

Concepts, Definitions, and Inheritance

Interpreting the atoms of lexical decomposition

by

DAVID ZORNEK

Applying the Conceptual Interpretation of Atoms to the Generative Lexicon

Thus far, we have approached the linguistic issue, i.e. the problem of interpreting the atoms of lexical decomposition, at a very high theoretical level. An abstract mathematical system has been offered to facilitate the interpretation, complete with a theorem that establishes a way in which semantic atoms and concepts can be regarded as identical. Conversely, we have approached the cognitive science issue from a mostly empirical perspective; the mathematical tools introduced here are offered as a model of empirical results about conceptual hierarchy and typicality. Little has been done to *apply* the mathematics to theories of lexical decomposition, in order to gain understanding of linguistic phenomena, and even less has been done to make use of the formalism as a set of theoretical tools to answer questions for cognitive scientists. In this chapter, we will fill in some remaining gaps and demonstrate, via case study, how to *use* inheritance networks to answer real questions. James Pustejovsky’s *Generative Lexicon* [2] is the case we will study.

In ??, I have already explained the basics of the Generative Lexicon. In the present chapter, I will expand on the earlier explanation by looking more closely at two of Pustejovsky’s generative rules. Where linguistics is concerned, Pustejovsky provides a formal model of how word senses are generated according to systematic polysemy, but the model is just that—purely formal. Equipped with the theorem of Chapter 4, we will be able to look “under the hood,” so to speak, in order to gain a better understanding of the cognitive mechanism that enables us to understand novel word senses, at least in the cases where word senses are generated according to Pustejovsky’s rules.

Cognitive scientists often use linguistic stimuli as a way to study concepts. In some of the experiments outlined in Chapter 3, subjects were primed with sentences and asked to respond. But words (and therefore sentences) are only an intermediary for what cognitive scientists are really interested in, which is concepts. There is very good reason to believe that

the same word can stand for different concepts in different contexts (more on this below), and therefore cognitive scientists are faced with the problem of identifying cases where two instances of a word do stand for the same concept, as well as cases where they do not.

For example, suppose a cognitive scientist studying the influence of hierarchy on inference primes subjects with the following sentences:¹

(i) Josh is about to begin reading.

(ii) Josh is holding a novel.

The scientist might then ask test subjects to rate their willingness to infer:

(iii) Josh is about to begin a book.

The idea is that we might test whether subjects are willing to accept this inference on the grounds that **novel** is subordinate to **book** in the hierarchy. But this only tests the influence of hierarchy if the senses of “novel” and “book” are senses that correspond to use of the concepts **novel** and **book**. We shall soon see, by means of a slightly modified example, that there is good reason to think that the sense of “book” in (iii) does not correspond to the concept **book**.

The theorem of Chapter 4 is an isomorphism, i.e. a relation-preserving bijection between two domains. These domains, called isomorphs, are regarded as mathematically indistinguishable, and therefore a main value of an isomorphism is that things learned in one domain can be extended into the other. We can answer the linguistic question by looking at how functional roles of concepts vary under generative rules, and we can answer (part of) the cognitive science question by making use of the formalism offered in decompositional lexical semantics. These questions will be answered simultaneously, not due to careful selection of examples, necessarily, but rather

¹In reality, it would probably be more likely to see cognitive scientists asking subjects to rate their willingness to infer (iii') “Josh is reading a book.” from (ii') “Josh is reading a novel.” I have constructed this artificial example because it is consistent with the problems that will be raised in this chapter. Although somewhat artificial, it is not entirely implausible, since cognitive scientists may want to compare the effects of hierarchy in inferences of different logical form.

because it is a consequence of the theorem that they are in fact the same question approached from different angles.

1 “Josh began the novel.”

Let’s begin with an example.² Consider the sentence “Josh began the novel.” The kinds of things we begin are actions; beginning is always beginning *to do* something. But the object of “began,” in this sentence, is “the novel,” which is not an action. Nevertheless, any native speaker of English will find this to be a perfectly comprehensible sentence. We are, therefore, faced with something of a puzzle: our understanding of the concepts represented by the words in this sentence implies that the sentence should be conceptually incoherent, but the sentence is obviously coherent.

