Category Development Using Complete Semantic Networks

Technical Report 97-01

March 1997

CL Research
20239 Lea Pond Place
Gaithersburg, Maryland 20879

Category Development Using Complete Semantic Networks

Kenneth C. Litkowski
CL Research
20239 Lea Pond Place
Gaithersburg, MD 20879
(Email: ken@clres.com)
(Web site: http://www.clres.com)

Mary Dee Harris
Language Technology, Inc.
2153 California St., NW
Washington, DC 20008
(Email: mdharris@acm.org)

1 Abstract

In developing lexicons to support a scoring system for open-ended test questions and a content analysis system, we found it necessary to extend the conventional notion of a semantic network as consisting of concept nodes and (perhaps several kinds of) labeled links by adding nodes for pseudoentries capturing semantic features (as in typed feature structures) and by viewing links as nodes with selectional restrictions on their arguments. For the test scoring system, we incorporated a module for maintaining lexical cohesion, based on the extended semantic network, into a general discourse analysis algorithm. The module accomplishes this objective by requiring discourse entities to be instantiations of lexical relations licensed by the network. For the content analysis dictionary, which assigns categories to entries, we found that category definition required use of lexical semantic information beyond the ISA (and other standard relational) backbone of a semantic network. We conclude that these more "complete" semantic networks are necessary and desirable for various NLP tasks. We found that several principles for category development and extension readily emerged, including principles for characterizing the "category" of a text. These principles have some correspondence with techniques for identifying the topic of a text. We describe some summary procedures for category development and naming.

2 Introduction

Two specific tasks involving lexicon development and analysis have led to the need for the broadened conceptualization of the knowledge contained in semantic networks of lexicons. The first task was to build a knowledge representation of texts (5-6 sentences of 30-100 words) for the Formulating-Hypotheses test item developed at the Educational Testing Service (ETS); this task required the use of lexical cohesion principles to build the discourse structure of the text. The second task involved the analysis of a content-analysis dictionary of 11,000 words.

We describe these tasks, both of which involved use of WordNet (Miller, et al., 1990) and consideration of other semantic networks. We identified shortcomings of these networks in completing the tasks. We present extensions of the nodes and relations in semantic networks, principally by converting labels on links into lexicon entries and by adding pseudoentries corresponding to pieces of lexical semantic information, to enable completion of the tasks. In both tasks, we found that sublexicons can be characterized in a way resembling categorization schemes used in information retrieval (Hearst & Schütze, 1996). We compare these extended networks with previous conceptions of semantic networks and their relation to other classification research and discourse modeling. We suggest, as a result of this analysis, that extension of semantic networks along these lines is necessary and desirable for various NLP tasks, including discourse analysis, tagging, information retrieval (specifically, classification research), topic identification, and content analysis, and present principles for using them for characterization, definition, and naming of categories.

3 Analysis of Text Sample Using Lexical Cohesion

(Burstein, et al., 1996) describe techniques for using lexical semantics to classify responses to test questions. We describe here a separate analysis toward the development of a computer-assisted scoring system.

We base our discussion on the following test item[1]:

> Many seventeenth-century Dutch landscape paintings survive today. The genre was very popular: many artists made their living by supplying such paintings and a few major artists commanded very high commissions for their work. Attributing these paintings (identifying the artist who painted them) is usually done by identifying the signature on the painting. In the absence of signatures, documentary evidence, such as bills of sale from seventeenth-century art dealers, may provide evidence on which to base attributions. In some cases, scholars must rely solely on a work's stylistic characteristics to make attributions. Of the landscape paintings that survive from this period, most have attributions. Paintings attributed to the few major artists painting in this genre outnumber those attributed to minor artists by a surprisingly large margin.
>
> Think of hypotheses (possible explanations) that would account for the disproportionate number of works attributed to major artists.
>
> Write each hypothesis as a separate answer of no more than 15 words.

We created separate sublexicons for the text constituting the test item and the verbs of several hundred responses, extracting pertinent entries from WordNet and a machine-readable dictionary (MRD) using DIMAP. For the text, we had a 59-word dictionary created from WordNet and the MRD (only 25 of which had WordNet information, with the remainder closed-class words not in WordNet). The response set yielded a dictionary of 130 verbs with information from WordNet.

We first examined the 25 nouns, verbs, adjectives, and adverbs in the text to identify obvious relations among them. Despite the small number, we found 5 relations of 2 and 3 nodes, supporting the hypothesis that even with a small text, some lexical cohesion must be present to maintain the comprehensibility of the text.

