteq \mathcal{B} \times \mathcal{B}$: the basic *disjointness* relation

(Inheritance/Disjointness) The *inheritance* relation $\sqsubseteq^* \subseteq \mathcal{B} \times \mathcal{B}$ is the smallest such that:

- $P \sqsubseteq^* P$ (Reflexivity)
- if $P \sqsubseteq Q$ and $Q \sqsubseteq^* R$ then $P \sqsubseteq^* R$ (Transitivity)

The *disjointness* relation $\#^* \subseteq \mathcal{B} \times \mathcal{B}$ is the smallest such that:

- if $P\#Q$ or $Q\#P$ then $P\#^*Q$ (Symmetry)
- if $P \sqsubseteq^* Q$ and $Q\#^*R$ then $P\#^*R$ (Chaining)

In diagramming inheritance networks, I will use directed edges to indicate inheritance and undirected edges to indicate disjointness. When it is useful to distinguish between basic relations and derived relations, I will indicate that a relation is derived with a dotted edge; however, when there is nothing gained by the distinction between basic and derived relations, I will not use dotted edges. For example, given $\mathcal{B} = \{a, b, c, d, e, f\}$, $\sqsubseteq = \{\langle a, d\rangle, \langle d, f\rangle, \langle e, c\rangle, \langle b, e\rangle, \langle e, f\rangle\}$, and $\# = \{\langle b, c\rangle, \langle d, e\rangle\}$, Fig. 1 shows all of the basic relations and some of the derived relations.

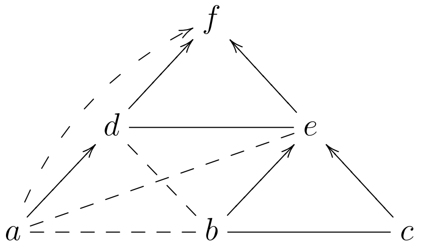

Fig. 1

(Consistent Definition)(cf. [?], *Conjunctive Concept*) A set $D \subseteq \mathcal{B}$ is a *consistent definition* on $\mathcal{B}$ iff:

(1) For all $x, y \in D$, it is not the case that $x\#y$.
(2) For all $x \in D$, $y \in \mathcal{B}$, if $x \sqsubseteq y$, then $y \in D$.

Formally, a consistent defnition is a subset of $\mathcal{B}$ containing some atoms $\alpha_1, \ldots, \alpha_n$, all of the atoms $\beta_1, \ldots, \beta_m$ from which the α_i inherit, and no disjoint atoms. In Fig. 1, the following is a (non-exhaustive) list of consistent definitions:

(i) {a,d,f}
(ii) {d,f}
(iii) {f}
(iv) {b,e,f}

Note that $\{a, d\}$ is not a consistent definition, since it does not contain f, and both a and d inherit from f. Neither is $\{b, e, d, f\}$ a consistent definition, since $e\#d$. $\{b, d, f\}$ is not a consistent definition, since there is a derived disjointness relation between b and d.

For a somewhat less abstract view of things, lets plug in some semantic atoms:

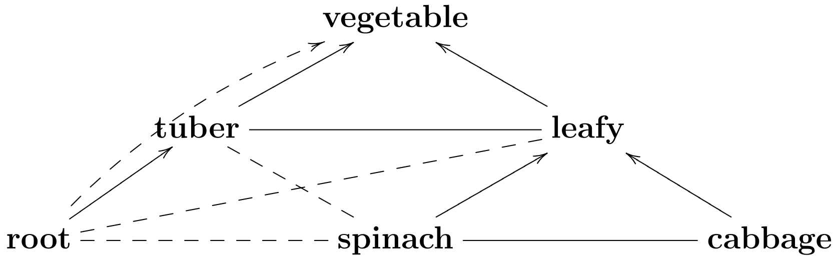

Fig. 2

Now, suppose we want a genus-differentia definition for "potato." Assuming potatoes are the only root tuber (a false assumption, but its falsity shouldn't matter for the present purpose), we might take "root tuber" as a definition, where **tuber** serves as a genus and **root** serves as a differentia. This locates "potato" within the inheritance network, from which we can plainly see that it is part of the meaning of "potato" that any potato is a vegetable. Therefore, a full consistent definition on $\mathcal{B}$ of potato is $D_{\text{potato}} = \{\textbf{root}, \textbf{tuber}, \textbf{vegetable}\}$. However, there is no consistent definition $D = \{\textbf{spinach}, \textbf{tuber}, \textbf{vegetable}\}$ because **tuber** and **spinach** are disjoint. Remember Cruse's diagnostic text for the excluded trait: "It's spinach, therefore it's a tuber" provokes intuitions of oddness, but "It's spinach, therefore it's a leafy vegetable" does not.

(Consistent Definability). We will say that a set S of lexical items is *consistently definable* on $\mathcal{B}$ if and only if there exists a D for every member of S. A lexical item $s \in S$ is said to be *consistently definable* at a point $b \in \mathcal{B}$ if and only if it has a consistent definition D_s containing b. An atom $b \in \mathcal{B}$ is *consistently definable* if and only if there exists a consistent definition D such that $b \in D$. A set of atoms is *consistently definable* if and only if all of its members are consistently definable.

Lemma 3.1. *For* $a, b \in \mathcal{B}$ *such that it is not the case that* $a\#b$, *let* $D_a = \{c | a \sqsubseteq c\}$ *and* $D_b = \{c | b \sqsubseteq c\}$. $D^* = D_a \cup D_b$ *is a consistent definition on* $\mathcal{B}$.

Proof. Because of the reflexivity of $\sqsubseteq$, it is obvious that D^* meets condition (2) above. Suppose (1) does not hold of D^*, i.e. there exists $x, y \in D^*$ such that $x\#y$. Then, by chaining we know that $a\#b$, since for all $x \in D^*$, either $a \sqsubseteq x$ or $b \sqsubseteq x$. But we have already said that it is not the case that $a\#b$, so (1) must hold of D^*. Therefore D^* is a consistent definition on $\mathcal{B}$. □

Lemma 3.2 Let $a, b \in \mathcal{B}$ be such that it is not the case that $a \sqsubseteq b$, and let $\mathcal{F} = \{D | D \text{ is a consistent definition}\}$. If a is consistently definable, then there exists some $D^a_{\neg b} \in \mathcal{F}$ such that $a \in D^a_{\neg b}$ and $b \notin D^a_{\neg b}$.

Proof. Assume a is consistently definable. Then there exists some $D^a \in \mathcal{F}$ such that $a \in D^a$. Either $a\#b$ or not. Suppose $a\#b$. Then $b \notin D^a$, but condition (1) for consistent definition. Suppose it is not the case that $a\#b$. Then there is no relation between a and b, which means that if D^a is a consistent definition, then $D^a_{\neg b} = D^a \setminus \{b\}$ is also a consistent definition. By the definition of $D^a_{\neg b}$, $b \notin D^a_{\neg b}$ and $a \in D^a_{\neg b}$. □

Now, the punchline. I will first state and prove a theorem in its abstract formulation. Once this is done, I will connect it to our more concrete domain.

The genus-differentia inheritance theorem (GDIT):[5] Given $\sqsubseteq$ and $\#$, such that $\sqsubseteq$ is inclusion and $\#$ is disjointness over $\mathcal{B}$, a consistently definable set of atoms, there exists a set $\mathcal{F}$, a set $\mathrm{Rep}_{\mathcal{B}} \subseteq \mathcal{P}(\mathcal{F})$ (where $\mathcal{P}(\mathcal{F})$ is the power set of $\mathcal{F}$), and a function $i : \mathcal{B} \to \mathrm{Rep}_{\mathcal{B}}$ such that $a \sqsubseteq b \Leftrightarrow i(a) \subseteq i(b)$ and $a\#b \Leftrightarrow i(a) \cap i(b) = \emptyset$.

