

## 4. Choices for Lexical Semantics

In this chapter, we discuss the positions taken by ontological semantics on certain current issues and fashions in lexical semantics.

### 4.1 Generativity

A popular idea in lexical semantics has been to make the lexicon “generative.” The reasons for this were both theoretical—to extend the idea of generativity from grammar to the lexicon—and practical—looking for ways of saving effort in acquisition of lexicons through the use of automatic devices. Pustejovsky (1991, 1995) introduces the generative lexicon (GL) in opposition to the lexicons in which all the senses are independent and simply enumerated. In this section, we attempt to demonstrate that, while GL may indeed be superior to an enumerative lexicon based exclusively on corpus-attested usages, it has no special advantages over a well-compiled broad-coverage enumerative lexicon suitable for realistic applications. In particular, the claimed ability of GL to account for the so-called novel word senses is matched by good-quality enumerative lexicons. The difference between generative and enumerative lexicons is, then, reduced to a preference for using some lexical knowledge at runtime or at lexicon acquisition time. The generativity of a lexicon turns out to be synonymous with (striving for) high quality of a lexicon, and GL is a popular but by no means necessarily the only way to achieve this goal.

#### 4.1.1 Generative Lexicon: Main Idea

There are several theoretical and descriptive avenues that the quest for automating lexical acquisition can explore:

- using paradigmatic lexical relations of a lexeme, such as synonymy, antonymy, hyperonymy and hyponymy to specify the lexical meaning of another lexeme; in other words, if a lexical entry is acquired, it should serve as largely filled template for the entries of words that stand in the above lexical relations to the original item;
- using a broader set of paradigmatic relations for the above task, such as the one between an organization and its leader (e.g., *company: commander, department: head, chair, manager*);
- using syntagmatic lexical relations for the above task, for instance, those between an object and typical actions involving it (e.g., *key: unlock, lock,...*).

The paradigmatic and syntagmatic relations among word meanings have been explored and implemented in dictionaries of various sizes and for various languages by the members of the Meaning-Text school of thought since the mid-1960s (Zholkovsky *et al.* 1961, Apresyan *et al.* 1969, 1973, Mel’čuk 1974, 1979). These scholars vastly enriched the list of paradigmatic relations beyond the familiar synonymy, antonymy, and hypo-/hyperonymy. Givón (1967) and McCawley (1968: 130-132) came up with similar ideas independently.

The emphasis in the above work has been on describing meanings of words in terms of those of other words. In the late 1980s and early 1990, the group of scholars in the Aquilex project<sup>48</sup> focused their attention on regular polysemy which explored how to apply paradigmatic and syntagmatic relations to the task of formulating meanings of word senses in terms of other senses of the same lexeme. They proposed to do it with the help of lexical rules that mapped lexicon entries

for new senses to those of the existing senses. Each rule corresponded to a specific relation between senses, such as the well-known “grinding” rule. The idea of regular polysemy ascends to some ideas of Apresyan (1974), where the term was actually introduced. Pustejovsky’s work can be seen as a part and an extension of the Aquilex effort on systematic polysemy. His idea of generativity in the lexicon was, therefore, that

- senses of a polysemous lexical item can be related in a systematic way, with types of such relations recurring across various lexical items;
- by identifying these relations, it is possible to list fewer senses in a lexical entry and to derive all the other senses with the help of (lexical) rules based on these relations.

Our own experience in lexical semantics and particularly in large-scale lexical acquisition since the mid-1980s<sup>49</sup> also confirms that it is much more productive to derive as many entries as possible from others according to as many lexical rules as can be found: clearly, it is common sense that acquiring a whole new entry by a ready-made formula is a lot faster. In the Mikrokosmos implementation of ontological semantics, a set of lexical rules was developed and used to automatically augment the size of an ontological semantic lexicon for Spanish from about 7,000 manually acquired entries to about 38,000 entries (Viegas *et al.* 1996b; see also Section 9.3.3).

#### 4.1.2 Generative vs. Enumerative?

Some claims made about the generative lexicon do not seem essential for its enterprise. In this and the next section, we critically examine them, in the spirit of freeing a good idea of unnecessary ballast.

The generative lexicon is motivated, in part, by the shortcomings of the entity it is juxtaposed against, the enumerative lexicon. The enumerative lexicon is criticized for:

- just listing the senses for each lexical item, without any relations established among them;
- the arbitrariness of (or, at least, a lack of a consistent criterion for) sense selection and coverage;
- failing to cover the complete range of usages for a lexical item;
- inability to cover novel, unattested senses.

Such enumerative lexicons are certainly real enough (most human-oriented dictionaries conform to the description to some extent), and there are quite a few of them around. However, there may be good enumerative lexicons, which cannot serve as foils for the generative lexicon. Enumera-

---

48. The works that we consider as belonging to this approach, some more loosely than others, include Asher and Lascarides (1995), Atkins (1991), Briscoe (1993), Briscoe and Copestake (1991, 1996), Briscoe *et al.* (1990, 1993, 1995), Copestake (1990, 1992, 1995), Copestake and Briscoe (1992), Copestake *et al.* (1994/1995), Johnston *et al.* (1995), Lascarides (1995), Nunberg and Zaenen (1992), Ostler and Atkins (1992), Pustejovsky (1991, 1993, 1995), Pustejovsky and Boguraev (1993), Saint-Dizier (1995), Sanfilippo (1995), Sanfilippo *et al.* (1992).

49. See, for instance, Nirenburg *et al.* (1985, 1987, 1989, 1995), Nirenburg and Raskin (1986, 1987a,b), Raskin (1987a,b, 1990), Carlson and Nirenburg (1990), Meyer *et al.* (1990), Nirenburg and Goodman (1990), Nirenburg and Defrise (1991), Nirenburg and L. Levin (1992), Onyshkevych and Nirenburg (1992, 1994), Raskin *et al.* (1994a,b), Raskin and Nirenburg (1995, 1996a,b), Viegas (1999).

tive lexicons could, in fact, be acquired using a well thought-out and carefully planned procedure based on a sound and efficient methodology, underlain, in turn, by a theory. There is no reason whatsoever to believe that such an enumerative lexicon will be unable to cover exactly the same senses as the generative lexicon, with the relations among these senses as clearly marked.

