Redefining the "Level" of the "Word"

Melissa Macpherson

EDS Research
5951 Jefferson NE
Albuquerque, NM 87109-3432
melissa@edsr.eds.com

Abstract

Using dictionaries as a model for lexicon development perpetuates the notion that the level of "the word", as structurally defined, is the right starting place for semantic representation. Difficulties stemming from that assumption are sufficiently serious that they may require a re-evaluation of common notions about lexical representation.

1 Introduction

In a recent paper Boguraev and Levin (1990) point out inadequacies in common conceptions of what a Lexical Knowledge Base (LKB) should be, inadequacies which stem from the assumption that a machine-readable dictionary (MRD) is not only the right source for acquisition of a lexicon but also the right model for its form. Their points about the kinds of generalizations that the standardly conceived LKB does not incorporate, especially if it is built from a dictionary source, are well-taken. They cite a need for the representation of various kinds of dynamic potential—specifically, the capacity to participate in predictable syntactic alternations and regular meaning extensions—which words of particular semantic classes possess in common, and which constitute the basis for lexical productivity. Representation of this kind of dynamic potential is missing from standard dictionaries, they suggest, because it is seen as being more predictable to a native speaker than are specific *-nym* relations. The same representation is missing from most LKBs both because it was not in the source dictionary and because it cannot be encoded in the usual LKB format. Boguraev and Levin conclude that the conception of the LKB must be extended to include both representation of the semantic classes whose members typically participate in these productive alternations and extensions, and sets of inference rules which capture the productive potential of members of those classes. They see this as a partial answer to the open-endedness problem; when a system of the sort they envision is given a new word, we need only specify its semantic class, and we will then have many aspects of both its syntactic and its meaning behavior in hand.

However perhaps a more serious difficulty arising from the adoption of the dictionary as model for the lexicon, and one which Boguraev and Levin do not discuss, is the implicit assumption that the level of the word, as structurally defined, is a semantically well-defined level also.[1] In the sections that follow, I will examine this assumption from several angles. Section 2 reviews the semantic behavior of what are structurally words,

[1] Boguraev and Levin do mention the possibility that generalizations of the sort they require may hold only over partial nodes in a lexical network; this would seem to indicate that semantic classes may be defined in terms of bundles of features, only some of which might be relevant in determining the applicability of a given extension or alternation.

illustrating cases where it seems to take more than one "word" to make a single meaning, and where a single "word" is semantically decomposable into several meaning components. Section 3 discusses the problem of extricating a functor's argument requirements from other aspects of its meaning. In Section 4 the discussion turns to difficulties inherent in distinguishing individual word senses. Sections 5 explores some possibilities for escaping from the dictionary model and solving some of these problems.

2 Morphemes and "Semantemes"

We are led as readers of English, and as dictionary users, to see white space between words as indicating a tangible boundary between units of some sort. We base much of our linguistic analysis on the assumption that there is such a thing as the "word level". If you go "below" this level of structure you are dealing with morphology or with lists of idiosyncratic information; "above" it is the level of syntax. Under this assumption semantics is also of two sorts: the largely classificatory operations of lexical semantics, and the compositional functions of syntactic structure-building. As a working assumption, this division of linguistic knowledge has some merit, and has allowed much useful work to proceed.

But in fact, even on the most superficial examination of real language, it is common to find gross disruptions of this convenient pattern. Unfortunately, lexicalization is an arbitrary process even within a single language, and it is probably only the merest accident when it happens that an atomic chunk of meaning is linked bi-uniquely to an invariant chunk of phonology. The more common run of structure-meaning correspondences are much more complicated.

For one thing, combinations of two or more surface words often function synchronically as more or less absolute semantic units.[2] These collocations may be able to be taken apart for syntactic purposes, but semantically they are not decomposable; they link to meanings which are more than, less than, or different from, the sum of the meanings of their individual parts. The commonly cited examples are phrasal verbs ('hold up', 'kid oneself' 'brim over with') and idiomatic expressions of various sorts ('house of cards', 'lie low', 'eat [someone] out of house and home'). The difficulties here, including locating the boundaries of true idiomaticity, are much discussed, and I will not dwell on them here.

