

**Automatic Creation of Lexical Knowledge Bases:
New Developments in Computational Lexicology**

Technical Report 97-03

July 1997

**Kenneth C. Litkowski
CL Research
20239 Lea Pond Place
Gaithersburg, Maryland 20879**

DRAFT

**Automatic Creation of Lexical Knowledge Bases:
New Developments in Computational Lexicology**

Abstract

Text processing technologies require increasing amounts of information about words and phrases to cope with the massive amounts of textual material available today. Information retrieval search engines provide greater and greater coverage, but do not provide a capability for identifying the specific content that is sought. Greater reliance is placed on natural language processing (NLP) technologies, which, in turn, are placing an increasing reliance on semantic information in addition to syntactic information about lexical items. The structure and content of lexical entries has been increasing rapidly to meet these needs, but obtaining the necessary information for these lexical knowledge bases (LKBs) is a major problem. Computational lexicology, which began in somewhat halting attempts to extract lexical information from machine-readable dictionaries (MRDs) for use in NLP, is seeing the emergence of new techniques that offer considerable promise for populating and organizing LKBs. Many of these techniques involve computations within the LKBs themselves to create, propagate, and organize the lexical information.

1 Introduction

Computational lexicology began in the late 1960s and 1970s with attempts to extract lexical information from machine-readable dictionaries (MRDs) for use in natural language processing (NLP), primarily in extracting hierarchies of verbs and nouns. During the 1980s, NLP began reaching beyond syntactic information with a greater reliance on semantic information, locating this information within the lexicon. After reaching a conclusion (in the early 1990s) that insufficient information could be obtained about lexical items from MRDs, new techniques have emerged to offer considerable promise for populating and organizing lexical knowledge bases (LKBs). An underlying reason for the realization of these techniques seems to be the increasing capability to deal with the large amount of data that must be digested to deal with the overall content and complexity of semantics.

This discussion begins with the assumptions about large amounts of information in lexical entries and particular computations made with this information in NLP. From this starting point, the paper describes emerging techniques for populating and propagating information to lexical entries derived from existing information with the LKB. The primary motivations for extending lexical entries comes from a need to provide greater internal consistency in the LKB and from an apparently insatiable requirement for greater amounts of information to support demands from very unlikely sources.

The first set of techniques that are described revolve around more detailed analysis of definitions from MRDs, focusing on research from Microsoft, with elaborations in attempts to articulate conceptual clusters. Next, several avenues of research have developed techniques for creating new categories out of existing hierarchies, dynamically cutting across hierarchical links, frequently in response to domain-specific processing of text. The status of lexical rules, which provide characterizations of how new entries and senses are derived from existing entries and senses, has been refined in ways that are closer to the way language uses these rules and that permit the variation in phrase structure. The last section discusses the potential of an overall theory of the lexicon arising from a formalization of semantic networks with the theory of labeled directed graphs.

2 Assumptions about contents of lexical entries

A lexicon begins with a simple listing of word forms, and may be initially extended to include phrasal entries. We would expect a next extension to include information found in an ordinary paper dictionary: inflectional forms, parts of speech, definitions, and perhaps usage information, pronunciation, and etymology. Lexicons used in some form of computerized text processing (such as information retrieval, natural language processing, machine translation, and content analysis) are requiring ever-increasing amounts of structure and content associated with each entry.

Information retrieval lexicons (thesauruses) create links between items, indicating that one entry is broader than, narrower than, or equivalent to another entry. Natural language processing requires syntactic information about each entry, primarily in the specification of subcategorization patterns (that is, syntactic structures likely to appear in the surrounding context). Machine translation makes use of simple correspondences, much like thesauruses, merely equating words in the source language to words in the target language (the transfer model), but this model doesn't always hold true because concepts are expressed differently in different languages, thus requiring more information about the conceptual and structural content of lexical entries (the interlingua model). Content analysis requires lexicons that are broken down into categories, themes, or subject areas.

As a result of developments in the fields noted above, lexical entries today may include categorical information (part of speech), inflectional and perhaps morphologically derived forms, syntactic and semantic features (typically boolean information), information about syntactic structure, semantic information that places the lexical item with a world view (an ontology), and miscellaneous information that characterizes a word's pragmatic usage. (Nirenburg, et al. 1992) provide the most complete range of information in a lexical entry, including category, orthography, phonology, morphology (irregular forms, paradigm, and stem-variants), annotations, applicability (such as field and language), syntactic features (binary values such as *count*, multiple values such as *number*), syntactic structure (subcategorization patterns), semantics (semantic class and lexical mapping from syntactic patterns to role values), lexical

relations, lexical rules, and pragmatics (including stylistic information and analysis triggers to characterize domain and text relations). As described in (Nirenburg, et al. 1995), entries from other systems may be mappable into an ontologically-based lexical entry.

There are four aspects of the Text Meaning Representation and Ontology of Nirenburg's MikroKosmos system where extension of the information may be possible: (1) semantic relations with other entries, perhaps not highlighted as well as in other systems that are overtly characterized as semantic networks, such as the Unified Medical Language System at the National Library of Medicine (this semantic network includes a highly elaborated set of 56 semantic relations, themselves presented in a hierarchy); (2) identification of collocational patterns associated with a lexical entry (such as Mel'čuk's functional specifications); (3) internal structure of the different senses of a lexeme, particularly showing any derivational relationships between senses and allowing for underspecification (that is, supersenses that are ambiguous with respect to particular features present in subsenses); and (4) identification of the logical constraints, preconditions, effects, and decomposition of meaning associated with use of the lexical item.

Based on the foregoing, a general assumption is that all possible information about each lexical item is to be obtained and placed in the lexicon. If there are additional types of information beyond that identified thus far, the assumption is that it will be useful to include such information in the lexicon. Typically, the specific information included in the lexicon is driven by the application and may be optimized in some way to facilitate use within that application. This means that only pertinent information for an application is extracted from the lexical knowledge base. (Of course, many applications may never need to develop all the information that may be associated with a lexical item.)

3 Assumptions about current computations in the lexicon

Historically, information in a lexicon has simply been accessed for subsequent processing in an application area. In the mid 1980s, an observation was made in the development of Generalized and Head-Driven Phrase Structure Grammars (GPSG, HPSG) that the lexicon could be the repository of information that could replace and facilitate many of the control structures used in natural language processing. Since that time, many systems have been developed that have placed increasing reliance on the lexicon. This led to the development of binding and unification techniques that make it possible for information from separate lexical entries to combine with one another. In addition, these techniques made it possible to structure the lexicon into an inheritance hierarchy, so that it is not necessary to put redundant information in every lexical entry. (The precise form of inheritance is an area of considerable research today, with (Davis 1996) providing a semantic hierarchy.)