The verbs in the response set yielded many more subfragments of WordNet. There was also considerably more clustering, with 85 of the verbs rooted at only 5 synsets. The 45 remaining verbs were almost all from what would be labeled a core lexicon of general terms (see (McRoy, 1992)).

We next subjected the text of 7 sentences to a hand discourse analysis that generally followed 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 cohesion module is an elaboration of Allen's step to determine if reference constraints can be satisfied for anaphors to permit the extension of a discourse state and is invoked after following standard methods of resolving anaphors. In the sample text, the NP "the genre", a definite reference, has no obvious referent, so the reference constraint is not satisfied. However, by requiring lexical cohesion, we must find some semantic relationship among previous discourse entities. Here, we establish that "genre" is "a general noun in cohesive function" (Halliday & Hasan, 1976: 275) with "paintings" in the first sentence, using WordNet links to identify a repetition, a synonym or near-synonym, a superordinate, or an acceptable general word. The cohesion module enables us to satisfy the reference constraint and extend the discourse structure.

Some discourse entities ("artists" in sentence 2) are brought into discourse focus without any overt anaphoric indication. The colon in sentence 2 indicates a cataphoric reference, that is, a forward reference (Halliday & Hasan, 1976: 17); its first subordinate clause is a case of non-restrictive apposition that has an identificational equivalence to the matrix (Quirk, et al., 1985: 1308-9). However, to deal with this in the cohesion module, we need to determine how "artists" maintains lexical cohesion with the preceding text. In this case, we find the desired cohesion in the AGENTIVE *quale* of "paintings" (one of (Pustejovsky, 1995: 76)'s four qualia used to characterize the meaning of a noun, in this case, who is the agent or producer of the painting). (This information does not appear in WordNet.) A similar problem occurs for the NP "the signature" in the third sentence. Although not indicated in WordNet, "signature" is part of the CONSTITUTIVE *quale* of "paintings" (that is, one of the physical or abstract components of a painting, as seen in more complete dictionaries). Thus, we see that the semantic

---

[1] Reprinted by permission of Educational Testing Service, the copyright owner. No further reproduction is permitted without written permission of ETS. ETS is not responsible for any analyses and conclusions presented in this paper.

representation of previous discourse entities opens up slots (for agents or other characteristics) that is bound by the discourse entity with focus.

These indirect anaphoric references are frequently characterized as resolvable by world knowledge (Halliday & Hasan, 1976: 70-4; Quirk, et al., 1985: 267- 8), but this does not help us here unless we have such a repository. Instead, invoking the need for lexical cohesion, and making use of information in a dictionary, we were essentially able to establish (or at least posit) world knowledge on the fly.

Incorporation of the lexical cohesion module provides a much richer semantic representation of the discourse. As indicated above, we made use of various slots opened up by considering richer structures for the lexical items. This same process appears to operate with attempts to complete the automated scoring system. We have found (qualitatively) that correct responses fit and extend the discourse structure built from the text, while incorrect responses conflict. This is an issue for further exploration.

4 Content Analysis Category System

Content analysis provides distributional methods for analyzing characteristics of textual material. One content analysis approach, Minnesota Contextual Content Analysis (MCCA) (McTavish & Pirro, 1990), has been used to analyze many types of texts. Subject material has included husbands and wives talking about their relationship, financial services proposed by a bank, interviews of ex-prison-inmates, Likert scale analysis, interviews of nursing home residents and staff, interviews of family caregivers for Alzheimer's patients, and literature.

(McTavish, et al., 1995; McTavish, et al., 1997) suggest that MCCA categories recapitulate WordNet synsets. We used WordNet synsets in examining these categories to determine their coherence and to characterize their relations with WordNet synsets.

We began with the hypothesis that the categories correspond to those developed by (Hearst & Schütze, 1996) in creating categories from the WordNet noun hierarchy. We found that the MCCA categories were generally internally consistent, but with characteristics not intuitively obvious.[2] As a result, we needed to articulate firm principles for characterizing the categories.

In the MCCA dictionary of 11,000 words,[3] the average number of words in a category is 95, with a range from 1 to about 300. Heuristic labels (e.g., *activity, fellow feeling, about changing, human roles, expression arena*) suggest the overall character of word meanings in the category. Using the DIMAP software (CL Research, 1997 - in preparation),[4] we created sublexicons of individual categories, extracted WordNet synsets for these sublexicons, extracted information from the Merriam-Webster Concise Electronic Dictionary integrated with DIMAP, and attached lexical semantic information from other resources to entries in these sublexicons.