Second, even what are structurally single words can exhibit the kind of compositionality ordinarily expected in syntactic constructions. The most interesting category here is that of lexicalized deverbal compounds, such as *clambake*, *carwash*, *ill-advised*, and *time-saving*. Roeper and Siegel (1978), and Selkirk (1982), among others, have explored the compositional regularities found in this highly productive component of the lexicon, and found that the construction and interpretation of novel verbal compounds is constrained according to criteria whose canonical field of application is syntax.

Both these cross-over phenomena can be accommodated structurally by means of analytical tricks of one sort or another. We have ways of representing verb-particle combinations, for example, so that they are looked up in the same fashion as structurally single words, even while their parts can be separated by theoretically indefinite stretches of the sentence. It is also possible to install vestiges of verbal structure into the representation of

[2]This is true even if we leave aside such troublesome orthographic conventions as the fact that fully lexicalized compounds are often written as separate words.

derivationally deverbal items, so that the regularities which they exhibit in compounding are predicted.

A more severe problem than either of these for locating the "level" of the "word" is exemplified by these examples of perfectly legitimate paraphrase (from Talmy, 1985):

be of interest to somebody = interest somebody

carry out an investigation into something = investigate something

Certainly we would want a natural language understanding system to arrive at the same conclusions from either variant in each pair. It is not at all clear, however, whether we should cause the single-word versions to be represented as decomposable into the phrasal expressions, or analyze the phrasal expressions as stylistically or pragmatically inspired expansions of the more succinct variants, or take some other other tack, such as just finding a way to list both variants as meaning the same thing.[3]

Even when they are not so naturally expanded to phrasal versions, many—maybe most, depending on the granularity of the analysis— of what we consider to be monomorphemic words have flagrantly compositional meanings; the most timeworn examples are *kill* (= 'cause-to-die') and other verbs in which a change-of-state (one meaning component) is brought about by a cause (the other meaning component). Accurate analysis of event descriptions containing such words would seem to require such sub-word semantic analysis, whether it is done as such or as an external listing of presuppositions and entailments of the description as a whole. The trouble, however, is sorting out some set of possible meaning components which is adequate to the compositional task and yet smaller (preferably both in cardinality and in size required for storage) than the set of English words.

Compare, for example, the pairs *kill / murder* and *send_away / banish*. Both pairs incorporate an additional meaning element in the second word of each pair which is at least somewhat similar in both cases; it would mean something like 'with malice aforethought and with the intention of instituting a permanent solution to a perceived problem and probably done with heightened actional intensity'. It is possible that we could find other pairs of lexical expressions whose only difference in meaning is the presence of this additional element in one of them; this would be a good indication that a productive meaning component was at work. But if this component itself can only be expressed in English by means of a very long phrase, we have a considerable difficulty of representation.

It is not that the notions of "word" and "morpheme" make no sense, where the latter is defined as "the minimal meaningful unit of language" (Pei and Gaynor, 1980:140); rather, both are invaluable from the vantagepoint of surface structure. But when we use semantics instead as the starting point, and we define something called a "semanteme",[4] say, as "the minimal unit of meaning that has a sound", then it is not at all clear that the two constructs would meet in the middle and we would arrive at the same set of words again.

Nor is this to say that the expanse of meaning encompassed by a structural word cannot be a unit. It must be true, as Jackendoff (1983, 1990) claims, that linguistic categories correspond fairly directly to conceptual categories. However his formulation of the mapping leaves the status of the structural word itself somewhat indeterminate:

[3] One of the pervasive fallacies inherent in the notion of the "level" of the "word" is that below this everything must be listed, while above this level everything must be representable as sets of rules.

[4] My apologies to anyone who has used this made-up term previously for a different purpose.

> Every content-bearing major *phrasal* constituent of a sentence (S, NP, AP, PP, etc.) corresponds to a conceptual constituent of some major conceptual category. (1990:44) [emphasis added]

We are still left with the question of the proper representation of lexical semantics *in a lexicon* when the borders of what is "lexical" are so fuzzy.