In a separate vein, a considerable industry had evolved for analyzing machine-readable dictionaries (MRDs). It had been found that ordinary dictionaries contained much information

that could be used in a variety of natural language processing tasks, and so, attempts were made to convert such information into appropriate forms. Along with these attempts, it was found possible to extract hierarchies from these MRDs (although fraught with a major difficulty in identifying the particular sense in which words were used to ensure the validity of the hierarchy).

4 Computations for populating and propagating lexical entries

The development of an LKB is generally considered to be an extremely labor-intensive effort, with each entry hand-crafted. Analysis of MRDs has attempted to automate some of this effort, but it is difficult to see where results from such efforts have actually been used. It seems as if no progress is being made, so that each new report in the literature may provide new observations, but there is little sense of an accumulation of knowledge, of the establishment of an LKB that is amenable to evolution and expansion. Moreover, (Richardson 1997: 132) stated that the import of (Ide & Veronis 1993) and (Yarowsky 1992) was to suggest that "LKBs created from MRDs provide spotty coverage of a language at best." Except for the efforts at Microsoft, it appears that there are presently no major projects aimed at extracting LKB material from MRDs. To some extent, dictionary publishers are making more direct electronic use of their materials, but this work generally is merely an electronic version of the paper dictionaries, with little view of an entirely different structure optimized for text processing applications.

Perhaps these difficulties require a different perspective on the nature of a lexicon. Personal and general (i.e., dictionary) lexicons undergo continuing evolution and extension. This suggests that computational lexicons need to be engineered with this in mind. LKBs are dynamic entities that will undergo almost continual revision. An LKB is an entity that sits apart from any use we make of it, and while it is sitting there off-line, and should be undergoing a continual process of expansion and reorganization. At any time, subsets of the information from the LKB are extracted for use in a particular application.

This process of expansion and reorganization can be very dynamic; lexicon update should be able to occur within the span of analysis of a single document. A single document can contain its own sublanguage, and may introduce new ontological facts and relations and may use existing lexical items in novel ways that are not present in the current LKB. There are reasonably well-known lexical processes by which these new ontological and sense data are added manually. We may now be at a sufficient state of progress that these processes can be automated to provide the kind of dynamic LKB that we need.

5 Motivations

The greatest problem of computational linguistics seems to be the acquisition bottleneck, specifically the acquisition of new lexical items (mostly new senses of existing words, that is, uses of existing words in ways that are only slightly different from what may be present in the LKB) and new pieces of knowledge unknown to our knowledge base. (These are items added to

semantic memory and to episodic memory.) To deal with this problem, it seems necessary to design bootstrapping techniques into the knowledge bases. These bootstrapping techniques require an almost continual re-evaluation of the data in our lexicons and knowledge bases, to make new computations on this data in order to reassess and reconsider each component part.

5.1 Greater amount of information available

Developments in NLP have required increasing amounts of information in the lexicon. In addition, there is an increasing requirement that this information be amenable to dynamic processing. Research with LKBs that have a static structure and content, such as WordNet, increasing move toward expansion of information and cross-cutting use of the existing structure and organization. Different types of applications make use of this information in unanticipated ways. Data dictionaries for database applications, articulation of primitives for such things as the Knowledge Query Manipulation Language and Knowledge Interchange Format, and terminological databases may each require a different cut on an LKB.

The development of an LKB should be able to encompass all of the applications that may eventually rely on it. A particular application would be able to extract only the necessary information and may take advantage of particular storage, representation, and access mechanisms for efficiency optimization. Every opportunity for processing text can be considered as an opportunity for expanding the LKB. Every opportunity should be taken to increase the amounts and types of information included in the LKB.

5.2 Consistency of lexicon

Guidelines are generally prepared for developing lexicons and LKBs. As much as possible, these guidelines should be automated. More specifically, an LKB should exhibit a considerable amount of internal consistency. At least three types of consistency can be envisioned: (1) circularity should be rooted out; (2) consistency and correctness of inheritance should be tested; and (3) compositional characteristics of lexical items should be checked. Compositional characteristics can be further checked externally by examination of actual data.

6 Definition analysis (forward and backward)

(Amsler 1980) provided the first rigorous attempt to analyze dictionary definitions, building a taxonomic hierarchy based on the genus words of a definition. This work was continued at IBM in the early 1980s, described in (Chodorow, et al. 1985), further attempting automatic extraction of these taxonomies. This was done through heuristic pattern-matching techniques to identify the genus terms in definitions and then to structure them in a hierarchy.

Several other research efforts during the later 1980s continued analysis of dictionary definitions to extract information. (Markowitz, et al. 1986) investigated "semantically"

significant patterns based on parsing definitions (with the linguistic string parser); these included taxonomy-inducing patterns, member-set relations, generic agents (in noun definitions), suffix definitions, identifying action verbs from noun definitions ("the act of Ving"), selectional information for verb definitions, and recognizing action vs. stative adjectives. Other work focused on extracting taxonomies (Klavans, et al. 1990; Copestake 1990; Vossen 1991; Bruce & Guthrie 1992).

(Richardson 1997) says that this work overlooks "the true tangled, circular nature of the taxonomies actually defined by many of the dictionary genus terms." Further, he cites (Ide & Veronis 1993) as observing that "attempts to create formal taxonomies automatically from MRDs had failed to some extent," citing "problems with circularity and inconsistency ... in the resulting hierarchies."

6.1 Microsoft techniques

(Richardson 1997) extracts and creates 16 bi-directional relations for its LKB (called MindNet). Microsoft has analyzed 147,000 definitions and example sentences from (Longman Dictionary of Contemporary English 1978) (LDOCE) and the (The American Heritage Dictionary of the English Language 1992) to create 1.4 million semantic links between lexical entries. The basis for the specific links is the use of structural patterns rather than just string matching, as performed in earlier work (Montemagni & Vanderwende 1993). Table 1 shows the relations automatically created by parsing in creating Microsoft's MindNet. There are two key steps in what Microsoft has done: (1) parsed the definitions and example sentences with a broad-coverage parser and (2) included, in characterizing a word's meaning, all instances in which that word has been used in defining other words, not only where that word is the genus term. An example of the significance of the latter is for creating meronymic ("part-of") links between entries. As (Richardson 1997) indicates, the parts of an object (say "car") are seldom described in the dictionary entry for that object. However, other entries (for example, "fender") make use of the object in their definitions (a *fender* is a "guard over the wheel of a **car**"). Richardson distinguishes between semantic relations derived by analyzing a word's definitions (forward-linking) and those derived from definitions of other words (backward-linking). Backward-linking relations are known as "inverted semantic relation structures" and are stored with a main entry; they are used for disambiguation in parsing and measurement of similarity. (When a definition is parsed, the relations structure is stored at that entry. An "inverted" structure is stored at all other words identified as related.)