In ontological semantics, the acquisition methodology allows for the application of lexical rules and other means of automating lexical acquisition both at the time when the lexicon is acquired (acquisition time) and when it is used (runtime). In the generative lexicon, only the latter option is presupposed. Whether, in a computational application, lexical rules are triggered at acquisition or run time may have a computational significance, but their generative capacity, e.g., in the sense of Chomsky (1965: 60), i.e., their output, is not affected by that, one way or another (see Viegas *et al.* 1996b).

#### 4.1.3 Generative Lexicon and Novel Senses

In a modern enumerative approach, such as that used in ontological semantics, text corpora are routinely used as sources of heuristics for establishing both the boundaries of a word sense and the number of different word senses inside a lexeme. However, unlike in the generative lexicon, an ontological semantic lexicon will include senses obtained by other means, including lexical rules: all the applicable lexical rules are applied to all eligible lexical entries, thus creating entries for all the derived senses, many of them not attested in the corpora.

Assuming the potential equivalence of the content of the generative lexicon, on the one hand, and a high-quality enumerative lexicon, on the other, the claimed ability of the generative lexicon to generate novel, creative senses of lexical items needs to be examined more closely. What does this claim mean? What counts as a novel sense? Theoretically, it is a sense which has not been previously attested to and which is a new, original usage. This, of course, is something that occurs rather rarely. Practically, it is a sense which does not occur in a corpus and in the lexicon based on this corpus. Neither the generative lexicon nor a good enumerative lexicon will—or should—list all the senses overtly. Many, if not actually most senses are derived through the application of lexical rules. But even if not listed, such a derived sense is present in the lexicon virtually, as it were, because it is fully determined by the pre-existing domain of a pre-existing lexical rule.

Does the claim of novelty mean that senses are novel and creative if they are not recorded in some given enumerative lexicon? If so, then the object chosen for comparison is low-quality (unless it was built based exclusively on a given corpus of texts) and therefore not the most appropriate one, as one should assume a similar quality of the lexicons under comparison. While the literature is not quite explicit on this point, several contributions (e.g., Johnston *et al.* 1995, Copestake 1995) seem to indicate the implicit existence of a given inferior lexicon or a non-representative corpus against which the comparison is made.

The other line of reasoning for justifying the claim of novelty involves the phenomena of type shifting and type coercion. A creative usage is one which arises from a rule that would overcome a sortal or other incongruity to avoid having to reject an input sentence as ill-formed. But there are rules that make type shifting and type coercion work. They are all pre-existing, not *post-hoc* rules, and, therefore, just as other lexical rules, fully determine, or enumerate (see below), their output in advance.<sup>50</sup>

The above both clarifies the notion of a novel, creative sense as used in the generative lexicon approach and raises serious doubts about its validity. One wonders whether the phenomenon is, really, simply the incompleteness of the corpus and the lexicon relative to which these senses are claimed to be novel. The claim of novelty is then reduced to a statement that it is better to have a high-quality corpus or lexicon than a lower-quality one, and, obviously, nobody will argue with that! A truly novel and creative usage will not have a ready-made generative device for which it is a possible output, and this is precisely what will make this sense novel and creative. Such a usage will present a problem for a generative lexicon, just as it will for an enumerative one or, as a matter of fact, for a human trying to treat creative usage as metaphorical, allusive, ironic, or humorous at text processing time.

#### 4.1.4 Permeative Usage?

Another claimed advantage of the generative lexicon is that it “remembers” all the lexical rules

50. It is perhaps appropriate here to resort to simple formalism to ~~obfuscate~~ clarify this point further. Let  $L$  be the finite set of all lexical rules,  $l$ , used to derive senses from other senses; let  $T$  be the finite set of all type-shifting and coercion rules,  $t$ ; let  $S$  be the (much smaller) set of the senses,  $s$ , of a lexical entry,  $e$ , in the generative lexicon  $G$ . Then,  $G = \{e_1^G, e_2^G, \dots, e_n^G\}$  and  $S_e = \{s_1^e, s_2^e, \dots, s_m^e\}$ . If  $l(s_e)$  is a sense of an entry derived with the help of lexical rule  $l$  and  $t(s_e)$  is a sense of an entry derived with the help of type-shifting, or coercion, rule  $t$ , then let us define  $V_e$  as the set of all such derived senses of an entry:  $V_e = \{v: \forall v \exists s \exists e \exists l \exists t v = l(s_e) \vee v = t(s_e)\}$ . Let  $W^G$  be the set of all derived senses for all the entries in  $G$ :  $W^{GLT} = \{w: \forall w \exists s \exists e \exists l \exists t w = l(s_e) \vee w = t(s_e)\}$ . Finally, let  $U^{GLT}$  be the set of all senses, listed or derived in  $G$ :  $U^{GLT} = W^{GLT} \cup C^G$ , where  $C^G = \{c: \forall c \exists s \exists e c = s_e\}$ .  $U^{GLT}$  represents the weak generative capacity of  $G$ , given the pre-defined sets  $L^G$  and  $T^G$  of lexical and type-shifting rules associated with the generative lexicon.