Eleven categories (such as *Have, Prepositions, You, I-Me, He, A-An, The*) consist of only a few words from closed classes. About 20 categories (*Implication, If, Colors, Object, Being*) consist of a relatively small number of words (34, 22, 65, 11, 12, respectively) taken primarily from syntactically or semantically closed-class words (subordinating conjunctions, relativizers, the tops of WordNet, colors).

The remaining 80 or so categories consist primarily of open-class words (nouns, verbs, adjectives, and adverbs), sprinkled with closed-class words (auxiliaries, subordinating conjunctions). These categories require more detailed analyses.[5]

The words and senses are first characterized as to part of speech, open- vs. closed-class, nature of morphological relationships. Next, if all words in a category have the same part of speech, we identified whether there are particular (Quirk, et al., 1985) characterizations of the words and whether there are particular subcategorization patterns in COMLEX (Macleod & Grishman, 1994; Wolff, et al., 1995) or the database by (Dorr, in press) based on (Levin, 1993). If

---

[2] In general, we have found that assignment of only about 5 to 10 percent of the words in a category is questionable.

[3] This dictionary has tagged 85 to 95 percent of the words in about 1500 analyses covering 45 million words over 15 years.

[4] A suite of programs for creating and maintaining lexicons for natural language processing, available from CL Research. Procedures used in this paper, applicable to any category analysis using DIMAP, are available at http://www.clres.com.

[5] Analysis of MCCA categories is a continuing process.

the words in the category do not have the same part of speech, we look for common semantic and/or morphological components specializing to each part of speech. In particular, we looked in WordNet for the presence of taxonomic and other relationships in WordNet, asking whether synonyms or other -nyms were in the same or different categories.

In some cases, all the words in a category come together into a tree or a shrub (a small number of small trees, indicating several roots) rooted at a single or a small number of synsets in the WordNet hierarchy. In other cases, the comparisons and contrasts of the words in the category require the use of other parts of a lexical entry than the ISA backbone, as well as other types of information that can be included in a semantic network. Specifically, these include particular semantic features, semantic components, or subcategorization patterns (including specific semantic roles or roles that have specific selectional restrictions on, e.g., the AGENT or PATIENT).

As these pieces of information are added to entries in the subdictionaries, various checks are made on the consistency of the entries, ensuring, e.g., that there are no loops. In this way, the semantic network corresponding to the category is consistently and coherently put together, so that any hierarchical or other relations between senses are clearly identified.

Several categories correspond well to the Hearst & Schütze model, that is, rooted at a single or a small number of synsets in the WordNet hierarchy. The categories *Functional roles, Detached roles,* and *Human roles* align with subtrees rooted at particular nodes in the WordNet hierarchies. For example, *Detached roles* includes (among a total of 66 words):

> ACADEMIC, ARTIST, BIOLOGIST, CREATOR, CRITIC, HISTORIAN, INSTRUCTOR, OBSERVER, PHILOSOPHER, PHYSICIST, PROFESSOR, RESEARCHER, REVIEWER, SCIENTIST, SOCIOLOGIST

These words appear in WordNet synsets headed by (CREATOR; EXPERT: AUTHORITY: PROFESSIONAL; and INTELLECTUAL).[6]

Other categories do not fall out so neatly. The category *Sanction* (120 words) includes the following words (and their inflected forms):

> APPLAUD, APPLAUSE, APPROVE, CONGRATULATE, CONGRATULATION, CONVICT, CONVICTION, DISAPPROVAL, DISAPPROVE, HONOR, JUDGE, JUDGMENT, JUDGMENTAL, MERIT, MISTREAT, REJECT, REJECTION, RIDICULE, SANCTION, SCORN, SCORNFUL, SHAME, SHAMEFULLY

Examination of the WordNet synsets is similarly successful here, identifying many words (particularly verbs) in a subtree rooted at JUDGE. However, the set is confounded by the presence of derived forms from other parts of speech. Another meaning component is seen in APPROVE and DISAPPROVE, namely, the negative or pejorative prefix. This set of words (rooted primarily in the verbs of the set) corresponds to the (Levin, 1993) *Characterize* (class 29.2), *Declare* (29.4), *Admire* (31.2), and Judgment verbs (33) and hence may have particular syntactic and semantic patterning. The verb frames attached to WordNet verb synsets are not sufficiently detailed to cover the granularity necessary to characterize an MCCA category.