Table 1: Relations Automatically Created in Microsoft Analysis

Cause	Domain	Hypernym	Location
Manner	Material	Means	Part
Possessor	Purpose	Quasi-Hypernym	Synonym
Time	Typical Object	Typical Subject	User

(Richardson 1997) notes that much of the work attempting to create networks from dictionary definitions in building LKBs has focused on quantitative information (that is, measuring distance between nodes in the network or measuring semantic relatedness based on co-occurrence statistics). Instead, he focuses on labeling semantic relations over simple co-occurrence relations and distinguishing between paradigmatic relatedness (substitutional similarity) and syntagmatic relatedness (occurring in similar contexts).

This important component of the Microsoft use of MindNet is a procedure for determining similarity between words based on semantic relations between them. A semantic relation path between word1 and word2 exists when word2 appears in word1's forward-linked structure or in any of word1's inverted relation structures. Richardson distinguishes between paradigmatic similarity (*magazine* may be substituted for *book* in many contexts) and syntagmatic similarity (*walk* and *park* frequently occur in the same context, e.g., "a walk in the park," but cannot be substituted for one another). Richardson builds similarity measures after studying the predominant semantic relation paths between entries (that is, path patterns).

6.2 Conceptual clusters

(Schank & Abelson 1977) describes an elaborate structure of scripts (e.g., a scenario of eating in a restaurant), intended to capture events made up of more than one element and identifying objects that play roles in the events. (McRoy 1992) says that a text will generally exhibit lexical cohesion and describes conceptual clusters, defined as "a set of senses associated with some central concept." She distinguishes three types of clusters: *categorical* (senses sharing a conceptual parent), *functional* (senses sharing a specified functional relationship such as part-whole), and *situational* (encoding "general relationships among senses on the basis of their being associated with a common setting, event, or purpose"). Thus, the situational cluster for *courtroom* includes senses for words such as *prison*, *crime*, *defendant*, *testify*, *perjure*, *testimony*, and *defend*. (Carlson & Nirenburg 1990), in describing lexical entries that can be used in world modeling, envision most of the components associated with scripts and conceptual clusters, particularly identifying semantic roles (with selectional restrictions) and decomposition of event verbs. (Richardson 1997) describes the process by which conceptual clusters can be identified from MindNet based on identifying the top 20 paths between query words. He notes that such

clusters are useful not only in word sense disambiguation but also in the expansion of queries in information retrieval. The specificity of the relations is an addition to previous work.

6.3 Fillmore's Frames

(Lowe, et al. 1997) outline the conceptual underpinnings of an effort to create a database called FrameNet. Their primary purpose is to produce frame-semantic descriptions of lexical items. They note the lack of agreement on semantic (case) roles and observe each field seems to bring a new set of more specific roles. They suggest that many lexical items evoke generic events with more specific characterizations of the roles and that they instantiate particular elements of the frames. They state that "any description of word meanings must begin by identifying underlying conceptual structures" which can be encoded in frames characterizing stereotyped scenarios. They recognize the importance of inheritance in encoding lexical items in this way.

They note that a frame (for generic medical events, for example) might involve detailed frame elements for *healer*, *patient*, *disease*, *wound*, *bodypart*, *symptom*, *treatment*, and *medicine*. A key new element is the examination from corpus analysis of the frame elements from a given frame that occur in a phrase or sentence headed by a given word (calling these *set frame element groups*). They would identify which elements of a frame element group are optional or implied but unmentioned. They would recognize that some lexical items may encode multiple frame elements (for example, *diabetic* identifies both the disorder and the patient). In summary, they envision that lexical entries will include full semantic/syntactic valence descriptions, with frame elements (for at least verbs) linked to a specification of sortal features, indicating the selectional and syntactic properties of the constituents that can instantiate them.

(UMLS knowledge sources 1996), with its elaborate semantic network and semantic relation hierarchy, identifies semantic types linked by the various relations, and thus would clearly satisfy some of the requirements for identifying frame elements in the medical field.

6.4 Barriere techniques

Richardson (personal communication) has stated that Microsoft's MindNet, with its forward-linked and backward-linked relational structures, essentially identifies conceptual clusters associated with lexical items. Indeed, viewing a graphical representation of some elements of MindNet, with lexical entries as nodes and the various relations as labels on directed arcs between nodes, it is clear that the concepts clustered about a lexical item capture the ways in which that lexical item may be used in ordinary text.

(Barrière & Popowich 1996b) have also extracted semantic structures from dictionary definitions, with the specific objective of identifying conceptual clusters. They note that much earlier work with MRDs has a localist orientation, with primary concern on providing

information for the main entries, without concern for the relations between entries. They provide a bootstrapping technique to create Concept Clustering Knowledge Graphs, based on using the conceptual graphs of (Sowa 1984). They start with a trigger word and expand a forward search through its definitions and example sentences to incorporate related words. They note that the clusters formed through this process are similar to the (Schank & Abelson 1977) scripts; however, they make no assumptions about primitives.

They start by forming a temporary graph using information from closed class words (*with* is subsumed by *instrument* in a relation hierarchy), relations extracted using defining formulas, and relations extracted from the syntactic analysis of the definition or sample sentence. They make use of a concept hierarchy and rules that provide predictable meaning shifts (from lexical implication rules). The key step in their procedure for combining temporary graphs is a maximal join operation formed around the maximal common subgraph using the most specific concepts of each graph. After forming a graph from analysis of a word's definition, they search the dictionary for places where that word is used in defining other words; this information is combined with the graphs already formed. While these clusters are similar to those developed by Microsoft, they are based on more rigorous criteria in requiring subsumption relationships between the temporary graphs and involve use of only semantically significant words. This information is useful in analyzing the entire network of definitions in a dictionary, as described below in the section on digraphs.

7 Higher order category formation

(Nida 1975) indicates that a semantic domain may be defined based on any semantic features associated with lexical items. He used this observation to assert that any attempt to identify a single hierarchy or ontology was somewhat arbitrary and dependent on a user's need. Problems with direct use of WordNet synsets in information retrieval (q.v. (Voorhees 1994)) may reflect the difficulty in using a single hierarchy.

(Nida 1975: 174) characterized a semantic domain as consisting of words sharing semantic components. (Litkowski 1997) suggests that dynamic category systems reflecting more of the underlying features and semantic components of lexical entries may be more useful in many NLP applications, thus providing importance to the addition of this information wherever possible. Several techniques have been developed in the past few years to create categorization schemes that cut across the static WordNet synsets.

7.1 Supercategories of Hearst

(Hearst & Schütze 1996) provide the starting point for creating new categories out of WordNet synsets. They recognized that a given lexicon may not suit the requirements of a given NLP task and investigated ways of customizing WordNet based on the texts at hand. They adjusted WordNet in two ways: (1) collapsing the fine-grained structure into a coarser structure,

but keeping semantically-related categories together and letting the text define the new structure and (2) combining categories from distant parts of the hierarchy.