$U^{GLT}$  is also an enumerable set in the calculus,  $I$ , defined by the set of rules  $L^G \cup T^G$  applied to  $C^G$  in the sense that there is a finite procedure,  $P$ , of (typically, one-step) application of a rule to a listed (or, rarely, derived) sense, such that each element in  $U^{GLT}$  is generated by  $P$  ( $P$  includes zero, or non-application, of any rule, so as to include  $C^G$  in the calculus). In fact,  $U^{GLT}$  is also decidable in the sense that for each of its elements,  $i$ , there is an algorithm in  $I$ , which determines how it is generated, i.e., an algorithm, which identifies, typically, a listed entry and a rule applied to it to generate  $i$ . The set of all those identified pairs of listed entries and rules applied to them determines the strong generative capacity of  $G$ .

Then, the only way the lexicon may be able to generate, i.e., define, a sense  $s$  is if  $s \in U^{GLT}$ . In what way can such a sense,  $h$ , be novel or creative if it is already predetermined in  $G$  by  $L$  and  $T$ ? This notion makes sense only if the existence of a proper subset  $B$  of  $U^{GLT}$  is implied, such that  $h \in U^{GLT} \wedge h \notin B$ . Then, a deficient enumerative lexicon,  $M$ , would list all the senses of  $B$  and not use any lexical or type-shifting rules:  $E = \{e_1^e, e_2^e, \dots, e_k^e\}$ ,  $B = \{b: \forall b \exists s \exists e b = s_e\}$  and  $L^E = T^E = \emptyset$ .

Obviously, if a lexicon,  $O$ , does enumerate some senses and derives others in such a way that every sense in  $U^{GLT}$  is either listed or derived in  $O$  as well, so that both the weak and strong generative capacities of  $O$  equal—or exceed—those of  $U^{GLT}$ , then  $G$  does not generate any novel, creative senses with regard to  $O$ . It also follows that the generative lexicon approach must specify explicitly, about each sense claimed to be novel and creative, relative to what corpus or lexicon is it claimed to be novel and creative.

that relate its senses. We submit, however, that, after all these rules have worked, the computational applications using the lexicon would have no use for them or any memory—or, to use a loaded term, trace—of them whatsoever; in other words, the decidability of the fully deployed set of all listed and derived senses is of no computational consequence.

Pustejovsky (1995: 47-50) comes up with the notion of permeability of word senses to support this lexical-rule memory claim. Comparing *John baked the potatoes* and *Mary baked a cake*, he wants both the change-of-state sense of *bake*, in the former example, and the creation sense in the latter to be present, to overlap, to permeate each other. The desire to see both of these meanings present is linked, of course, to a presupposition that these two meanings of *bake* should not be both listed in the lexicon but rather that one of them should be derived from the other. The argument, then, runs as follows: see these two distinct senses? Well, they are both present in each of the examples above, thus permeating each other. Therefore, they should not be listed as two distinct senses. Or, putting it more schematically: See these two senses? Now, you don't!

Our position on this issue is simple. Yes, there are perhaps two distinct senses—if one can justify the distinction (see Section 9.3.5 for a detailed discussion of methods to justify the introduction of a separate sense in a lexeme). No, they do not, in our estimation, both appear in the same normal (not deliberately ambiguous) usage. Yes, we do think that the two senses of *bake* may be listed as distinct, with their semantics dependent on the semantic properties of their themes. Yes, they can also be derived from each other, but what for and at what price?

We also think the permeative analysis of the data is open to debate because it seems to jeopardize what seems to us to be the most basic principle of language as practiced by its speakers, namely, that each felicitous speech act is unambiguous. It is known that native speakers, while adept at understanding the meaning of natural language text, find it very hard to detect ambiguity<sup>51</sup>. It stands to reason that it would be equally difficult for them to register permeation, and we submit that they actually do not, and that the permeating senses are an artifact of the generative lexicon approach. This, we guess, is a cognitive argument against permeation.

Encouraging permeative usage amounts to introducing something very similar to deliberate ambiguity, a kind of a “sense-and-a-half” situation, into semantic theory, both at the word-meaning level as permeability and at the sentence-meaning level as co-compositionality (see also Sections 3.3-4, and 3.7 below). It seems especially redundant when an alternative analysis is possible. One of the senses of *cake* should and would indicate that it often is a result of baking—there are, however, cold, uncooked dishes that are referred to as cakes as well. No sense of *potato* would indicate that—instead, *potato*, unlike *cake*, would be identified as a possible theme of *cook*, and *cook*

---

51. See, for instance, Raskin 1977a and references there. The reason for the native speaker's unconscious blocking of ambiguity is that it is a complication for our communication and it raises the cognitive processing load (see, e.g., Gibson 1991). So the hearer settles on the one sense which happens to be obvious at the moment (see, again, Raskin 1977a and references there), and blocks the others. There are “non-bona-fide” modes of communication which are based on deliberate ambiguity, such as humor (see, for instance, Raskin 1985c: xiii, 115; cf. Raskin 1992), but functioning in these modes requires additional efforts and skills, and there are native speakers of languages who do not possess those skills without, arguably, being judged incompetent.

will have *bake* and many other verbs as its hyponyms. This analysis takes good care of disambiguating the two senses of *bake* via the meaning of their respective themes, if a need for such disambiguation arises. In fact, it still needs to be demonstrated that it is necessary or, for that matter, possible, to disambiguate between these two senses for any practical or theoretical purpose, other than to support the claim of permeability of senses in the generative lexicon approach. And, circularly, this claim is subordinate to the imperative, implicit in the generative lexicon approach, to reduce the number of senses in a lexicon entry to a preferable minimum of one.

#### 4.1.5 Generative Vs. Enumerative “Yardage”

To summarize, some central claims associated with the generative lexicon seem to juxtapose it against low-quality or badly acquired enumerative lexicons and to disregard the fact that any reasonable acquisition procedure for an enumerative lexicon will subsume, and has subsumed in practice, the generative devices of the generative lexicon.

When all is said and done, it appears that the difference between the generative lexicon and the high-quality enumerative lexicon is only in some relatively unimportant numbers. The former aspires to minimize the number of listed senses for each entry, reducing it ideally to one. The latter has no such ambitions, and the minimization of the number of listed entries in it is affected by the practical consideration of the minimization of the acquisition effort as mentioned in Section 4.1.1 above.