This example of words in the *Sanction* category (repeated in other categories) indicates a need to define categories not only in terms of supercategories using the Hearst & Schütze model, but also with additional lexical semantic information not present in WordNet. In particular, we see the need for encoding derivational and morphological relations, finer-grained characterization of government patterns, and primitive semantic components.

5 Extension of Nodes and Relations in Semantic Networks

In performing the two tasks, we were required to use much more information than contained in the ISA backbone and other relations of WordNet. We now describe how this additional information may be included in an expanded conceptualization of a semantic network. In particular, we will extend the nodes to include pseudoentries encapsulizing semantic features (e.g., ±MALE) and various lexical relations (including derivational, inflectional, and metonymic sense extensions). We will transform the relations into nodes (entries) of the network, with semantic roles (associated with subcategorization patterns) and selectional restrictions on these roles.

In our two tasks, we began with the synsets in WordNet, which consist primarily of a tangled hierarchy organized around a backbone of ISA links and other relations between nodes. Some use was made of a similar semantic network, the MetaThesaurus of the Unified Medical Language System (UMLS)

---

[6] Identification of these synsets facilitates extension of the MCCA dictionary to include further hyponyms of these synsets.

(Lindberg, et al., 1993; UMLS knowledge sources, 1996). WordNet contains 23 semantic relations to link nodes, while the UMLS defines and uses a hierarchy of 54 semantic relations.

WordNet includes antonymic, hypernymic, meronymic, troponymic, pertainymic, entailment, and cause relations. The UMLS hierarchy includes the ISA relation and an "associated with" relation that is subdivided into physical, spatial, functional, temporal, and conceptual relations. While some of these relations correspond to the WordNet relations, many of them are realized as verbs (such as *affects, exhibits, indicates, precedes, measures*). Further, the definitions of these relations (e.g., *adjacent to* as including "adjoins, abuts, is contiguous to, is juxtaposed, and is close to" and *surrounds* as including "limits, bounds, confines, encloses, and circumscribes") suggests that lexical entries (that is, nodes) in the verb hierarchy and adjective network can themselves be used as relations.

We were required to include lexical relations and rules in the semantic network. (Flickinger, 1987) first overtly included inflectional relations as distinct nodes in a lexical hierarchy. There are several types of derivational rules that operate in the lexicon. At the most basic, we have morphological relations between lexical entries (e.g., see (Pentheroudakis & Vanderwende, 1993) and (Quirk, et al., 1985: Appendix I) on word formation). These rules operate at the lexical entry level, rather than at the sense level. These rules have not been seen as distinct lexical entries, but that they can and perhaps should be is based on consideration of sense-extension or lexical rules. These rules operate primarily on what have been identified as individual senses of lexical entries (e.g., the *grinding* rule which converts an animal sense of a lexical item into a meat sense, for which see (Copestake & Briscoe, 1991), where the rule itself is placed among the lexical entries and is viewed as providing the conditions for the transformation of one sense into another).

We also were required to consider semantic features associated with lexical items (Katz & Fodor, 1963). (Copestake & Briscoe, 1991) showed that such features could be identified as distinct lexical items (hereafter termed *pseudoentries*) that can be inherited by regular lexical entries. We also used the qualia structures of (Pustejovsky, 1995). The formal (that is, what is this) aspect of a qualia structure corresponds to the ISA backbone of WordNet, while the constitutive aspect corresponds to its meronymic links. The agentive and telic (the purpose associated with the noun) qualia are not currently included in WordNet. We may view these aspects as relations in the UMLS hierarchy and hence give them independent status as lexical entries with selectional restrictions identifying which nodes of the network may satisfy the restrictions.

The final relation is the semantic role. For lexical items having subcategorization patterns, such as verbs identifying cases, these relations provide links to other entries. This is done by identifying selectional restrictions, which are nothing more than the restriction of, say, the AGENT or INSTRUMENT to a subset of the semantic network.

In general, then, nodes of the semantic network may be viewed as lexical entries or pseudoentries representing concepts. Semantic relations may be viewed as labels on the links between nodes. However, these relations are better viewed as distinct nodes in the network, with the property that they act as a linking mechanism between other nodes; semantic relations may also exist within their own hierarchy. (These are similar to role nodes in (Brachman & Schmolze, 1985).)