Proof. Let $\mathcal{F} = \{D | D \text{ is a consistent definition on } \mathcal{B}\}$, and let $i : \mathcal{B} \to \mathcal{P}(\mathcal{F})$ be a function such that $i(x) = \{D \in \mathcal{F} | x \in D\}$.

Assume $a \sqsubseteq b$. Suppose $D \in i(a)$. Then $a \in D$, by the definition of i, which means that $b \in D$ by (2). So $D \in i(b)$, by the definition of i. We therefore have $a \sqsubseteq b \Rightarrow i(a) \subseteq i(b)$.

Assume that it is not the case that $a \sqsubseteq b$. Then, by (3.2) there exists some $D^a_{\neg b} \in \mathcal{F}$ such that $a \in D^a_{\neg b}$ and $b \notin D^a_{\neg b}$. Then $D^a_{\neg b} \in i(a)$, but $D^a_{\neg b} \notin i(b)$, which means that $i(a) \not\subseteq i(b)$. So we have: If it is not the case that $a \sqsubseteq b$, then it is not the case that $i(a) \subseteq i(b)$, which contraposits to $i(a) \subseteq i(b) \Rightarrow a \sqsubseteq b$.

We have therefore shown that $a \sqsubseteq b \Leftrightarrow i(a) \subseteq i(b)$, and we now turn to proving that $a\#b \Leftrightarrow i(a) \cap i(b) = \emptyset$.

Assume $a\#b$. Now suppose $i(a) \cap i(b) \neq \emptyset$. Then there exists some $D \in \mathcal{F}$ such that $a \in D$ and $b \in D$. But, by (1), this cannot be the case. So, $a\#b \Rightarrow i(a) \cap i(b) = \emptyset$.

Assume $i(a) \cap i(b) = \emptyset$. Then there exists no $D \in \mathcal{F}$ such that $a \in D$ and $b \in D$. Now suppose it is not the case that $a\#b$. Let D_a, D_b, and D^* be defined as in (3.1). Then D^* is a consistent definition on $\mathcal{B}$. But, by the reflexivity of $\sqsubseteq$, $a \in D_a$ and $b \in D_b$, which means that $a, b \in D^*$. So there is some $D \in \mathcal{F}$ such that $a \in D$ and $b \in D$, a contradiction. Therefore, $i(a) \cap i(b) = \emptyset \Rightarrow a\#b$.

We have therefore shown that $i(a) \cap i(b) = \emptyset \Leftrightarrow a\#b$. □

There are a number of ways of understanding what the GDIT does for us. I take it that these are simply different perspectives on the GDIT, not actually distinct facts that the GDIT entails: What the theorem says is that (a) inheritance networks are isomorphic with some function on consitently definable atoms. When we state things this way, (b) the GDIT can be seen as a representation theorem for inheritance networks. Since mathematically, inheritance networks are a class of posets, a representation theorem can do a lot of theoretical work for us. One very important aspect of this is that (c) we can also view the GDIT as a completeness

[5] Thanks to Cody Roux for helping me to understand the appropriate mathematical framework for stating and proving this theorem.

theorem for inheritance networks in the domain of semantic atoms. Inheritance networks are a complete formalization of the function i, which is a function that maps atoms to the set of all consistent definitions in which they participate, i.e. their *functional role* in the content-system. This is especially appealing for functionalist readers, who will be apt to say that that the content of atoms simply is their functional role in the system. For functionalists, then, the GDIT states (d) the content of any set of consistently definable atoms can be modelled by an inheritance network, and any inheritance network can be interpreted by some set of consistently definable atoms.

There is reason to believe that this connection between semantics and type-inheritance quite generally. If we asked Joe-on-the-street what meaning is, he would likely respond with some variant of "The meaning of a word is its definition." When we look into meaning more closely, of course, we find out that this notion of meaning is inadequate, but we might search for an aspect of meaning that gives rise to the intuition. And the genus-differentia construction of definition undoubtedly implies the inheritance relation. The consistency with which the type-inheritance relations seems to appear in models of meaning should lead us to believe that there is something fundamental about this relation, which should be reflected in any "real semantics for Markerese."