To reach the same generative capacity from a smaller range of listed senses, the generative lexicon will have to discover, or postulate, more lexical rules, and our practical experience shows that this effort may exceed, in many cases, the effort involved in listing more senses, even though each such sense may have to be created from scratch.

A final note on generativity in the lexicon: in an otherwise pretty confused argument against Pustejovsky’s treatment of *bake* and his efforts to reduce the two meanings to one (see Section 1.4 above),<sup>52</sup> Fodor and Lepore (1996) manage to demonstrate that any gain from that reduction will be counterbalanced by the need to deal both with the process of attaining this goal and with the consequences of such treatment of polysemy. We cannot help agreeing with their conclusion, albeit achieved from questionable premises, that “the total yardage gained would appear to be negligible or nil” (*op. cit.*: 7).

## 4.2 Syntax vs. Semantics

The principal choice for lexical semantics with respect to its relations with syntax is whether to assume that each syntactic distinction suggests a semantic difference. Similarly to the situation in compositional semantics (see Section 3.5.2 above), a theoretical proposal in lexical semantics may occasionally claim not to assume a complete isomorphism between the two, but in practice, most lexical semanticists accept this simplifying assumption.

---

52. Coming from a very different disciplinary background, the authors put forward a line of reasoning similar to ours at some times, but also take unnecessary detours and make some unnecessary claims of their own in the process of pursuing totally different goals—different not only from ours but also from Pustejovsky’s. We had a chance to comment on this rather irrelevant review in (Section 2.3.1).

GL's position on this issue, shared with many lexical semanticists, is expressed variously as the dependence of semantics on "basic lexical categories" (Pustejovsky 1995: 1), on "syntactic patterns" and "grammatical alternations" (*ibid.*: 8), as the search for "semantic discriminants leading to the distinct behavior of the transitive verbs" in the examples (*ibid.*: 10), or as an "approach [that] would allow variation in complement selection to be represented as distinct senses" (*ibid.*: 35). The apparently thorough and constant dependence of lexical semantics on syntax comes through most clearly in the analyses of examples.

Thus, introducing a variation of Chomsky's (1957) famous examples of *John is eager to please* and *John is easy to please* and analyzing them in terms of *tough*-movement and the availability or non-availability of alternating constructions (*op.cit.*: 21-22), Pustejovsky makes it clear that these different syntactic behaviors, essentially, constitute the semantic difference between adjectives like *eager* and adjectives like *easy*. We have demonstrated elsewhere (Raskin and Nirenburg 1995) that much more semantics is involved in the analysis of differences between these two adjectives and that these differences are not at all syntax dependent. *Easy* is a typical scalar, whose value is a range on the ease/difficulty scale and which modifies events; *eager* is an event-derived adjective modifying the agent of the event. This semantic analysis does explain the different syntactic behaviors of these adjectives but not the other way around.

One interesting offshoot of the earlier syntax vs. semantics debates has been a recent strong interest in "grammatical semantics," the subset of the semantics of natural languages which is overtly grammaticalized (see, for instance, Frawley 1992—cf. Raskin 1994; in computational-semantic literature, B. Levin 1993 and Nirenburg and L. Levin 1992—who call this field "syntax-driven lexical semantics"—are noteworthy). This is a perfectly legitimate enterprise as long as one keeps in mind that semantics does not end there.

Wilks (1996) presents another example of an intelligent division of labor between syntax and semantics. He shows that up to 92% of homography recorded in Longman Dictionary of Contemporary English (LDOCE 1987) can be disambiguated based exclusively on the knowledge of the part of speech marker of a homograph. Homography is, of course, a form of polysemy and it is useful to know that the labor-intensive semantic methods are not necessary to process all of it. Thus, semantics can focus on the residual polysemy where syntax does not help. In a system not relying on LDOCE, a comparable result may be achieved if word senses are arranged in a hierarchy, with homography at top levels, and if disambiguation is required only down to some nonterminal node in it.

It is also very important to understand that, ideally, grammatical semantics should not assume that each syntactic distinction is reflected in semantic distinction—instead, it should look at grammaticalized semantic distinctions, that is, such semantic phenomena that have overt morphological or syntactic realizations. Consequently, work in grammatical semantics should not consist in detecting semantic distinctions for classes of lexical items with different values on a given syntactic feature (see, for instance, Briscoe *et al.* 1995, Copestake 1995, or Briscoe and Copestake 1996).

The dependence on syntax in lexical semantics may lead to artificially constrained and misleading analyses. Thus, the analysis of the sense of *fast* in *fast motorway* (see, for instance, Lascarides 1995: 75) as a new and creative sense of the adjective as opposed, say, to its sense in *fast runner*,

ignores the important difference between syntactic and semantic modification. It is predicated on the implicit conviction that the use of the adjective with a different noun subcategory—which constitutes, since Chomsky (1965), a different syntactic environment for the adjective—automatically creates a different sense for *fast*. As shown in Raskin and Nirenburg (1995), however, many adjectives do not modify semantically the nouns they modify syntactically, and this phenomenon covers many more examples than the well-known *occasional pizza* or *relentless miles*. Separating syntactic and semantic modification in the case of *fast* shows that it is, in fact, a modifier for an event, whose surface realization can be, at least in English, syntactically attached to the realizations of several semantic roles of, for instance, *run* or *drive*, namely, AGENT in *fast runner*, INSTRUMENT in *fast car*, and LOCATION (or PATH) in *fast motorway*. Throughout these examples, *fast* is used in exactly the same sense, and letting syntax drive semantics distorts the latter seriously. We maintain that it is incorrect and unnecessary either to postulate a new sense of *fast* in this case or to relegate it to “the dustbin of pragmatics” which amounts in practice to justifying never treating this phenomenon at all. In Section 8.4.4 below, we show how ontological semantics proposes to treat this phenomenon as a standard case of semantic ellipsis.