3 Argument Structure and Valence

The problem of the imperfect mapping between structural words and atomic or even bounded meanings extends into what might have been expected to be one of the simpler aspects of representing lexical semantics, that of the exposition of argument structure. Attempts to assign thematic role structure to a wide range of verbs run immediately into two major problems. One is that, as Jackendoff (1990) points out, it may be only part of the meaning of a verb which is responsible for assigning a particular thematic role to a given argument. For instance, the verb *pass* (in the construction *pass the house*) is analyzed as incorporating a Path functor, similar in its function to a surface preposition *by*, which is part of the meaning of the verb. What appears in surface structure therefore as the object of *pass* is better analyzed semantically as the object of the underlying *by*; this is why there is no very satisfactory thematic role label for the object of *pass*. It is also possible for a verb to assign a role to only part of the referent of an argument. For example, the verb *climb*, in *climb the mountain*, has as its Goal not the mountain but only the *top* of the mountain. In general, Jackendoff's arguments that the standard conception of thematic roles requires an accompanying commitment to lexical decomposition are difficult to refute.

At the same time, thematic role labels, or something like them, are necessary for the description of events whether we are willing to commit to lexical decomposition or not. The fact that the set of labels does not seem to be able to be cleanly inventoried at the level of the structural word does not change the fact that some such set of labels is necessary to let us individuate event types, that is, to distinguish events in which 'X verbs Y' from ones in which 'Y verbs X' (Carlson 1984). While we may not be able to arrive at a satisfactorily comprehensive set of labels without complete semantic decomposition, we must have a sufficient set to be able to extract the basic 'who did what to whom' structure from a sentence.

However, finding a boundary between the semantic contribution of the verb itself and that of its expected arguments is often practically impossible. Boguraev and Levin's contention that semantic class determines syntactic behavior, while arguable on other grounds, surely stems from the accurate observation that verbs which describe similar situations naturally *tend* to govern similar numbers of arguments in similar ways.[5] Par-

[5] Unfortunately, the arbitrariness of lexicalization rears its ugly head here also, and the notion of "semantic class" necessary to make this linkage between syntactic behavior and meaning may be somewhat circular; that is, in doubtful cases will we decide whether an item is a change-of-state verb on the basis of whether it undergoes the causative/inchoative alternation? Carlson and Tanenhaus (1988) would not have been able to construct their experiments comparing the effects of thematic structure ambiguity with those of semantic ambiguity, if there did not exist verbs of the same semantic class which do not share the same possibilities for thematic structure alternation. An example pair from their study is *load / fill*:

load / fill the truck (with bricks)

load / *fill the bricks

ticularly where it is true that a whole semantic class of verbs exhibits similar syntactic behavior, it becomes extremely difficult to separate argument structure as an independent phenomenon from the core meaning, if there is such a thing, of the assigning verb. What would it mean to 'eat' without 'eating something'?—it seems pointless even to consider such a question.

Even the definition of individual thematic roles is susceptible to influence from the meanings of expected fillers for those roles; this fact, along with the the difficulty of discovering how these roles are assigned to arguments by word-level functors, may help to explain the persistent lack of consensus on the proper set of thematic roles, as exemplified in Wilkins (1988) and elsewhere. Jackendoff faces one aspect of this problem by removing Agent from the set of primary thematic roles; rather, Agent and Patient are derived role types, designated on a separate Action Tier and superimposed on basic thematic structure through an interaction of the semantics of the verb and the nature of the entities filling potential Agent and Patient role slots. That is, a sentence with *kill* always implies a killer, but it has an Agent just if the killer is an entity capable of Agent-like activity, for instance a person.

Syntax too has an effect on the definition of thematic roles, and that in turn has a pervasive effect on verbal semantics. Talmy (1985) discusses the fact that subjecthood, because of its frequent association with Agenthood, "may tend to confer upon any semantic category expressed in it some initiatory or instigative characteristics." Consequently a Stimulus expressed as Subject may be seen to be acting purposefully or at least actively upon the Experiencer, while where the Experiencer is Subject, "the mental event may be felt to arise autonomously and to direct itself outward toward a selected object." Verbs of experience whose canonical syntactic pattern includes either Stimulus or Experiencer as Subject (for example, *please* and *admire*, respectively) are thus under the influence of this effect in one direction or the other; this may be one source of the ambiguity in the sentence 'He pleases me' and hence in the verb *please*.

4 Senses of words

This blurring of what we would rather see as a clear division between the semantic contributions of functor and argument also complicates the proper treatment of word senses. Compare the following uses of the verb *introduce*:

1. We then introduced a catalyst into the solution.
2. She wants to introduce a little culture into his life.
3. I introduced Pete to green chili / green chili to Pete.[6]
4. I introduced Joe to Pete / Joe and Pete.