There appears to be a deep relationship between semantic content and abstract order theory, in particular inheritance ordering. Looking at semantic content through the lens of abstract orders, we can see why object-oriented programming is so valuable in computational linguistics: object-oriented languages make heavy use of the inheritance relation. Also, it is a perspective with a great enough degree of mathematical precision to facilitate rigorous investigation of linguistic questions, while being abstract and general enough to accomodate a very wide range of specific formal semantic theories. We can understand semantics *more generally*, apart from the formal semantic systems that were created for the purpose of analyzing particular semantic problems. In fact, abstract order theory is *so* general that it will afford us a framework for comparing two distinct fields of research: cognitive science and formal semantics.[6] Therefore, this perspective allows us to represent facts about decompositional theories in a way that is especially disposed toward a cognitive interpretation of atoms. Quite simply, inheritance networks are the *correct* set of formal tools for answering this question.

Therefore, the third and final necessary condition to be placed on a cognitive structure if it is to be viable as an interpretation of atoms: (iii) It must be modelable by inheritance networks.

[6] Cognitive linguistics is, to a very large degree, a response to formal linguistics. In drawing a connection between cognitive science and formal semantics, it is possible to read this essay as the beginnings of an analysis of the relationship between formal and cognitive linguistics. Although I don't represent the question in this way to myself, I would not argue with such a reading.

In summary and stated differently, we have a minimal set of three necessary conditions that must be met by a cognitive structure, if it is to be a viable interpretation for the atoms of a decompositional theory:

(i′) It must not rely on linguistic meaning in any way for content.
(ii′) It must explain *aboutness.*
(iii′) It must exhibit inheritance behavior.

4. Concepts satisfy the above conditions

Concepts are a kind of cognitive representation; but not all cognitive representations are concepts. Concepts have been called "glue that holds our mental world together" [23] and "units of thought, contitutents of belief and theories" [5]. Representations of particular objects, e.g. this particular coffee mug on my desk, my grandmother, etc., are not concepts. Going back to Plato, concepts have been regarded as *general* in character. A concept is a cognitive representation of some general category. In general, I will follow Greg Murphy's convention of using "concept" to refer to mental representations of categories and using "category" to refer to the real-world members of the category.

Under the *classical view* of concepts, a concept is fully analyzable by a genus-differentia definition. This view was implicit in Plato and Socrates, but was first made explicit by Aristotle. For 2300 years or so, the classical view went unchallenged (and in fact, most early work on concepts assumed and made use of the classical picture), until a debate between C. L. Hull and K. L. Smoke in the 1920s and 1930s [16] [30]. Hull explicitly adopted a form of genus-differentia definition, to which Smoke responded that the simple genus-differentia description failed to capture the complexity of a conceptual representation. Smoke argued that our concepts do not come to closely approximate a single genus-differentia definition, but that they rather become richer over time, so that a full description required stating multiple properties analyzable by more complex genus-differentia relations. Smoke's objections to Hull were not a rejection of genus-differentia analysis, but did foreshadow later work that did lead to the downfall of the classical view.

Eleanor Rosch's work in the 1970s showed us the fundamental inadequacy of the classical view as a total theory of concepts [27] [28] [26]. Her model offered an analysis of concepts by means of *typicality*, which ends up being an analog of Wittgenstein's family resemblances. Where Wittgenstein offered us a family resemblance picture of language, justified on theoretical grounds [31], Rosch followed with a family resemblance picture of concepts, justified on empirical grounds. The overturn of the classical view led to the establishment of new theories: the prototype and exemplar theories of concepts are different ways of cashing out family resemblance in terms of some notion of overall similarity,[7] while the later "theory

[7] According to prototype theories, family resemblance is a matter of overall similarity to some "prototype" or best example of the concept, while exemplar theories capture family resemblance as overall similarity among a whole list of all or many exemplars of the category.

theories" build on the prototype and exemplar theories by representing concepts as consolidations of all of the information and relations associated with category members, not unlike a scientific theory. I will not here be committing to any of the above theories. Rather, it is my aim to sit entirely outside of the debate over prototype vs. exemplar vs. theory theories, instead focusing on the relations between concepts, rather than the internal structure of concepts. What we can see here, however, is that exemplar, prototype and theory theorists alike can all agree that there is something that concepts are *about*, with the difference being over the structure of what concepts are *about*.