Distinguishing word senses on the basis of differences in syntactic behavior does not seem to be a very promising practice (cf. the Dorr *et al.* 1994/1995 attempt to develop B. Levin’s approach into doing precisely this) also because such an endeavor can only be based on the implicit assumption of isomorphism between the set of syntactic constructions and the set of lexical meanings. But it seems obvious that there are more lexical meanings than syntactic distinctions, orders of magnitude more. That means that syntactic distinctions can at best define classes of lexical meanings, and indeed that is precisely what the earlier incursions from syntax into semantics achieved: rather coarse-grained taxonomies of meanings in terms of a rather small set of features.

### **4.3 Lexical Semantics and Sentential Meaning.**

Semantics as a whole can be said to be the study of lexical and sentential meaning. When the work of lexical semantics is finished, the question arises, how word meanings are combined into the meaning of a sentence. In many lexical semantic approaches, including GL, it is assumed that deriving sentential meaning is the task of formal semantics (see Section 3.5.1 above). The other choice would be developing a dedicated theory for this purpose. An orthogonal choice is whether simply to acknowledge the need for treating sentential meaning as the continuation of work in lexical semantics or actively to develop the means of doing so. In what follows, we will discuss these choices. We will not reiterate here our discussion of sentential semantics in Section 3.5 above: what we are interested in here is how (and, actually, whether) the proposer of a lexical semantic approach addresses its integration with an approach to sentential semantics.

We should mention here, without developing it further—because we consider it unsustainable and because no realistic semantic theory has been put forth on this basis—a possible extreme point of view which denies the existence of lexical semantics. What is at issue is the tension between the meaning of text and word meaning. The compositional approach assumes the latter as a given, but one has to be mindful of the fact that word meaning is, for many linguists, only a definitional construct for semantic theory, “an artifact of theory and training” (Wilks 1996). Throughout the millennia, there have been views in linguistic and philosophical thought that only sentences are real and basic, and words acquire their meanings only in sentences (see, for instance, Gardiner 1951,

who traces this tradition back to the earliest Indian thinkers; Firth 1957, Zvegintzev 1968, and Raskin 1971 treat word meaning as a function of the usage of a word with other words in sentences but without denying the existence of word meaning; Grice 1975).

### 4.3.1 Formal Semantics for Sentential Meaning

In spite of Pustejovsky's (1995: 1) initial and fully justified rejection of formal semantics as a basis of achieving the GL goals with respect to sentential meaning, all that the approach found in contemporary linguistic semantics for dealing with sentential meaning was the analyses of quantifiers and other closed-class phenomena. Formal semantics currently pretty much claims a monopoly on compositionality and extends itself into lexical semantics with regard to a number of mostly closed-class phenomena, especially the quantifiers.<sup>53</sup>

This creates a problem for the GL approach: there is no ready-made semantic theory it can use for the task of sentential meaning representation of a sufficiently fine granularity that NLP requires. This situation is familiar to all lexical semanticists. In GL, Pustejovsky tries to enhance the concept of compositionality as an alternative to standard formal semantics. In the GL approach, compositionality ends up as a part of lexical semantics proper, while formal semantics takes over in the realm of sentential meaning.

As we argued in Section 3.5.1 above, however, formal semantics is not necessarily the best candidate for the theory of sentential meaning. It is a direct application of mathematical logic to natural language. All the central concepts in logic are taken from outside natural language, and the fit between these concepts and the language phenomena is not natural. Formal semantics, thus, follows a method-driven approach, exploring all the language phenomena to which it is applicable and by necessity ignoring the rest. An alternative to such an approach is an investigation of all relevant language phenomena, with methods and formalisms derived for the express purpose of such an investigation (see Nirenburg and Raskin 1999).

### 4.3.2 Ontological Semantics for Sentential Meaning

These latter problem-driven approaches include conceptual dependency (e.g., Schank 1975), preference semantics (Wilks 1975a) and our own ontological semantics (e.g., Onyshkevych and Nirenburg 1995). In ontological semantics, to recapitulate briefly, sentential meaning is defined as an expression, text meaning representation, obtained through the application of the sets of rules for syntactic analysis of the source text, for linking syntactic dependencies into ontological dependencies and for establishing the meaning of source text lexical units. The crucial element of this theory is a formal world model, or ontology, which also underlies the lexicon and is thus the basis of the lexical semantic component. The ontology is, then, the metalanguage for ontological lexical semantics and the foundation of its integration with ontological sentential semantics.

We are not ready to go as far as claiming that lexical semantics and sentential semantics must always have the same metalanguage, but we do claim that each must have a metalanguage. We

---

53. See, for instance, Lewis (1972), Parsons (1972, 1980, 1985, 1990), Stalnaker and Thomason (1973), Montague (1974), Dowty (1979), Barwise and Perry (1983), Keenan and Faltz (1985), Partee *et al.* (1990), Chierchia and McConnell-Ginet (1990), Cann (1991), Chierchia (1995), Hornstein (1995), Heim and Kratzer (1998).

know that not all approaches introduce such a metalanguage explicitly (see 2.4.1.3 and 2.4.5, especially Table 2, above). In lexical semantics, this means, quite simply, that every theory must make a choice concerning the conceptual status of its metalanguage. The introduction of an explicit ontology is one way to make this choice. Other choices also exist, as exemplified by the GL approach, in which “nonlinguistic conceptual organizing principles” (Pustejovsky 1995: 6) are considered useful, though remain undeveloped.

We believe that the notational elements that are treated as theory in GL can be legitimately considered as elements of semantic theory only if they are anchored in a well-designed model of the world, or ontology. Without an ontology, the status of these notions becomes uncertain, which may license an osmosis- or emulation-based usage of them: a new feature and certainly a new value for a feature can always be expected to be produced if needed, the *ad hoc* way. A good example of this state of affairs is the basic concept of qualia in GL.