To collapse the hierarchy, they use a size cutoff. They formed a new category if a synset had a number of children (hyponyms) between a lower and upper bound (25 and 60 were used). They formed a new category from a synset if it had a number of hyponyms greater than the lower bound, bundling together the synset and its descendants. They identified 726 categories and used these as the basis for assigning topic labels to texts, following (Yarowsky 1992) (collecting representative contexts, identifying salient words in the contexts and determining a weight for each word, and predicting the appropriate category for a word appearing in a novel context). To extend their category system, they computed the closeness of two categories based on co-occurrence statistics for the words in the category (using large corpora). They then used the mutual ranking between categories (both categories had to be highly ranked as being close to the other). As a result, they combined the original 726 categories into 106 new supercategories. (Names for the new supercategories were chosen by the authors.) The results in characterizing texts was observably better. They also noted that their approach could be used at a narrower level in order to achieve greater specificity.

7.2 Basili supercategories

(Basili, et al. 1997) describe a method for tuning an existing word hierarchy (in their case, WordNet) to an application domain. The technique creates new categories as a merging of WordNet synsets in such a way as to facilitate elimination of particular WordNet senses, thus reducing ambiguity.

They make several observations about the nature of domain-specific vocabularies. They note that a number of lexical acquisition techniques become more viable when corpora have a domain-specific semantic bias, particularly allowing the identification of domain-specific semantic classes. They suggest that modeling semantic information is very corpus and domain dependent, and general-purpose sources (MRDs and static LKBs, including WordNet) may be too generic.

A domain-specific approach can take advantage of several findings: (1) ambiguity is reduced in a specific domain, (2) some words act as sense primers for others, and (3) raw contexts of words can guide disambiguation. They use a classifier that tunes WordNet to a given domain, with the resulting classification more specific to the sublanguage and then able to be used more appropriately to guide the disambiguation task. There are four components to this process: (1) tuning the hierarchy rather than attempting to select the best category for a word; (2) using local context to reduce spurious contexts and improve reliability; (3) not making any initial hypothesis on the subset of consistent categories of a word; and (4) considering globally all contexts to compute a domain-specific probability distribution.

To develop the classifier, they make use of WordNet tops (unique beginners) as classes. They first compute the *typicality* of a word (to which class does most of a word's synsets belong), the *synonymy* of a word in a class (the number of words in the corpus appearing in at least one of the synsets of the word that belong to the class divided by the number of words in the corpus that appear in at least one of the synsets of the word), and the *saliency* of a word in a class (the product of the absolute occurrences of the word in the corpus, the typicality, and the synonymy). A *kernel* is formed for a class by selecting words with a high saliency. This kernel appears to be clearly distinctive for the domain (shown in the example).

In the next step, the kernel words are used to build a probabilistic model of a class, that is, distributions of class relevance of the surrounding terms in typical contexts for each class are built. Then, a word is assigned a class according to the contexts in which it appears in order to develop a *domain sense*. These steps reduce the WordNet ambiguity (from 3.5 to 2.2 in the material presented). Finally, each word is assigned a class based on maximizing a normalized score of the domain senses over the set of kernel words.

The system described above has been used as the basis for inductively acquiring syntactic argument structure, selectional restrictions on the arguments, and thematic assignments. This information allows further clustering of the senses, which would enable further refinement of a category system like WordNet, that is, as information is added to WordNet entries, all the steps above could be performed more effectively.

7.3 Buitelaar's techniques

(Buitelaar 1997) argues that a lexical item should be assigned a representation of all its systematically related senses, from which further semantic processing can derive discourse dependent interpretations. This type of representation is known as underspecification. In this case, it is based on the development of systematic polysemous classes with a class-based acquisition of lexical knowledge for specific domains. The general approach for identifying the classes stems from the Generative Lexicon theory of (Pustejovsky 1995), with qualia roles enabling type coercion for semantic interpretation.

An important basis for this approach is disambiguation between senses is not always possible (the problem of *multiple reference*) and may in fact not be appropriate, since an utterance may need to convey only part of the meaning of a word, without requiring specification down to a final nuance (the *sense enumeration* problem). One may think of representing the different senses of a word in its own hierarchy, with leaves corresponding to fully-distinguished senses and with internal nodes corresponding to decision points on particular semantic features. The meaning at these internal nodes is thus underspecified for the semantic features at the leaves.

Buitelaar suggests that much polysemy is systematic and uses WordNet classes to identify the systematicity. For an individual word with multiple WordNet senses, he notes that the senses

may group together on the basis of the WordNet tops or unique beginners and that even within the groups the senses may be related as instantiations of particular qualia (*formal*, *constitutive*, *telic*, and *agentive*) of an overarching sense.

Buitelaar reduces all of WordNet's sense assignments to a set of 32 basic senses (corresponding to, but not exactly identical to, WordNet's 26 tops). He identifies 442 polysemous classes in WordNet, each of which is induced by words having more than one top. Some of these do not correspond to systematic polysemy, but are rather derived from homonyms that are ambiguous in similar ways and that hence are eliminated from further study.

Qualia roles are typed to a specific class of lexical items. The types are simple (**human**, **artifact**) or complex (**information•physical**), also called "dotted types." There are two complex types: (1) systematically related (where an utterance simultaneously and necessarily incorporates both of the simple types of which it is composed, e.g., *book*, *journal*, *scoreboard* are **information** and **physical** at the same time, a "closed dot") and (2) related but not simultaneously (only one aspect is (usually) true in an utterance, e.g., *fish* is **animal•food**, but is only one of these in a given utterance, an "open dot"). Open-dot types generally seem to correspond to systematic polysemy, such as induced by the *animal-grinding* lexical relation. Identification of such lexical relations is still an open area of research.

The underspecified types enumerated above can be adapted to domain-specific corpora. The underspecified type is a basic lexical-semantic structure into which specific information for each lexical item can be put, that is, provides variables which can be instantiated. Buitelaar suggests that the manner of instantiation is domain- and corpus-specific. He first tags each word in a corpus with the underspecified type. The next step involves pattern-matching on general syntactic structures, along with heuristics to determine whether a specific type is appropriate for the application of the pattern. For example, the pattern "NP Prep NP", where Prep = "of", indicates a "part-whole" relation if the head noun of the first NP has a type either the same as that of the second NP or is one of the composing types of the second NP. Thus, "the second paragraph of a journal," with "paragraph" of type **information** and "journal" of type **information•physical**, allows the inference that the "paragraph" is a part of the "journal."

The information gathered in the second step is used to classify unknown words. Results of the classifier seem to relate to the homogeneity of the corpus. Finally, the underspecified lexicon is adapted to a specific domain by using the observed patterns and translating them into semantic ones and generating a semantic lexicon representing that information. Particular patterns are viewed as identifying hypernyms (the formal quale), meronyms (the constitutive quale), and predicate-argument structure (the telic and agentive qualia).