Although genus-differentia definition fails as a full description of the internal sctruture of concepts, it seems highly implausible that genus-differentia definition has no relation to concepts whatsoever. After all, if an idea pervades for over 2000 years, it must have *something* going for it; and despite its inadequacy, many fruitful results were obtained under the classical view, in particular by Piaget's influential work from the 1960s [23, pg. 15].

Research in cognitive science has established that concepts are organized more-or-less *hierarchically* [17] [7] [3]. A hierarchy is a network in which members are related by the set-inclusion relation. If one concept C is higher than another concept D in the hierarchy, then the category of which D is a mental representation is a subset of the category represented by C; we say that C is *superordinate* to D and D is *subordinate* to C. It is not too difficult to see how genus-differentia definition naturally falls out of hierarchy. If we take $\phi_D(g, d)$ to be a genus-differentia definition for D, where g is the genus and d is the differentia, then we see that the C-category is a candidate value for g. C is superordinate to D, i.e. C represents a broader category of which D represents a subset, i.e. C is a genus of D. d is some property possessed by all members of the D-category, but no other members of the C-category. Since hierarchy relations are subset relations, they are reflexive and transitive, but not symmetric. But these are the exact properties of the inheritance relation, and so we can regard the conceptual hierarchy as a kind of inheritance ordering. In fact, Murphy has explicitly called the fundamental hierarchy relation *property inheritance* [23].

This is not surprising, given the GDIT. The GDIT states that inheritance networks are isomorphic to a function on consistently definable atoms. But the fact that they are called "atoms" in the GDIT is not essential to the theorem or its proofs. We can just as easily replace the word "atom" with "concept" everywhere it appears in the GDIT and its proof and regard the GDIT as a completeness theorem for inheritance networks in the domain of concepts, instead of in the domain of semantic atoms. Since the GDIT hinges on two things: the notion of consistent definition and on the function i, we should then inquire after how they apply to concepts.

In the domain of semantic atoms, a consistent definition specified the location of an atom within the inheritance lattice; likewise, in the domain of concepts, a consistent definition locates a concept C within the hierarchy, by specifying a set-inclusion path running from C to its most general superordinate. Note that, as multiple consistent definitions are possible for the same atom, and that the function i maps each atom α to the set of all consistent definitions containing α. Then, in the domain of concepts, i maps each concept C to the set of all set-inclusion paths containing C, i.e. the set of all hierarchy relations in which C participates. Therefore, $i(C)$ is the *functional role* of C in the hierarchy.

The analogy between concepts and atoms is quite direct. One *apparent* difference between a conceptual hierarchy and inheritance networks is that, under the way a conceptual hierarchy has been described here, there is not analog of the disjointness relation. I do think that a disjointness relation can be built into the hierarchy; however, this would require engaging in a discussion of the *internal* structure of concepts, entering into the debate over prototype vs. exemplar vs. theory theories, which has been set aside for the purpose of the present discussion. Nevertheless, this is not problematic, since nothing in our definition of inheritance network required that the disjointness relation be a non-empty set; by failing to include disjointness in hierarchy, we have simply confined ourselves to modeling hierarchy as an inheritance network whose disjointness relation is the empty set.

Steven Sloman has argued that concepts are not, in fact, organized hierarchically. In [29], he offered a number of experiments in which subjects were asked to make a logical inference about hierarchy relations. He found that, in a variety of experiments, subjects did not make inferences that conformed to hierarchy. These cases are sort of "fringe" cases where the process of inference is complicated, such as by making category membership probabilistic. By adding probabilistic reasoning into the mix, it makes sense that test subjects would not do as well in making inferences than they otherwise would, since it makes the inferences more complex. In most cases, especially simple ones, subjects do tend to conform to what we would expect when making inferences about hierarchy relations. Therefore, I take it that hierarchy is at least a near enough approximation to how our concepts are actually organized that, for simplicity's sake, we can take the hierarchical view to be more-or-less right.