The qualia structure in GL consists of a prescribed set of four roles with an open-ended set of values. The enterprise carries an unintended resemblance to the type of work fashionable in AI NLP in the late 1960s and 1970s: proposing sets of properties (notably, semantic cases or case roles) for characterizing the semantic dependency behavior of argument-taking lexical units (see, e.g., Bruce 1975). That tradition also involved proposals for systems of semantic atoms, primitives, used for describing actual meanings of lexical units. This latter issue is outside the sphere of interest of GL, though not, in our opinion, of lexical semantic theory.

The definitions of the four qualia roles are in terms of meaning and share all the difficulties of circumscribing the meaning of case roles. Assignment of values to roles is not discussed by Pustejovsky in any detail, and some of the assignments are problematic, as, for instance, the value “narrative” for the constitutive role (which is defined as “the relation between an object and its constitutive parts” (1995: 76)) for the lexicon entry of *novel* (*ibid.*: 78). The usage of ‘telic’ has been made quite plastic as well (*ibid.*: 99-100), by introducing ‘direct’ and ‘purpose’ telicity, without specifying a rule about how to understand whether a particular value is direct or purpose.

One would expect to have all such elements as the four qualia specified explicitly with regard to their scope, and this is, in fact, what theories are for. What is the conceptual space, from which the qualia and other notational elements of the approach emerge? Why does GL miss an opportunity to define that space explicitly in such a way that the necessity and sufficiency of the notational concepts introduced becomes clear—including, of course, an opportunity to falsify its conclusions on the basis of its own explicitly stated rules?<sup>54</sup> An explicit ontology would have done all of the above for GL.

To be fair, some suggestions have been made for generalizing meaning descriptions in GL using the concept of lexical conceptual paradigms (e.g., Pustejovsky and Boguraev 1993, Pustejovsky and Anick 1988, Pustejovsky *et al.* 1993). These paradigms “encode basic lexical knowledge that is not associated with individual entries but with sets of entries or concepts” (Bergler 1995: 169). Such “meta-lexical” paradigms combine with linking information through an associated syntactic schema to supply each lexical entry with information necessary for semantic processing. While it

---

54. An examination of the Aristotelian roots of the qualia theory fails to fill the vacuum either.

is possible to view this simply as a convenience device that allows the lexicographer to specify a set of constraints for a group of lexical entries at once (as was, for instance, done in the KBMT-89 project (Nirenburg *et al.* 1991), this approach can be seen as a step toward incorporating an ontology.

Bergler (1995) extends the amount of these “meta-lexical” structures recognized by the generative lexicon to include many elements that are required for actual text understanding. Thus, she presents a set of properties she calls a “style sheet,” whose genesis can be traced to the “pragmatic factors” of PAULINE (Hovy 1988). She stops short, however, of incorporating a full-fledged ontology and instead introduces nine features, in terms of which she describes reporting verbs in English. A similar approach to semantic analysis with a set number of disjoint semantic features playing the role of the underlying meaning model was used in the Panglyzer analyzer (see, for instance, Nirenburg 1994).

There is a great deal of apprehension and, we think, miscomprehension about the nature of ontology in the literature, and we addressed some of these and related issues in Section 2.6.2.2 above, Chapter 5 below as well as in Nirenburg *et al.* (1995). One recurring trend in the writings of scholars from the AI tradition is toward erasing the boundaries between ontologies and taxonomies of natural language concepts. This can be found in Hirst (1995), who acknowledges the insights of Kay (1971). Both papers treat ontology as the lexicon of a natural (though invented) language, and Hirst objects to it, basically, along the lines of the redundancy and awkwardness of treating one natural language in terms of another. Similarly, Wilks *et al.* (1996: 59) see ontological efforts as adding another natural language (see also Johnston *et al.* 1995: 72), albeit artificially concocted, to the existing ones, while somehow claiming its priority.

By contrast, in ontological semantics, an ontology for NLP purposes is seen not at all as a natural language but rather as a language-neutral “body of knowledge about the world (or a domain) that a) is a repository of primitive symbols used in meaning representation; b) organizes these symbols in a tangled subsumption hierarchy; and c) further interconnects these symbols using a rich system of semantic and discourse-pragmatic relations defined among the concepts” (Mahesh and Nirenburg 1995: 1; see also Section 7.1). The names of concepts in the ontology may look like English words or phrases but their semantics is quite different and is defined in terms of explicitly stated interrelationships among these concepts. The function of the ontology is to supply “world knowledge to lexical, syntactic, and semantic processes” (*ibid*), and, in fact, we use exactly the same ontology for supporting multilingual machine translation.

An ontology like that comes at a considerable cost—it requires a deep commitment in time, effort, and intellectual engagement. It requires a dedicated methodology based on a theoretical foundation (see Chapter 5 below). The rewards, however, are also huge: a powerful base of primitives, with a rich content and connectivity made available for specifying the semantics of lexical entries, contributing to their consistency and non-arbitrariness.<sup>55</sup>

### 4.3.3 Lexical Semantics and Pragmatics

In much of lexical and formal semantics, three major post-syntactic modules are often distinguished, though not at all often developed: lexical semantics, compositional semantics and pragmatics. Pragmatics is variously characterized as “commonsense knowledge,” “world knowledge,”

or, even more vaguely, context. It is perceived as complex, and, alternatively, not worth doing or not possible to do, at least for now (see, for instance, Pustejovsky 1995: 4, Copestake 1995). Occasionally, brief incursions into this *terra incognita* are undertaken in the framework of syntax-driven lexical semantics (see, for instance, Asher and Lascarides 1995, Lascarides 1995) in order to account for difficulties in specific lexical descriptions. Pragmatic information is, then, added to corresponding lexical entries to explain the contextual meanings of words. Curiously, pragmatics is, on this view, related to lexical semantics but not to sentential semantics.