7.4 Intersective sets

(Palmer, et al. 1997) are concerned with lexical acquisition and have described an implementation of lexical organization that may have increased potential for adaptable lexical processing. They explicitly represent a lexical hierarchy that captures fine-grained classes of lexical items, as well as their associations with other classes that share similar semantic and syntactic features. This approach is being applied to the Lexicalized Tree Adjoining Grammar. They hypothesize that syntactic frames can be used to extend verb meanings and thus acquire new senses for lexical items.

(Levin 1993) verb classes are based on regularities in diathesis alternations, as specified by several pairs of syntactic frames. There is an underlying hypothesis that these classes correspond to some underlying semantic components, which are discussed in general terms but not yet made explicit. For an unknown verb in a text, being able to recognize its syntactic pattern provides a reasonable prediction of its verb class, thus providing a first attempt to characterize its semantic features. This may sometimes enable a sense extension for an existing verb.

Palmer, et al. have examined Levin's verbs in conjunction with the WordNet synsets. In particular, they observed that many verbs fall into multiple Levin classes. They augmented Levin classes with so-called *intersective* classes, grouping existing classes that share at least three members, with the hypothesis that such an overlap might correspond to a systematic relationship. The intersective class names consist of the Levin class numbers from which they were formed. (Since Levin includes only 4,000 verbs, with 20,000 identified in a large dictionary, each set may conceivably be extended, allowing reapplication of this technique. The analysis could also be extended to overlaps containing only two members.) Palmer, et al. identified 129 intersective classes; they then reclassified the verbs, removing them from the Levin classes if they occurred in an intersective class. This reduced the "ambiguity" of the verbs (that is, the number of classes to which a verb belongs). Moreover, the resulting intersective classes had face validity, seeming to correspond to intuitively apparent idiosyncratic ambiguities.

As mentioned above, the Levin classes, even though capturing common syntactic patterning, are thought to correspond to semantic differences. So, the intersective classes were examined in conjunction with WordNet synsets. Although the analysis was performed mostly by hand and with intuitive judgments, the comparison apparently is made by identifying WordNet synsets that have hyponyms in the intersective class and the two classes from which it was formed. Thus, with the intersective class "cut/split," it was possible to identify WordNet distinctions of synsets "cut into, incise" and "cut, separate with an instrument" (and coincidentally, indicating that the first of these synsets is a hyponym of the second).

Palmer, et al. indicate that they are building frames to represent the meanings of their lexical entries, capturing syntactic and semantic distinctions. By examining the relationships of these entries with the information obtained from the intersective class analysis and the WordNet

synsets, they can more easily identify the specific syntactic and semantic distinctions (that is, disambiguate one class with another and vice versa). Moreover, it then becomes easier to arrange the lexical items into an inheritance hierarchy where specific syntactic and semantic components are expressed as templates.

Based on the inheritance hierarchy, they can then measure the proximity of classes in the lattice in terms of the degree of overlap between each class's defining features. Conversely, but not mentioned by the authors, it seems possible to go the other way. If lexical entries have a bundle of syntactic and semantic features, they can be examined for common components to identify templates (e.g., containing a field for number with a set of possible values).

7.5 Abstraction

Abstraction is the process of identifying these underlying features and relaxing and removing the subsidiary features to create a more general characterization of a set of words or a text. (Litkowski & Harris 1997; Litkowski 1997) describe principles and procedures for category development, particularly noting the similarity to (Hearst & Schütze 1996) in providing supercategories. A general theme in these principles and procedures was the importance of characterizing lexical entries in terms of their syntactic and semantic features. Another theme was that existing categorizations, such as WordNet, should not be viewed as static entities. This stems not from the fact that one may quibble with WordNet entries and hierarchies, but rather from the hypothesis that characterization of a categorization scheme or a text may cut across WordNet synsets because the characterization involves highlighting of different underlying syntactic, semantic, or other lexical features.

(Litkowski & Harris 1997) particularly dealt with category development for textual material, that is, characterizing the discourse structure of a text. There, a discourse analysis was performed generally following Allen's algorithm for managing the attentional stack in discourse structure analysis (Allen 1995: 526-9), with an extension to incorporate lexical cohesion principles (Halliday & Hasan 1976). The algorithms involved identifying discourse segments, discourse entities, local discourse contexts (for anaphora resolution), and eventualities. The result was a set of discourse segments related to one another (with many identified as subsidiary), discourse entities and eventualities, and various role and ontological relations between these entities. The concepts and relations (including the discourse relations) were essentially present in and licensed by the lexicon, and then instantiated by the given text to carve out a subnetwork of the lexicon. The definition of this subnetwork was then constructed by identifying the highest nodes in the ISA backbone and the additional relations that operate on the backbone, along with selectional restrictions that are used.

Characterizing this subnetwork was a matter of identifying the topmost ISA nodes (and perhaps more importantly, identifying descendants that to be excluded). Naming this subnetwork is based on the set of topmost nodes, any relations (semantic roles or other semantic relations),

and selectional restrictions. This process of characterizing a subnetwork is quite similar to the development of supercategories in . Thus, to at least that extent, this process may be viewed as leading to identification of the topic of a text. (It is assumed that the network nodes are organized in the same way as WordNet synsets, that is, several lemmas expressing the same concept. This would constitute a thematic characterization of a text. The exclusion of descendants would perhaps increase precision in information retrieval, a significant problem with search engines that allow thesaural substitutions or expand queries based on themes.)

8 Extension of lexical entries

An important characteristic of a lexicon is that the entries and senses are frequently systematically related to one another. Many lexical entries are derived from existing ones. Lexical rules can cover a variety of situations: derivational morphological processes, change of syntactic class (conversion), argument structure of the derived predicate, affixation, and metonymic sense extensions. Thus, lexical rules should "express sense extension processes, and indeed derivational ones, as fully productive processes which apply to finely specified subsets of the lexicon, defined in terms of both syntactic and semantic properties expressed in the type system underlying the organization of the lexicon" (Copestake & Briscoe 1991). The most basic of these derivational relations is the one in which inflected forms are generated. These are generally quite simple, and include the formation of plural forms of nouns, the formation of tensed (past, past participle, gerund) forms of verbs, and the formation of comparative and superlative forms of adjectives.

Derivational relations may form verbs from nouns, nouns from verbs, adjectives from nouns and verbs, nouns from adjectives, and adverbs from adjectives. Many of these relations have morphological implications, with the addition of prefixes and suffixes to base forms. These relations generally operate at the level of the lexical entries.

In a lexicon where entries are broken down into distinct senses, the senses may be systematically related to one another without any morphological consequences. The *animal-grinding* lexical relation mentioned above is such an example.