It now remains to give an argument that concepts meet condition (i′), which is to say that concepts do not rely on linguistic meaning in any way for their existence or description. Now, we need not be concerned with showing this for *all* concepts. We need only show that this holds for a certain class of concepts, which are viable as an interpretation of atoms. Atoms are supposed to be the most basic components of lexical meaning. It seems right, then, that we might look for some most basic set of concepts. In fact, there is a very well-studied set of concepts with exactly this feature.

There is a *basic level of categorization*, which was first studied by Roger Brown [4], later followed by Berlin [3], Malt [20], and others. These researchers have found, over and over, that the most basic level of categorization is neither the most nor the least general level of the hierarchy; instead, it lies somewhere in the middle. The first precise operational description of the basic level was given by Rosch, et al., in a series of experiments from the 1970s [28] [26]. Her original description was found to be problematic [22], but a number of others have been offered. In particular, Murphy and Brownell have offered a metric in which basicality is determined by *informativeness* and *distinctiveness* [24]. Informativeness, a measure of the amount of information associated with the concept, will tend to push concepts down to more subordinate levels, while distinctiveness, a measure of the dissimilarity to other concepts with a common superordinate category or genus, will tend to push concepts up to more superordinate levels. Informativeness and distinctiveness combine into *differentiation*. The most robust advantages for a basic level that maximizes differentiation have been found when using artificial, purely perceptual categories that are unfamiliar to test subjects and are not associated with any known word, which is very strong evidence in support of the non-linguistic or pre-linguistic nature of basic concepts [23, pg. 220].

Of course, we need atoms at more than just the basic level of hierarchy. Unfortunately, the empirical research is not particularly helpful here. Subordinate concepts have not been very well-studied. But, subordinate concepts aren't a major concern at any rate. In the inheritance hierarchies of linguistic theories, there is a definite bottom level, which is not the case with concepts. Moreover, the bottom level usually given in examples of lexical decomposition is the kind of thing that is normally seen at the basic level of concepts, e.g. **book**, **dog**. More important are the superordinate concepts, which are obviously an important part of the inheritance network; we see such general atoms as **event** and **thing**. But research on superordinate concepts is unfortunately linguistic at the outset. Most research into this area is about the linguistic behavior of words that are taken to represent superordinate concepts (such as the fact that the names given to such concepts are normally mass nouns [21]). This research, it mut be pointed out, does not establish that superordinate concepts *are* dependent on some linguistic content, but that the question under consideration is itself linguitic: *What sort of linguistic representation will we tend to give for superordinate concepts?*

Since the empirical research in this area is not helpful for the present purpose, I limit myself to giving a plausible description of how superordinates might not depend on linguistic content. Superordinate concepts are representations of the genera for the basic concepts, united by some feature possessed by all members of the genus. The basic concepts united under a superordinate concept are similarity groups, in much the same way that the basic level concepts are similarity groups between objects; they are simply broader similarity groups. I wager that most of us have recognized similarity between abstract ideas without being able to put in words the way in which they are similar. This experience lends some credence to the idea that superordinate concepts might be obtainable without reliance on linguistic meaning. Ahn and Medin [1] [2] have provided a model of concept formation that does not rely on linguistic content; they do not distinguish between basic and superordinate levels in their research, and it appears that the concepts formed by their subjects are basic level concepts (which makes sense, since the basic level concepts are the ones we would expect them to form most readily), but their model is extendable to the superordinate domain, provided similarity does not necessarily appeal to linguistic meaning.[8]

5. Conclusion

References

[1] Woo-Kyoung Ahn. *A two-stage model of category construction.* Ph.D. dissertation, University of Illinois, Urbana-Champagne, 1990.