An important point for us in understanding this position is that scholars firmly committed to formality (and formalism, see Section 2.4.1.4) felt compelled to venture into an area admittedly much less formalizable, because without this, it would not have been possible to account for certain lexical semantic phenomena. The next logical step, then, would be to come up with a comprehensive theory and methodology for combining all kinds of pertinent information with lexical meaning and characteristics of the process of deriving sentential meaning. We believe that continued reliance on truth-conditional formal semantics as a theory of sentential meaning would make such an enterprise even more difficult than it actually is.

Ontological semantics does not see any reason even to distinguish pragmatics in the above sense from deriving and representing meaning in context—after all, any kind of language- or world-related information may, and does, provide clues for semantic analysis. The sentential semantics in our approach is designed to accommodate both types of information. As Wilks and Fass (1992: 1183) put it, “knowledge of language and the world are not separable,<sup>56</sup> just as they are not separable into databases called, respectively, dictionaries and encyclopedias” (see also Nirenburg 1986). Practically, world knowledge, commonsense knowledge, or contextual knowledge, is recorded in the language-independent static knowledge sources of ontological semantics—the ontology and the Fact DB. The main question that an ontological semanticist faces with respect to that type of knowledge is not whether it should be recorded but rather how this is done best.

#### 4.4 Description Coverage

In principle, any theory prefers to seek general and elegant solutions to an entire set of phenomena in its purview. In practice, lexical semantics has to choose whether account only for those phenomena that lend themselves to generalization or to hold itself responsible for describing the entire set of phenomena required by a domain or an application.

GL shares with theoretical linguistics the practice of high selectivity with regard to its material.

---

55. Ontological semantic lexicons fit Fillmore and Atkins' (1992: 75) vision of an ideal dictionary of the future: “...we imagine, for some distant future, an online lexical resource, which we can refer to as a “frame-based” dictionary, which will be adequate to our aims. In such a dictionary,... individual word senses, relationships among the senses of the polysemous words, and relationships between (senses of) semantically related words will be linked with the cognitive structures (or ‘frames’), knowledge of which is presupposed by the concepts encoded by the words.”

56. The very existence of the distinction between lexical and pragmatic knowledge, the latter equated with “world knowledge” or “encyclopedic knowledge,” has been a subject of much debate (see Raskin 1985a,b, 1985c: 134-135, 2000; for more discussion of the issue—from both sides—see Hobbs 1987, Wilensky 1986, 1991, Peeters 2000; cf. Wilks 1975a, 1975b: 343).

This makes such works great fun to read: interesting phenomena are selected; borderline cases are examined. The tacit assumption is that the ordinary cases are easy to account for, and so they are not dealt with. As we mentioned elsewhere (Raskin and Nirenburg 1995), in the whole of transformational and post-transformational semantic theory, only a handful of examples has ever been actually described, with no emphasis on coverage. Lexical semantics is largely vulnerable on the same count.

Large-scale applications, on the other hand, require the description of every lexical-semantic phenomenon (and a finer-grained description than what can be provided by a handful of features, often conveniently borrowed from syntax), and the task is to develop a theory for such applications underlying a principled methodology for complete descriptive coverage of the material. The implementation of any such project would clearly demonstrate that the proverbial common case is not so common: there are many nontrivial decisions and choices to make, often involving large classes of data.

Good theorists carry out descriptive work in full expectation that a close scrutiny of data will lead to, often significant, modifications of their *a priori* notions. The task of complete coverage forces such modifications on pre-empirical theories. Thus, the need to describe the semantics of scalars forced the development of the previously underexplored phenomenon of scale, e.g., *big* (scale: SIZE), *good* (scale: QUALITY), or *beautiful* (scale: APPEARANCE) in the study of the semantics of adjectives.

There are many reasons to attempt to write language descriptions in the most general manner—the more generally applicable the rules, the fewer rules need to be written; the smaller the set of rules (of a given complexity) can be found to be sufficient for a particular task, the more elegant the solution, etc. In the area of the lexicon, for example, the ideal of generalizability and productivity is to devise simple entries which, when used as data by a set of syntactic and semantic analysis operations, regularly yield predictable results in a compositional manner. To be maximally general, much of the information in lexical entries should be inherited, based on class membership or should be predictable from general principles.

However, experience with NLP applications shows that the pursuit of generalization for its own sake promises only limited success. In a multitude of routine cases, it becomes difficult to use general rules—Briscoe and Copestake (1996) is an attempt to alleviate this problem through non-linguistic means. The enterprise of building a language description maximizing the role of generalizations is neatly encapsulated by Sparck Jones: “We may have a formalism with axioms, rules of inference, and so forth which is quite kosher as far as the manifest criteria for logics go, but which is a logic only in the letter, not the spirit. This is because, to do its job, it has to absorb the *ad hoc* miscellaneity that makes language only approximately systematic” (1991, p.137).

This state of affairs, all too familiar to anybody who has attempted even a medium-scale description of an actual language beyond the stages of morphology and syntax, leads to the necessity of directly representing, usually in the lexicon, information about how to process small classes of phenomena which could not be covered by general rules. An important goal for developers of NLP systems is, thus, to find the correct balance between what can be processed on general principles and what is idiosyncratic in language, what we can calculate and what we must know liter-

ally, what is compositional and what is conventional. In other words, the decision as to what to put into a set of general rules and what to store in a static knowledge base such as the lexicon becomes a crucial early decision in designing computational-linguistic theories and applications. Thus, the question is: to generalize or not to generalize?