Only *load* undergoes the locative alternation.

Talmy (1985) also gives numerous examples of verbs in which shared semantic class does not guarantee identity of valence. An example is the triple *emanate*, *emit*, and *radiate*. *Emanate* must have Figure as subject, *emit* must have Ground as subject, and *radiate* can have either. Likewise *steal*, *rob*, and *rip off*, which all have Agent prominently in focus as subject, differ in their treatment of Figure and Ground. *Steal* makes Figure the object, *rob* makes Ground the object, and *rip off* can put either one in that position.

[6] I do not know why the second version of this sentence is so much less acceptable than the first; I assume it has to do with a violation of focusing constraints.

5. Ronco has introduced several fascinating new products.

6. Joe introduced the coyote into Santa Fe stores.

As the definitions and examples in the following (partial) dictionary entries suggest, it is common to think of each of these sentences as utilizing a different sense of the verb *introduce*.

(from *OALD3E*)

```
  4883820, ..<ent h=introduce>
<def>bring in</def>
<def>bring forward</def>
<ip>introduce into / to</ip>
<ex>introduce a Bill before Parliament</ex>
<def>bring (sth) into use or into operation for the
     first time</def>
<def>cause (sb) to be acquainted with (sth)</def>
<ex>introduce new ideas into a business</ex>
<ex>Tobacco was introduced into Europe from America</ex>
<ex>The teacher introduced his young pupils to the
    intricacies of geometry</ex>
<ip>introduce sb (to sb)</ip>
<def>make (persons) known by name (to one another),
     esp in the usual formal way</def>
<ex>introduce two friends</ex>
<ex>He introduced me to his parents</ex>
<ex>The chairman introduced the lecturer to the
    audience</ex>
<ip>introduce (into)</ip>
<def>insert</def>
<ex>introduce a tube into a wound</ex>
<ex>introduce a subject into a conversation</ex>
......
```

(from *CED Fact Base*)

```
c_DEF([ "introduce",1,1,1 ], "(often foll| by to) to present
                              (someone) by name (to another person)
                              or (two or more people to each other)" ,2 ).
c_DEF([ "introduce",1,1,2 ], "(foll| by to) to cause to experience
                              for the first time" ,1 ).
c_SAMP([ "introduce",1,1,2 ], "to introduce a visitor to beer" ).
c_DEF([ "introduce",1,1,3 ], "to present for consideration or
                              approval, esp| before a legislative
                              body" ,1 ).
c_SAMP([ "introduce",1,1,3 ], "to introduce a draft bill in Congress" ).
c_DEF([ "introduce",1,1,4 ], "to bring in; establish" ,1 ).
c_SAMP([ "introduce",1,1,4 ], "to introduce decimal currency" ).
c_DEF([ "introduce",1,1,5 ], "to present (a radio or television
```

```
                                programme, etc|) verbally" ,1 ).
c_DEF([ "introduce",1,1,6 ], "(foll| by with) to start" ,1 ).
c_SAMP([ "introduce",1,1,6 ], "he introduced his talk with some music" ).
c_DEF([ "introduce",1,1,7 ], "(often foll| by into) to insert or
                                 inject" ,1 ).
c_SAMP([ "introduce",1,1,7 ], "he introduced the needle into his arm" ).
......
```

However the example sentences (1)–(6) seem to illustrate not separate senses but a natural continuum of meanings, in which the contribution of *introduce* itself does not perceptibly change. In sentence (1), *introduce* comes closest to what we would probably call its core meaning, something like 'bring / put something into a new place'. Sentence (2) relocates this action in a different kind of space; both the Theme 'something' and the Goal 'new place' are abstract in this case.[7] In Sentence (3), the Goal location is some sort of recognition space, associated either with the recognizer or the subject matter; the effect is that Pete now knows about green chili. Obviously this sense extension for *introduce* only works if at least one of its arguments is an entity which is capable of recognizing, i.e. a person. Moreover, where the *introduce-ee* is also a person as in (4), (or where the single Object of *introduce* refers to more than one person) the introduction becomes reciprocal; the "actional content" of the verb, in Talmy's words, has been doubled.