The status of lexical relations is currently undergoing substantial refinement (see (Helmreich & Farwell 1996) for example). Several useful developments have recently occurred that have implications for the content of lexical entries themselves.

8.1 Instantiation of lexical rules

(Flickinger 1987) first introduced the notion that lexical rules were important parts of a hierarchical lexicon. (Copestake & Briscoe 1991) describe types of noun phrase interpretations that may involve metonymy: individual-denoting NPs, event-denoting NPs (subdivided into those with telic roles and those with agentive roles, based on an underspecified predicate),

animal-denoting interpretation vs. food-denoting one, count nouns transformed into mass senses denoting a substance derived from the object. Perhaps as important as describing these processes, Copestake and Briscoe also were able to express these lexical rules as lexical entries themselves (in a typed feature structure). (These might be called "pseudoentries" to distinguish them from words and phrases that would be used in texts.)

The essence of the representation is that a lexical rule consists of two features (denominated <0> and <1>), where the first feature (<0>) has a value (which is itself a complex feature structure) that specifies the typed feature structures to be matched and the second feature (<1>) has a value that specifies the typed feature structure in the derived entry or sense (where, for example, a new value for an "orthography" feature would create a new entry in the lexicon).

This representational formalism could be used to extend a lexicon. One could take an existing lexicon and start a process to generate new entries and senses for each lexical rule. This process would simply iterate through a list of rules, find any entries and senses to which the <0> feature applies, and create new entries and senses based on the <1> feature of the lexical rule. Conversely, in a recognition system, for any unknown word or use of an existing word, one could create a tentative entry or sense (postulating various syntactic and semantic features), search the lexical rules to determine if any of them has a <1> feature matching the postulated entry or sense, and then determine if the corresponding <0> feature matches an existing entry or sense (thus validating the characterization of the unknown word or sense).

8.2 Probabilistic Finite State Machines in Lexical Entries

(Briscoe & Copestake 1996) recognize various efficiency issues that have arisen in connection with systems that rely heavily on lexical rules. They note the development of techniques for (1) 'on-demand' evaluation of lexical rules at parse time, (2) the storage of finite state machines in lexical entries to identify possible "follow relations" (an ordering of lexical rules that can apply to a lexical entry), and (3) an extension of entries with information common to all their derived variants. Notwithstanding, they state that "neither the interpretation of lexical rules as fully generative or as purely abbreviatory is adequate linguistically or as the basis for LKBs."

To deal with this problem, they create a notion of probabilistic lexical rules to correspond with language users' assessments of the degree of acceptability of a derived form. They introduce probabilities in both the lexical entries and the lexical rules. For the lexical entries, they assume a finite state machine that can represent the possible application of lexical rules, which are intended to encompass all entry and sense derivations from a base form. This is the conditional probability of a lexical entry of the given sense given the word form (the frequency of the derived form, e.g., a particular subcategorization pattern, divided by the frequency of the word form). Some states will have no associated probability if they are not attested. There is, of course, the

difficulty of acquiring reliable estimates, and they note the desirability of using smoothing techniques for rare words.

For unattested derived lexical entries, the relative productivity of the lexical rule can be used. To compute this, they identify all the forms to which the rule can apply and then determine how often it is used. (For example, they would determine how often the lexical rule transforming *vehicle* into *go using vehicle*, Levin's class 51.4.1, occurs. They would then determine from a noun hierarchy all nouns that identify vehicles)

8.3 Phrase variation

Idioms and phrases (multi-word terms) constitute a significant problem in lexicon development. This is an area in which many developments are emerging. There is a spectrum of non-random cooccurrences in language, loosely called collocations, that may be said to range from syntactic patterns to specific word combinations that must appear exactly in sequence and whose meaning is not composed from the meanings of its constituent words. At this latter end of the spectrum, the word combinations achieve the status of constituting a distinct lexical entry. The dividing line between what constitutes a lexical entry is not clearly drawn. The issue of how to recognize the word combinations is also not yet firmly established.

(Mel'čuk & Zholkovsky 1988) describe many functional relations that may give rise to collocations. (Smadja & McKeown 1990) categorized collocations as open compounds, predicative relations, and idiomatic expressions. (Smadja & McKeown 1991) describe procedures for lexical acquisition of multi-word terms and their variations. Generally, these procedures have been useful for proper nouns, particularly organizations and company names. Some recent developments suggest that a broadened view of the lexicon, its structure, and the contents of its entries may be useful in the further characterization of multi-word terms.

(Burstein, et al. 1996; Burstein, et al. 1997) developed domain-specific concept grammars which correspond to the inverse of the variant extension technique described for lexical rules. These grammars were used to classify 15- to 20-word phrases and essays (answers to test items) for use in an automatic scoring program. Automatic scoring must be able to recognize paraphrased information across essay responses and to identify similar words in consistent syntactic patterns, as suggested by (Montemagni & Vanderwende 1993).

They built a concept lexicon identifying words thought to convey the same concept (using only the relevant vocabulary in a set of training responses). They parsed the answers (using the Microsoft parser), and substituted superordinate concepts from the lexicon for words in the parse tree. They then extract the phrasal nodes containing these concepts. In the final stage, phrasal and clausal constituents are relaxed into a generalized representation (XP, rather than NP, VP, or AP). Their concept grammars for classifying answers were then formed on the basis of the

generalized representation. In part, these concept grammars are licensed by the fact that many concepts are realized in several parts of speech.

(Jacquemin, et al. 1997) describe a system for automatic production of index terms to achieve greater coverage of multi-word terms by incorporating derivational morphology and transformational rules with their lexicon. This is a domain independent system for automatic term recognition from unrestricted text. The system starts with a list of controlled terms, automatically adds morphological variants, and considers syntactic ways linguistic concepts are expressed.

They identify three major types of linguistic variation: (1) syntactic (the content words are found in a variant syntactic structure, e.g., *technique for performing volumetric measurements* is a variant of *measurement technique*); (2) morpho-syntactic (the content words or derivational variants are found in a different syntactic structure, e.g., *electrophoresed on a neutral polyacrylamide gel* is a variant of *gel electrophoresis*); and (3) semantic (synonyms are found in the variant, e.g., *kidney function* is a variant of *renal function*). The morphological analysis is more elaborate than simple stemming. First, inflectional morphology is performed to get the different analyses of word forms. Next, a part of speech tagger is applied to the text to perform morphosyntactic disambiguation of words. Finally, derivational morphology is applied (over)generate morphological variants. This overgeneration is not a problem because the term expansion process and collocational filtering will avoid incorrect links.