We take the sentence to mean is that Josh is beginning to do something to/with/etc. the novel. There is an implicit action that is being begun, which is not syntactically realized, and “the novel” is the direct object of the verb that refers to this action. For example, two very natural ways of understanding the sentence are “Josh began reading the novel” and “Josh began writing the novel.” The conceptual question is whether the concept **novel** is, in fact, an action concept (i.e., a representation of a category of actions, rather than a category of objects)—contrary to intuitions—or whether “the novel” stands for a more complex concept of **reading the novel** or **writing the novel** in this sentence. The linguistic question is how the semantics of “the novel” can allow it to be coherently used as an argument for “began.” James Pustejovsky has answered the linguistic question for us, by making use of inheritance networks to provide a formal model of lexical decomposition. By trading on the mathematical indistinguishability of the aspects of semantic content that are represented in inheritance networks and conceptual hierarchy, we can make use of his formal answer to the linguistic

²I have modified things slightly in this section from what was including in our toy example above. This was done because the semantics dealt with here will be more intuitive and natural to deal with than what would be required to deal directly with the sentences used above. The basic ideas, however, are the same, and it should not be too difficult for the reader to see how what is said here relates to what has been said above.

question in answering the conceptual question.

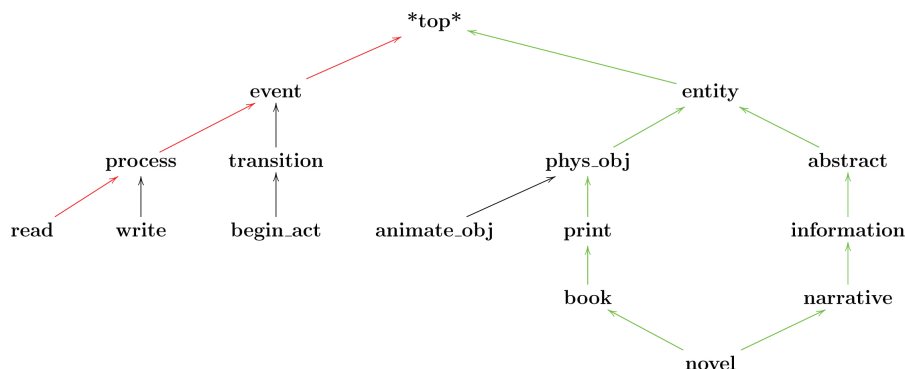
1.1 True Complement Coercion

Recall the following (partial) typed feature structures for “novel” and “begin” from Chapter 2:

$$\left[\begin{array}{l} \mathbf{novel} \\ \mathcal{A} = \left[\begin{array}{l} \text{ARG}_1 = \mathbf{x:book} \\ \text{CONST} = \mathbf{narrative} \\ \text{FORMAL} = \mathbf{book} \\ \text{TELIC} = \mathbf{read} \\ \text{AGENT} = \mathbf{write} \end{array} \right] \\ \mathcal{Q} = \end{array} \right]$$

$$\left[\begin{array}{l} \mathbf{begin} \\ \mathcal{A} = \left[\begin{array}{l} \text{ARG}_1 = \mathbf{animate_obj} \\ \text{ARG}_2 = \mathbf{event}_1 \end{array} \right] \\ \mathcal{E} = \left[\begin{array}{l} \text{E}_1 = \mathbf{transition} \\ \dots \end{array} \right] \\ \mathcal{Q} = \left[\begin{array}{l} \text{FORMAL} = \mathbf{P(event}_1, \mathbf{x)} \\ \text{AGENT} = \mathbf{begin_act(event}_1, \mathbf{x)} \end{array} \right] \end{array} \right]$$

The types given in the typed feature structures are supplied by the following inheritance network (ignore the colors for the time being; we will make use of them shortly):



We can read the argument expectations for “begin” directly off of the typed feature structure. In particular, we see that “begin” expects a direct object of type **event**; this type assignment is a representation of our intuition that actions³ are the kinds of things that can be begun. Since “novel” is not a verb, as has been explained, its argument structure specifies the default behavior of “novel” when it is used as an argument; it behaves as an argument of type **book**. But, as can be read off of the inheritance network, **book** does not inherit from **event**, the sentence “Josh began the novel” entails a violation of the type expectations for arguments of “begin.”