[2] Woo-Kyoung Ahn and Douglas L. Medin. A two-stage model of category construction. *Cognitive Science*, 16(1):81—121, 1992.

[3] B. Berlin. *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies.* Princeton University Press, Princeton, NJ, 1992.

[4] R. Brown. How shall a thing be called? *Pychological Review*, 65:14–21, 1958.

[5] Susan Carey. *The Origin of Concepts.* Oxford University Press, Oxford, 2009.

[6] Bob Carpenter. *The Logic of Typed Feature Structures.* Number 32 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, 1992.

[7] A.M. Collins and M.R. Quillian. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8:432–438, 1969.

[8] Ann Copestake. *Implementing Typed Feature Structure Grammars.* CSLI Publications, Stanford, 2002.

[9] Ann Copestake, Antonio Sanfillipo, Ted Briscoe, and Valeria De Paiva. The ACQUILEX LKB: an introduction. In Ted Briscoe, Valeria De Paiva, and Ann Copestake, editors, *Inheritance, Defaults, and the Lexicon*, Studies in Natural Language Processing, pages 148–163. Cambridge University Press, Cambridge, 1993.

[10] D.A. Cruse. *Lexical Semantics.* Cambridge University Press, Cambridge, 1986.

[11] Jerry Fodor and Jerrold Katz. The structure of a semantic theory. *Language*, 39(2):170—210, 1963.

[12] Jerry Fodor and Ernst Lepore. The emptiness of the lexicon: reflections on james pustejovsky's the generative lexicon. *Linguistic Inquiry*, 29(2):269—288, 1998.

[13] Jerry Fodor and Ernst Lepore. Impossible words? *Linguistic Inquiry*, 30(3):445—453, 1999.

[14] Jerry Fodor and Ernst Lepore. Morphemes matter; the continuing case against lexical decomposition. *RuCSS Tech Report*, 34, 2000.

[15] Jerry Fodor and Ernst Lepore. Impossible words: A reply to kent johnson. *Mind & Language*, 20(3):353—356, 2005.

[16] C. L. Hull. Quantitative apects of the evolution of concepts. *Pychological Monographs*, 28, 1920.

[17] B. Inhelder and J. Piaget. *The early grown of logic in the child: classification and seriation.* 1964.

[18] Ray Jackendoff. *Semantics and Cognition.* The MIT Press, Cambridge, MA, 1983.

[19] David Lewis. General semantics. *Synthese*, 22(1):18—67, 1970.

[8] In fact, in the same way that an inheritance ordering naturally falls out of genus-differentia definition, an inheritance ordering will naturally fall out out of Ahn's and Medin's model.

[20] B. C. Malt. Category coherence in cross-cultural perpective. *Cognitive Psychology*, 29:85–148, 1995.

[21] E. M. Markman. Why superordinate category terms can be mass nouns. *Cognition*, 19, 1985.

[22] G. L. Murphy. Cue vailidity and levels of categorization. *Psychological Bulletin*, 91, 1982.

[23] G. L. Murphy. *The Big Book of Concepts*. The MIT Press, Cambridge, MA, 2002.

[24] G. L. Murphy and Brownell H. H. Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 1985.

[25] James Pustejovsky. *The Generative Lexicon*. The MIT Press, 1998.

[26] E. Rosch. Principles of categorization. In E. Rosch and B. B. Lloyd, editors, *Cognition and Categorization*. Erlbaum, Hillsdale, NJ, 1978.

[27] E. Rosch and C.B. Mervis. Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.

[28] C. Simpson, E. Rosch, and Miller R. S. Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2:491–502, 1976.

[29] Steve Sloman. Categorical inference is not a tree: The myth of inheritance hierachies. *Cognitive Psychology*, 35:1—33, 1998.

[30] K. L. Smoke. An objective stuy of concept formation. *Pychological Monographs*, 42, 1932.

[31] Wittgenstein. *Philosophical Investigations*. Blackwell, Oxford, 1953.