The firmly negative answer (“never generalize”) is not common in NLP applications these days—after all, some generalizations are very easy to make and exceptions to some rules do not faze too many people: morphology rules are a good example.<sup>57</sup> A skeptical position on generalization, i.e., “generalize only when it is beneficial,” is usually taken by developers of large-scale applications, having to deal with deadlines and deliverables. Only rules with respectable-sized scopes are typically worth pursuing according to this position (see Viegas *et al.* 1996b). The “nasty” question here is: are you ready then to substitute “a bag of tricks” for the actual rules of language? Of course, the jury is still out on the issue of whether language can be fully explained or modeled—at least, until we learn what actually goes on in the mind of the native speaker—with anything which is not, at least to some extent, a bag of tricks.

Rules and generalizations can be not only expensive but also in need of corrective work due to overgeneralization; and this has been a legitimate recent concern (see, for instance, Copestake 1995, Briscoe *et al.* 1995). Indeed, a rule for forming the plurals of English nouns, though certainly justified in that its domain (scope) is vast, will produce, if not corrected, forms like *gooses* and *childs*. For this particular rule, providing a “stop list” of (around 200) irregular forms is relatively cheap and therefore acceptable on the grounds of overall economy. The rule for forming mass nouns determining meat (or fur) of an animal from count nouns denoting animals (as in *He doesn't like camel*), discussed in Copestake and Briscoe (1992) as the “grinding” rule, is an altogether different story. The delineation of the domain of the rule is rather difficult (e.g., one has to deal with its applicability to *shrimp* but not to *mussel*; possibly to *ox* but certainly not to *heifer* or *effe*; and, if one generalizes to non-animal food, its applicability to *cabbage* but not *carrot*). Some mechanisms were suggested for dealing with the issue, such as, for instance, the device of ‘blocking’ (see Briscoe *et al.* 1995), which prevents the application of a rule to a noun for which there is already a specific word in the language (e.g., *beef* for *cow*). Blocking can only work, of course, if the general lexicon is sufficiently complete, and even then a special connection between the appropriate senses of *cow* and *beef* must be overtly made, manually.

Other corrective measures may become necessary as well, such as constraints on the rules, counter-rules, etc. They need to be discovered. At a certain point, the specification of the domains of the rules loses its semantic validity, and complaints to this effect have been made within the approach itself (see, for instance, Briscoe and Copestake 1996 about such deficiencies in Pinker 1989 and B. Levin 1993; Pustejovsky 1995: 10 about B. Levin's 1993 classes).

A semantic lexicon that stresses generalization faces, therefore, the problem of having to deal with rules whose scope becomes progressively smaller, that is, the rule becomes applicable to

57. Even in morphology, however, generalization can go overboard. Using a very strict criterion of membership in a declension paradigm, Zaliznyak (1967) demonstrated that Russian has 76 declension paradigms for nouns. Traditional grammars define three. These three, however, cover all but a few hundred Russian nouns. One solution is to write rules for all 76 paradigms. The other is to write rules only for those paradigms with a huge membership and list all the other cases as exceptions.

fewer and fewer lexical units as the fight against overgeneration (including blocking and other means) is gradually won. At some point, it becomes methodologically unwise to continue to formulate rules for creation of just a handful of new senses. It becomes easier to define these senses extensionally, simply by enumerating the domain of the rule and writing the corresponding lexical entries overtly.

Even if it were not the case that the need to treat exceptions reduces the scope of the rules postulated to do that, the overall size of the original scope of a rule, such as the grinding rule (see also Atkins 1991, Briscoe and Copestake 1991, Ostler and Atkins 1992), should cause a considerable amount of apprehension. It is quite possible that the size of its domain is commensurate with the size of the set of nouns denoting animals or plants to which this rule is not applicable. That should raise a methodological question about the utility of this rule. Is it largely used as an example of what is possible or does it really bring about savings in the descriptive effort? Unless one claims and demonstrates the latter, one runs a serious risk of ending up where the early enthusiasts of componential analysis found themselves, after long years of perfecting their tool on the semantic field of terms of kinship (see, for instance, Goodenough 1956, Greenberg 1949, Kroeber 1952, Lounsbury 1956). The scholarship neglected the fact that this semantic field was unique in being an ideal fit for the method ( $n$  binary features describing  $2^n$  meanings). Other semantic fields, however, quickly ran the technique into the ground through the runaway proliferation of semantic features needed to be postulated for covering those fields adequately (see also Section 3.4.3 above). We have found no explicit claims in all the excellent articles on grinding and the blocking of grinding about extensibility of the approach to other rules or rule classes. In other words, the concern for maximum generalization within one narrow class of words is not coupled with a concern for developing a methodology of discovering other lexical rules.

We believe that the postulation and use of any small rule, without an explicit concern for its generalizability and portability, is not only bad methodology but also bad theory because a theory should not be littered with generalizations whose applicability is narrow. The greater the number of rules and the smaller their domains, the less manageable—and elegant—the theory becomes. Even more importantly, the smaller the scope and the size of a semantic class, the less likely it is that a formal syntactic criterion (test) can be found for delineating such a class, and the use of such a criterion for each rule seems to be a requirement in the generative lexicon paradigm. This means that other criteria must be introduced, those not based on surface syntax observations. These criteria are, then, semantic in nature (unless they are observations of frequency of occurrence in corpora). We suspect that if the enterprise of delineating classes of scopes for rules is taken in a consistent manner, the result will be the creation of an ontology. As there are no syntactic reasons for determining these classes, new criteria will have to be derived, specifically, the criteria used to justify ontological decisions in our approach.

This conclusion is further reinforced by the fact that the small classes set up in the battle against overgeneralization are extremely unlikely to be independently justifiable elsewhere within the approach, which goes against the principle of independent justification that has guided linguistic theory since Chomsky (1965), where the still reigning and, we believe, valid paradigm for the introduction of new categories, rules, and notational devices into a theory was introduced. Now, failure to justify a class independently, opens it to the charge of *ad hoc*-ness, which is indefensible within the paradigm. The only imaginable way out lies, again, in an independently motivated

ontology.