According to Pustejovsky, we resolve the type error by applying a generative rule that searches the qualia structure of “begin” for some type that inherits from **event**.⁴ In fact, we find that the qualia structure of “novel” includes two qualia, i.e. TELIC and AGENT, whose assigned types inherit from **event**; we read off of the graph that both **read** and **write** inherit from **event**. Coercing “novel” to behave as an argument of type **read** or **write** leads us to understand “Josh began the novel” as “Josh began *to read* the novel” or “Josh began *to write* the novel,” respectively.

³The distinction between actions and events is subtle and not important here. There are technical reasons in the GL for making the distinction, but I find the language of actions to be more natural. I will treat the terms as rather interchangeable.

⁴Pustejovsky does offer a formal definition of this rule. His formalization is quite opaque, however, and I think unpacking it will take us needlessly afield from the main point, since the general idea behind the rule can be understood to the degree required here without the formalization.

Pustejovsky’s type coercion is closely related to Barsalou’s *ad hoc* categories [1]. Barsalou makes the point that some categories are well-established in memory, and therefore we have familiar, well-established concepts to represent those categories. However, in practice we make use of novel categories all the time. For example, Barsalou points out that we might be faced with a situation in which we need to make use of the category “things to take from one’s home during a fire.” Surely, we have no well-worn concept to represent this category. (Much pity on anyone who does!) Because we lack such a concept, there seem to be no identifiable features among items in the category (such as children, pets, laptop computers, etc.) that serve to bind the category together by family resemblance. Nevertheless, we are able to form and make use of the necessary concepts “on-the-fly,” when the need arises.⁵

Likewise, we have a word “novel” with a well-worn sense that corresponds to the concept **novel**. Pustejovsky’s type coercion is a model of the mechanism by which we create an *ad hoc* word sense “on-the-fly” when the need arises. The default argument behavior of “novel” is given by the type assignment of the feature ARG₁. Since novel behaves as an argument of type **x:book**, we take this to mean that the default (or “well-worn”) sense of “novel” corresponds to some concept that inherits from **book**, i.e. the concept **novel**. When the argument behavior of “novel” is coerced to a new type, we are then faced with the question of whether the new type (read: concept) is identical with the original one.

We might recall from ?? that identity between functional roles is a necessary condition for identity between concepts. The functional role $i(\mathbf{novel})$ is highlighted in green in the above graph. A consistent definition $D_{\mathbf{read}}$ is highlight in red. We can read directly off of the graph that $D_{\mathbf{read}} \notin i(\mathbf{novel})$. But, by the definition of $i(\mathbf{read})$, we know that $D_{\mathbf{read}} \in i(\mathbf{read})$. Therefore, $i(\mathbf{read}) \neq i(\mathbf{novel})$, and therefore the concept being represented by “novel” in “Josh began the novel” cannot be the concept **novel**. We have answered

⁵Barsalou’s theory of ad hoc categories is, incidentally, predicated largely off of Rosch’s work. I have no particular opinion on how and whether this observation relates to the project of this thesis; I bring it up merely as a point of interest.

the question of whether the concepts corresponding to each sense of “novel” are identical.

2 Another example: “Mary drives a Honda.”

Consider the sentences:

- (i) Mary drives a Honda.
- (ii) Mary bought stock in Honda.

We only have one word “Honda,” which refers to both a category of cars and a car-manufacturing company. We might wonder, however, whether these are both different parts or aspects of the same concept **Honda**. After all, the two domains of reference are clearly related to one another, and in fact the very reason we use the word “Honda” to refer to the category of cars is because they are made by Honda, the car manufacturer. Does “Honda” correspond to a single concept, which is a representation of a category that includes both the cars and the manufacturer? Or do these two uses of “Honda” correspond to distinct concepts?