In (5), the meaning extension present in (2), (3), and (4) is further specialized so that the implicit Goal location is not only a "recognition space", but is understood to be the recognition space belonging collectively to the individuals making up a market. We make this additional extension, however, not because we are now using a different sense of the verb *introduce*, but simply because of what we know about *products*. That is, our understanding of the type of space in which the introduction is effected is governed by both the core meaning of the verb—we know *introduce* means putting something into some space just as we know that eating means eating something—and the semantic content of its first argument.

In (6) the inference from 'product' to 'market' is reversed. In the most plausible reading of the sentence, *the coyote* refers generically to a decorative motif or to some kind of typically salable item, not to an actual animal. This is true because we know that *Santa Fe stores* represents a particular kind of market, and so we prefer an interpretation in which the thing introduced is something which can typically be marketed.

If we try to decide then what the senses of *introduce* are, as the dictionary model would suggest that we should do, we can take several tacks. We could just take a dictionary's word for the number and definitions of separate senses (leaving aside for the moment the problems of mapping between the different sense divisions of different dictionaries), but then we have clearly missed a generalization about the extensions of meaning possible for the verb *introduce* and others like it. We could distinguish separate senses on the basis of the use in some of the examples of *into* and in others of *to*, but that distinction cross-cuts what are clearly more important semantic distinctions.

On the basis of the thematic structure that the verb *introduce* instantiates, which is more or less the same for all six examples, we might say that it has only one sense.[8]

[7] This is a legitimate and extremely common meaning extension, from physical space to abstract space; in fact it is exactly the sort of regularity that a LKB should include among its lexical rules.

[8] The one problematical construction here would be the one with a plural Theme and no explicit Goal, as in 'I introduced Joe and Pete.' In this case, as noted above, the introduction is reciprocal and therefore the thematic structure would apparently have to be copied over into another, simultaneous action.

Likewise it would decompose in every case to something like

CAUSE (GO (X (FROM (outside-some-space))(TO (inside-some-space))))

Under this analysis it is the characteristics of 'X', the locatum, and of the respective entities which serve as Goal that differentiate senses; sentence (1) makes reference to a real thing and a physical space, (2), (3), and (4) refer to recognizable entities and individual cognitive space, and (5) and (6) refer to something which can be sold and a particular kind of collective cognitive space. In (5), naming a product allows us to infer a market as the implicit Goal, while in (6) knowing that the Goal is a market lets us know that *the coyote* is to be construed as a product. What seems to be happening is what Cruse (1986) calls "modulation" of a word sense by context, except that in this case the "context" is an integral part of the argument structure of the verb.

In any case, it is clear that splitting the verb *introduce* into a set of separate senses *at the word level* will be arbitrary in one way or another.

5 Words as Worlds

The ideal, but computationally outlandish, solution to this problem would be to represent every word in English (not just each functor, but every word) as the union of all the situations in which it could potentially participate, so that all the combinatorial potential of each word would be an inherent part of its inventory of meaning(s). Under this system, a structural word would just be the entry point into a rich representation which might or might not observe that bit of structure further, being built instead around the way the word works in combination with others. All lexical categories, not just the prototypical functors, would have combinatorial preferences, as McCawley (1986) suggests. Semantic composition of functor and arguments in a sentence then would consist of unifying these representations in the most cognitively perspicuous way, in accordance with the bounds imposed by syntax, to create a single coherent scene. In the process of unification, gaps would be filled, construals or views chosen in accordance with constraints instantiated by specific lexical interactions and general cognitive conventions, and aspects of the unifying representations which are contradictory or not currently in focus would fall away. A new semantic combination constructed in this way would be acceptable in inverse proportion to the strain which constructing it placed on these general conventions, and selectional restrictions would simply be statistically based thresholds of strain.

Various efforts have been made toward constructing a realistic version of this ambitious model. Raskin (1986) discusses a system he calls Script-based Semantics, which captures many of the ideas described above. In his system, the appearance of a word in a sentence being analyzed evokes both a set of syntactic senses and one or more scripts, where a script is defined as "a large chunk of semantic information surrounding the word or evoked by it ... a cognitive structure internalized by the native speaker, [representing] knowledge of a small part of the world." These scripts are represented as graphs with lexical nodes and semantic relation arcs, thus bearing a strong resemblance to the Conceptual Structures of Sowa (1984). Under both schemes a set of combinatorial rules then unifies graphs in such a way that ambiguity between word senses disappears and a complete representation for a sentence is constructed.