The next phase deals with transformation-based term expansion. Transformations are inferred from the corpus based on linguistic variations (distinct from morphological variants). Two general types of variation are identified: (1) variations based on syntactic structure: (a) coordination (*chemical and physical properties* is a variation of *chemical properties*), (b) substitution/modification (*primary cell cultures* is a variation of *cell cultures*), (c) compounding/decompounding (*management of the water* is a variation of *water management*) and (2) variations according to the type of morphological variation: (a) noun-noun variations, (b) noun-verb variations (*initiate buds* is a variation of *bud initiation*), and (c) noun-adjective variations (*ionic exchange* is a variation of *ion exchange*). A grammar (a set of metarules) was devised to serve as the basis for filtering, using only regular expressions to identify permissible transformations.

8.4 Underspecified forms

The reverse of lexical extension through lexical rules leads to the notion of underspecified forms. As mentioned earlier, (Buitelaar 1997) suggested a notion of underspecification in the identification of categories. (Sanfilippo 1995) presented an approach to lexical ambiguity where sense extension regularities are represented by underspecifying meanings through lexical polymorphism. He particularly cited verb alternations (Levin 1993)

and qualia structures (Pustejovsky 1995) and suggested, since there is no control on the application of lexical rules, the use of underspecified forms.

Sanfilippo proposed to represent ambiguities arising from multiple subcategorizations using "polymorphic" subcategorization lexical entries with a typed-feature-structure formalization. An entry is created to represent all possible subcategorizations and then syntactic contextual information is used during language processing to identify (or ground) the underspecified form (binding particular variables). This was done by generating a list of resolving clauses (in Prolog) which identify how the terminal types are inferred from specific contextual information. Moreover, he noted that the resolving clauses could themselves be positioned within a thematic type hierarchy so that it would be unnecessary for this information to be specified within each lexical entry, allowing it to be inherited. Considerable research is presently under way to extend the notion of underspecification.

9 Digraph theory techniques

(Litkowski 1975; Litkowski 1976; Litkowski 1978; Litkowski 1980) studied the semantic structure of paper dictionaries as labeled directed graphs (digraphs) in an overall effort to identify semantic primitives. In these studies, the starting point was to view nodes in the digraphs as entries (and later as concepts) and arcs as definitional relations between entries (initially the simple relation "is used to define" and later as the various types of semantic relations). Digraph theory allows predictions about the semantic structure. In particular, it asserts that every digraph has a point basis (that is, primitives) from which every point in the digraph may be reached. It provides a rationale for moving toward those primitives (the development of "reduction rules" that allow the elimination of words and senses as non-primitive). It makes a prediction that primitive concepts are concepts that can be verbalized and lexicalized in several ways. (These predictions were well served in the development of WordNet, where unique beginners were identified as consisting of several words and phrases, that is, the synsets. Whether analysis of dictionary definitions in an unabridged would yield the same set is an open question.)

(Richardson 1997) commented on the "problems with circularity and inconsistency ... in the resulting hierarchies" noted in earlier studies (Amsler 1980; Chodorow, et al. 1985; Ide & Veronis 1993). He states that the massive network built at Microsoft invalidates this criticism. However, he did not examine this network to determine if it contained any circularities or inconsistencies. (Litkowski 1978) and (Barrière & Popowich 1996a) discussed this problem, with the latter noting that, for a well-constructed children's dictionary, with a relatively small number of definitions, the "taxonomy is a forest with multiple trees, each of which having at its root a group of words defined through a loop" containing a group of synonyms. The results from the study of digraphs, along with the techniques of Barrière, suggest that Microsoft's MindNet can be subjected to further analysis to organize the sets of structures.

The digraph techniques further substantiate the notion of lexical underspecification. When the definition of a node is expanded from representing an entry to representing the concepts in the senses, several observations immediately come into play. The first is that the senses themselves should be organized into their own hierarchy. The second is that nodes in the sense hierarchy frequently correspond to the common factors of the subsenses.

10 Conclusions

Population and propagation of information throughout an LKB is a valuable enterprise. It is intellectually stimulating in its own right, providing many insights into the ways in which humans structure concepts and knowledge. More importantly, the use of the techniques described provides mechanisms for filling out information that can be used in many applications. The techniques suggest that the more information contained in the LKB, the greater the number of applications that might make use of the information in novel ways. The techniques themselves may be useful in these applications. Many of the techniques involve bootstrapping operations, so that the evolution of the LKB and its use can begin small and grow incrementally. Finally, these techniques and information can be used in developing lexical acquisition procedures to obtain external information. Together, the internal lexicon computations and their application to external methods may contribute greatly to solving the bottleneck problem.

References

- Allen, J. (1995). *Natural language understanding* (2nd). Redwood City, CA: The Benjamin/Cummings Publishing Company, Inc.
- The American Heritage Dictionary of the English Language* (A. Soukhanov, Ed.) (3rd). (1992). Boston, MA: Houghton Mifflin Company.
- Amsler, R. A. (1980). The structure of the Merriam-Webster pocket dictionary [Diss], Austin: University of Texas.
- Barrière, C., & Popowich, F. (1996a). Building a noun taxonomy from a children's dictionary. EURALEX96. Göteborg, Sweden.
- Barrière, C., & Popowich, F. (1996b). Concept clustering and knowledge integration from a children's dictionary. COLING96.
- Basili, R., Rocca, M. D., & Pazienza, M. T. (1997). Towards a bootstrapping framework for corpus semantic tagging. 4th Meeting of the ACL Special Interest Group on the Lexicon. Washington, DC: Association for Computational Linguistics.
- Briscoe, T., & Copestake, A. (1996). Controlling the application of lexical rules. In E. Viegas & M. Palmer (Eds.), *Breadth and Depth of Semantic Lexicons*. Workshop Sponsored by the Special Interest Group on the Lexicon. Santa Cruz, CA: Association for Computational Linguistics.
- Bruce, R., & Guthrie, L. (1992). Genus disambiguation: A study of weighted preference. COLING92.