Some typed feature structures:

$$\left[\begin{array}{l} \mathbf{drive} \\ \mathcal{A} = \left[\begin{array}{l} \text{ARG}_1 = \mathbf{x} : \mathbf{human} \\ \text{ARG}_2 = \mathbf{y} : \mathbf{vehicle} \end{array} \right] \\ \mathcal{E} = \left[\begin{array}{l} \text{E}_1 = \mathbf{e}_1 : \mathbf{process} \\ \text{E}_2 = \mathbf{e}_2 : \mathbf{process} \\ \text{RESTR} = < \circ_\infty \end{array} \right] \\ \mathcal{Q} = \left[\begin{array}{l} \text{FORMAL} = \mathbf{move}(\mathbf{e}_2, \mathbf{y}) \\ \text{AGENT} = \mathbf{drive_act}(\mathbf{e}_1, \mathbf{x}, \mathbf{y}) \end{array} \right] \end{array} \right]$$

$$\left[\begin{array}{l} \text{buy stock} \\ \mathcal{A} = \\ \mathcal{E} = \\ \mathcal{Q} = \end{array} \left[\begin{array}{l} \text{ARG}_1 = \mathbf{x} : \mathbf{human} \text{ARG}_2 = \mathbf{y} : \mathbf{coporation} \\ \text{E}_1 = \mathbf{e}_1 : \mathbf{transition} \\ \text{FORMAL} = \mathbf{transfer_ownership}(\mathbf{e}, \mathbf{x}, \mathbf{stock_share}, \mathbf{y}) \\ \text{TELIC} = \mathbf{gain}(\mathbf{e}, \mathbf{x}, \mathbf{profit}) \\ \text{AGENT} = \mathbf{trade}(\mathbf{e}, \mathbf{x}, \mathbf{y}) \end{array} \right] \right]$$

I have omitted feature structures for the senses of “Honda” since it isn’t clear or important to understand whether and/or how one sense of “Honda” is generated from the other by type coercion. Creating a feature structure for “Honda” would require ironing out these issues and identifying an appropriate generative rule, which would be messy (to say the least) and would take us too far afield from the present task. Everything we need to know can be seen from the following inheritance hierarchy, which provides the types we need:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXX

We know from the argument structures of “drive” and “buy stock” that the sense of “Honda” in (i) must display argument behavior of type **vehicle**, while the sense of “Honda” in (ii) must display argument behavior of type **corporation**. Again, we can read directly off of the graph that $D_{\mathbf{Honda-Co}} \notin i(\mathbf{Honda})$, and therefore **Honda-Co** and **Honda** cannot be the same concept.

3 Some final remarks

It is important to realize that the diagnostic demonstrated here is only a diagnostic for non-identity between concepts. Since the framework of inheritance networks and consistent definitions does not provide a sufficient condition for identity of concepts, we cannot establish that two concepts α and β are identical unless we are able to check that *every possible* $D_\alpha \in i(\beta)$.

This would only be possible, in principle, if the number of possible D_α 's is finite, and there is no reason to suppose that this is true. (I suspect that it is not.)

Also, the Theorem and the conceptual interpretation of semantic atoms have no consequence for GL *qua* formal system. (Or for any theory of lexical decomposition *qua* formal system.) They do have consequences for the meta-theories of lexical decomposition, however, in that they give a deeper account of *why* such theories have the features they do, by interpreting the atoms as concepts. In the case of GL, we are now equipped to understand type coercion as remapping the relation between concepts and words (in cases where the diagnostic justifies this understanding). Also, since cognitive scientists have already developed a robust set of empirical tools for studying conceptual hierarchy, we can now make use of those tools to determine which inheritance networks are admissible in theories of lexical decomposition. For theories of lexical decomposition that dictate certain inheritance relations, we can turn toward cognitive science to determine whether these relations are empirically justified. For theories of lexical decomposition like GL, which leave wide open the question of what our inheritance networks should look like, we now have an empirical means for discovering the right inheritance networks to use. The details of how this can be done are outside the scope of this thesis.

4 Summary

We can make use of the isomorphism to study semantic atoms by looking at concepts and vice versa. Taking GL as a case study allows us to see how the functional role *i* can provide us with diagnostics to determine whether two distinct uses of a word represent the same concept. In making use of diagnostics to establish that the functional roles of the concepts being represented by words are not identical, we can infer that the concepts are not identical. However, the framework of inheritance networks affords us with no known diagnostic to establish that two concepts *are* identical. Such a diagnostic would require possession of a sufficient condition for identity of

concepts, but the tools defined here can only provide us with a necessary condition for identity of concepts.

References

- [1] Lawrence W. Barsalou. Ad hoc categories. *Memory & Cognition*, 11(3):211—227, 1983.
- [2] James Pustejovsky. *The Generative Lexicon*. The MIT Press, 1998.