What is not clear in either system is where the graphs for individual word senses come from. If we were to use the dictionary definitions given earlier as a source of the script(s)

or graph(s) for the verb *introduce*, in combination with the following entries for *product*, we would still have absolutely no basis for combining them to mean what we want them to mean in examples (5) and (6).

(from *OALD3E*)

```
  7296955, ..<ent h=product>
<def>sth produced (by nature or by man)</def>
<ex>farm products</ex>
<ex>the chief products of Scotland</ex>
<def>(maths) quantity obtained by multiplication</def>
<def>(chem) substance obtained by chemical reaction</def>
......
```

(from *CED Fact Base*)

```
c_DEF([ "product",1,1,1 ], "something produced by effort, or
                            some mechanical or industrial process" ,1 ).
c_DEF([ "product",1,1,2 ], "the result of some natural process" ,1 ).
c_DEF([ "product",1,1,3 ], "a result or consequence" ,1 ).
c_DEF([ "product",1,1,4 ], "a substance formed in a chemical
                            reaction" ,1 ).
c_DEF([ "product",1,1,4,1 ], "the result of the multiplication of
                              two or more numbers, quantities, etc" ,1 ).
c_DEF([ "product",1,1,4,2 ], "another name for intersection ,1,3" ,1 ).
......
```

Hand-building such scripts on a sense-by-sense basis for large-scale efforts, on the other hand, would require not only massive redundancy but also an enormous amount of sophisticated-yet-tedious lexical analysis. Not only that, but by pegging these scripts or conceptual graphs to individual word senses, we return to the problem of knowing how to arbitrarily distinguish those senses.

The most promising answer to many of these problems can be found in the approach of collocational semantics, in which the "meaning" of a word consists of sets of observed uses of that word. Practitioners of this approach have sometimes apologized for it (see for instance Pazienza and Velardi, 1989) as a practical but unprincipled substitute for "real" semantics. But in fact the collection, classification, and normalization of collocational patterns may constitute the most realistic way to sidestep various serious difficulties involved in other methods of building a large inventory of lexical knowledge; it might be the way to discover what it "means" for a text to talk about the *introducing a product*. Collocational semantics provides an objective methodology for detecting functional equivalence between expressions, which is really what we want when we attempt to encode synonymy, and for pinpointing functional distinctions, which is what we are trying to do when we attempt to distinguish word sense meanings. Above all, this approach offers promise precisely because it does not depend crucially on the level of the structural word for the definition of semantic units.

As for the operation of a language-understanding system including such a lexical store, obviously allowing the appearance of each word in a sentence to immediately evoke everything we know about its uses would be a computational nightmare. Instead, what is required is a mechanism for incrementally introducing lexical information into sentence

analysis in such a way that just enough, and no more, is present at any given stage, after the fashion of Hirst (1987) with his "Polaroid" words. Under such a methodology several stages of lexical lookup would be necessary. Syntactic patterns, morphologically-derived categorizations, and minimal thematic structure, as discussed above, might be available in the usual way at the first level. The matching of partially disambiguated, intermediate sentence graphs constructed at the first level against an inventory of highly schematic "entry" graphs, and then against successive layers of more completely specified model graphs, would funnel the analysis through the next levels of lookup.[9] A system of semantic classes, along with a set of redundancy rules encoding systematic relations between and alternations possible for members of those classes, would constitute yet another level of lexical knowledge; this level would be important in the organization of the lexicon, but might be accessed directly during lookup only in the case of the appearance of previously unknown words.

How would we construct this kind of multi-level lexical knowledge base? The first level described above can and should be constructed on the basis of information from machine-readable dictionaries. The next set of lookup levels described, in which the structural word is no longer the primary entity but just a handle for indexing graphs, can be built on the basis of knowledge assembled from large volumes of text; the methodology employed by Smadja and McKeown (1990) would be one of many techniques possible for obtaining and organizing the various levels of model graphs. Our idea is that the graded levels of specificity would be constructed by means of successive generalization operations over the most specific set of graphs, extracted more or less directly from text. The final level of knowledge—that is, what Boguraev and Levin have found missing from standard LKBs—must still be installed by linguists. Dictionaries do contain some clues, and certainly the ability to systematically and automatically pull together usage instances can help; but the real work must be done by people.