- Buitelaar, P. (1997). A lexicon for underspecified semantic tagging. 4th Meeting of the ACL Special Interest Group on the Lexicon. Washington, DC: Association for Computational Linguistics.
- Burstein, J., Kaplan, R., Wolff, S., & Lu, C. (1996). Using lexical semantic information techniques to classify free responses. In E. Viegas & M. Palmer (Eds.), *Breadth and Depth of Semantic Lexicons*. Workshop Sponsored by the Special Interest Group on the Lexicon. Santa Cruz, CA: Association for Computational Linguistics.
- Burstein, J., Wolff, S., Lu, C., & Kaplan, R. (1997). An automatic scoring system for Advanced Placement biology essays. Fifth Conference on Applied Natural Language Processing. Washington, DC: Association for Computational Linguistics.
- Carlson, L., & Nirenburg, S. (1990). *World Modeling for NLP* [CMU-CMT-90-121]. Pittsburgh, PA: Carnegie Mellon University, Center for Machine Translation.
- Chodorow, M., Byrd, R., & Heidorn, G. (1985). Extracting semantic hierarchies from a large on-line dictionary. 23rd Annual Meeting of the Association for Computational Linguistics. Chicago, IL: Association for Computational Linguistics.
- Copestake, A. (1990). An approach to building the hierarchical element of a lexical knowledge base from a machine-readable dictionary. First International Workshop on Inheritance in Natural Language Processing. Tilburg, The Netherlands.
- Copestake, A. A., & Briscoe, E. J. (1991, June 17). Lexical operations in a unification-based framework. ACL SIGLEX Workshop on Lexical Semantics and Knowledge Representation. Berkeley, CA: Association for Computational Linguistics.
- Davis, A. R. (1996). Lexical semantics and linking in the hierarchical lexicon [Diss], Stanford, CA: Stanford University.
- Flickinger, D. (1987). Lexical rules in the hierarchical lexicon [Diss], Stanford, CA: Stanford University.
- Halliday, M. A., K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hearst, M. A., & Schütze, H. (1996). Customizing a lexicon to better suit a computational task. In B. Boguraev & J. Pustejovsky (Eds.), *Corpus processing for lexical acquisition* (pp. 77-96). Cambridge, MA: The MIT Press.
- Helmreich, S., & Farwell, D. (1996). *Lexical Rules* is italicized. In E. Viegas & M. Palmer (Eds.), *Breadth and Depth of Semantic Lexicons*. Workshop Sponsored by the Special Interest Group on the Lexicon. Santa Cruz, CA: Association for Computational Linguistics.
- Ide, N., & Veronis, J. (1993). Extracting knowledge bases from machine-readable dictionaries: Have we wasted our time? KB&KS93. Tokyo.
- Jacquemin, C., Klavans, J. L., & Tzoukermann, E. (1997). Expansion of multi-word terms for indexing and retrieval using morphology and syntax. 35th Annual Meeting of the Association for Computational Linguistics. Madrid, Spain: Association for Computational Linguistics.
- Klavans, J., Chodorow, M., & Wacholder, N. (1990). From dictionary to knowledge base via taxonomy. 4th Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary: Electronic Text Research. University of Waterloo.

- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: The University of Chicago Press.
- Litkowski, K. C. (1975). *Toward semantic universals*. Delaware Working Papers in Language Studies, No. 18. Newark, Delaware: University of Delaware.
- Litkowski, K. C. (1976). *On Dictionaries and Definitions*. Delaware Working Papers in Language Studies, No. 17. Newark, Delaware: University of Delaware.
- Litkowski, K. C. (1978). Models of the semantic structure of dictionaries. *American Journal of Computational Linguistics*, *Mf.81*, 25-74.
- Litkowski, K. C. (1980, June 19-22). Requirements of text processing lexicons. 18th Annual Meeting of the Association for Computational Linguistics. Philadelphia, PA: Association for Computational Linguistics.
- Litkowski, K. C. (1997). Desiderata for tagging with WordNet synsets and MCCA categories. 4th Meeting of the ACL Special Interest Group on the Lexicon. Washington, DC: Association for Computational Linguistics.
- Litkowski, K. C., & Harris, M. D. (1997). *Category development using complete semantic networks*. Technical Report, vol. 97-01. Gaithersburg, MD: CL Research.
- Longman Dictionary of Contemporary English* (P. Proctor, Ed.). (1978). Harlow, Essex, England: Longman Group.
- Lowe, J. B., Baker, C. F., & Fillmore, C. J. (1997). A frame-semantic approach to semantic annotation. 4th Meeting of the ACL Special Interest Group on the Lexicon. Washington, DC: Association for Computational Linguistics.
- Markowitz, J., Ahlswede, T., & Evens, M. (1986, June 10-13). Semantically Significant Patterns in Dictionary Definitions. 24th Annual Meeting of the Association for Computational Linguistics. New York, NY: Association for Computational Linguistics.
- McRoy, S. W. (1992). Using multiple knowledge sources for word sense discrimination. *Computational Linguistics*, *18*(1), 1-30.
- Mel'čuk, I. A., & Zholkovskiy, A. (1988). The explanatory combinatorial dictionary. In M. W. Evens (Ed.), *Relational models of the lexicon* (pp. 41-74). Cambridge: Cambridge University Press.
- Montemagni, S., & Vanderwende, L. (1993). Structural patterns versus string patterns for extracting semantic information from dictionaries. In K. Jensen, G. Heidorn & S. Richardson (Eds.), *Natural language processing: The PLNLP approach* (pp. 149-159). Boston, MA: Kluwer Academic Publishers.
- Nida, E. A. (1975). *Componential analysis of meaning*. The Hague: Mouton.
- Nirenburg, S., Carbonell, J., Tomita, M., & Goodman, K. (1992, /). *Machine translation: A knowledge-based approach*. San Mateo, CA: Morgan Kaufmann.
- Nirenburg, S., Raskin, V., & Onyshkevych, B. (1995, March 27-29). *Apologiae ontologiae*. J. Klavans (Ed.). AAAI Spring Symposium Series: Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity. Stanford University: American Association for Artificial Intelligence.
- Palmer, M., Rosenzweig, J., Dang, H. T., & Xia, F. (1997). Capturing syntactic/semantic generalizations in a lexicalized grammar. University of Pennsylvania, Philadelphia, PA.

- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: The MIT Press.
- Richardson, S. D. (1997). Determining similarity and inferring relations in a lexical knowledge base [Diss], New York, NY: The City University of New York.
- Sanfilippo, A. (1995, March 27-29). Lexical polymorphism and word disambiguation. In J. Klavans (Ed.), *AAAI Spring Symposium Series: Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity*. Stanford University: American Association for Artificial Intelligence.
- Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum.
- Smadja, F. A., & McKeown, K. R. (1990). Automatically extracting and representing collocations for language generation. 28th Annual Meeting of the Association for Computational Linguistics. Pittsburgh, PA: Association for Computational Linguistics.
- Smadja, F. A., & McKeown, K. R. (1991). Using collocations for language generation. *Computational Intelligence*, 7(4).
- Sowa, J. F. (1984, /). *Conceptual structures: Information processing in mind and machine*. Menlo Park, Calif.: Addison-Wesley.
- UMLS knowledge sources* [7th Experimental Edition]. (1996). Bethesda, MD: National Library of Medicine.
- Voorhees, E. M. (1994, July 3-6). Query expansion using lexical-semantic relations. In W. B. Croft & C. J. van Rijsbergen (Eds.), *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval* (pp. 61-69). Dublin, Ireland: Springer-Verlag.
- Vossen, P. (1991). *Converting data from a lexical database to a knowledge base* [ESPRIT BRA-3030]. ACQUILEX Working Paper, vol. 027.
- Yarowsky, D. (1992). Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. 14th International Conference on Computational Linguistics (COLING92). Nantes, France.