6 Conclusion

The notion that the level of "the word", as structurally defined, is the appropriate starting place for semantic representation, has been implicit in the design of most lexical knowledge bases. The use of machine-readable dictionaries as a source of lexical knowledge reinforces this notion, at the cost of considerable descriptive loss. This discussion has also revealed a further detrimental assumption fostered by the use of the dictionary as the model for the lexicon: the idea that a single store of lexical knowledge, with a single lookup function and a unified structure, is the necessary mechanism for bringing word meanings into sentence analysis. On the contrary, the multi-faceted semantic behavior of the structural units we call "words" requires a base of knowledge consisting of multiple knowledge stores, each organized in a way that is appropriate to the knowledge it contains and the stage of processing to which it must contribute. The structural word can be used as the indexing key which relates the separate, differently structured stores, but it need not be the basic currency of all of "lexical" semantics.

[9] One or more of these levels of distillation would presumably be equivalent to the single set of canonical graphs envisioned by Sowa (1984).

References

[1] Boguraev, Bran, and Beth Levin (1990). "Models for lexical knowledge bases", in *Electronic Text Research: Proceedings of the Sixth Annual Conference of the UW Centre for the New OED*, pp. 65-78.

[2] Carlson, G. (1984). "Thematic roles and their role in semantic interpretation", *Linguistics 22*, pp. 259–279.

[3] Carlson, G. and Tanenhaus, M. (1988). "Thematic roles and language comprehension', in Wilkins, W. (ed.), *Syntax and Semantics 21: Thematic Relations*. San Diego, CA: Academic Press, pp. 263–291.

[4] *CED Prolog Fact Base*, extracted from the *Collins English Dictionary* (edited by Patrick Hanks, 1979) by E. A. Fox and R. France, 1987.

[5] Cruse, D. A. (1986). *Lexical Semantics*. Cambridge, England: Cambridge University Press.

[6] Hirst, G. (1987). *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge, England: Cambridge University Press.

[7] Jackendoff, Ray (1983). *Semantics and Cognition*. Cambridge, MA: MIT Press.

[8] Jackendoff, Ray (1990). *Semantic Structure*. Cambridge, MA: MIT Press.

[9] McCawley, James D. (1986). "What linguists might contribute to dictionary making if they could get their act together", in Bjarkman, Peter, and Victor Raskin (eds.), *The Real-World Linguist*. Norwood, NJ: Ablex Publishing Co., pp. 3–18.

[10] Pei, Mario A., and F. Gaynor (1980). *A Dictionary of Linguistics*. Totowa, NJ: Littlefield, Adams & Co.

[11] Raskin, Victor (1986). "On possible applications of Script-based Semantics", in Bjarkman, Peter, and Victor Raskin (eds.), *The Real-World Linguist*. Norwood, NJ: Ablex Publishing Co., pp. 19–45.

[12] Roeper, T., and M. E. A. Siegel (1978). "A lexical transformation for verbal compounds", *Linguistic Inquiry 9*, pp. 199–260.

[13] Selkirk, E. O. (1982). *The Syntax of Words*. Cambridge, MA: MIT Press.

[14] *OALD3e*. Electronic version of the *Oxford Advanced Learner's Dictionary of Current English*, edited by A. S. Hornby (1974), prepared in electronic form by F. W. Tompa and Oxford University Press.

[15] Sowa, J. F., 1984. *Conceptual Structures: Information Processing in Mind and Machine*. Reading, MA: Addison-Wesley Publishing Co.

[16] Smadja, F. A., and K. McKeown (1990). "Automatically Extracting and Representing Collocations for Language Generation", *Proceedings of the 28th Annual Meeting of the ACL*, pp. 252–259.

[17] Talmy, Leonard (1985). "Lexicalization patterns: Semantic structure in lexical forms", in Shopen, Timothy (ed.), *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*. Cambridge, England: Cambridge University Press, pp. 57–149.

[18] Velardi, P. and M. T. Pazienza (1989). "Computer aided interpretation of lexical cooccurrences", in *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pp. 185–192.

[19] Wilkins, Wendy (1988). *Syntax and Semantics 21: Thematic Relations*. San Diego, CA: Academic Press.