6 Comparison of Extended Semantic Networks with Previous Work

The sublexicons identified in the analysis of a cohesive text used as a test item, a cohesive set of responses to that test item, and a pre-defined category system clearly carve out portions of semantic networks. While some parts of these sublexicons correspond to subtrees rooted at particular nodes in these networks, several additional useful "cuts" through the network have been identified.

(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. The work done here supports that observation, but does so in the context of a much larger conception of semantic relations among nodes in a network.

The identification of sublexicons that cut across parts of speech corresponds to the presence of semantic components in identifying categories. This notion is supported in (Burstein, et al., 1996), where concept grammars are based on collapsing phrasal and constituent nodes into a generalized XP representation, This notion of using semantic components as the basis for category development has been invoked as well in (Laffal, 1995).

The semantic network carved out by adhering to principles of lexical cohesion is the most interesting. In the first place, the principal nodes of the network are not in themselves naturally related, thus following Nida in allowing the use of "on the fly" criteria for defining a sublexicon. In the second place, the need to invoke semantic roles (and their selectional restrictions) in analyzing the discourse structure clearly develops "on the fly" conceptual clusters, similar to those described in (McRoy, 1992), particularly situational clusters. The relationship between the sublexicon formed from the ETS response set and the ETS test item suggests further that the identified sub-semantic network has affinities to scripts as described by (Schank & Abelson, 1977). The procedures described here are consistent with those described in (Barrière & Popowich, 1996) and indeed would benefit from those procedures in further automating the identification of situational clusters.

7 Conclusions: Use of Extended Semantic Networks for Characterization, Definition, and Naming of Categories

Category development is important in many fields of research and is usually approached with statistical techniques of classification based on assignment of characteristics to units of analysis. In fields associated with linguistics, category development has been most often used for content analysis and information retrieval. In these fields, after the development of a category system and identification of lexemes judged to lie within each category, the analysis of text is then based on the frequency with which the lexemes of a category occur within units of analysis. The field of information retrieval performs this last step, usually with only a rudimentary category system based on identifying the root words of lexemes. However, within information retrieval, considerable promise (Liddy, et al., 1993) has been obtained by using a category system based on the subject codes in Longman's Dictionary of Contemporary English (LDOCE) (Longman Dictionary of Contemporary English, 1978).

Labeling texts with a pre-defined set of categories, even using such categories for identifying the principal topics of texts, has shown a clear benefit of customizing lexicons to suit computational tasks (Hearst & Schütze, 1996). This approach to customizing the lexicon built upon the semantic network embodied in WordNet, but made use of only its principal ISA backbone, that is, the hypernymic and hyponymic relations between synsets.

The most straightforward categories are those characterized as lists of subtrees in an ISA hierarchy. The definition of such a category is simply the set of highest nodes that cover all concepts in the category (as described for the MCCA category *Detached roles*). Each node should be a concept in the category or a node such that any of its children in the category have that node as its direct parent. It is legitimate to exclude subtrees rooted at the descendants of the nodes in the set. The name for the category is equivalent to the set of nodes defining the category. The category systems described in (Nida, 1975; Hearst & Schütze, 1996; Laffal, 1995) are of this type.

Another relatively straightforward type of category is that formed with lexical items that have well-defined morphological or derivational relations between pairs of the concepts. This group can also contain other types of lexical-rule relations. These morphological, derivational, and lexical-rule relations are subsidiary within the category. There is a set of semantic components that lie below them; these are the principal defining characteristics of the category, handled as in the previous type of category. The category system in (Burstein, et al., 1996), which generalizes part of speech, is of this type.

The most complex type of category system is the one necessary to describe the discourse structure of a text. When a text has been analyzed as described in the previous section, the result is a set of discourse segments related to one another (with many identified as subsidiary), discourse entities and eventualities, and various (primarily) role relations between these entities. We view the concepts and relations (including the discourse relations) as essentially present in and licensed by the lexicon, and then instantiated by the given text to carve out a subnetwork. The definition of this subnetwork is 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 is again a matter of identifying the topmost ISA nodes (and perhaps more importantly here, more accurately identifying descendants that are to be excluded). Naming this subnetwork is again based on the set of topmost nodes, which may in this case include relations (as from the UMLS tree of semantic relations). This process of characterizing a subnetwork is quite similar to the development of supercategories in (Hearst & Schütze, 1996). Thus, to at least that extent, this process may be viewed as leading to identification of the topic of a text.

8 Acknowledgements

We would like to thank Jill Burstein, Robert Amsler, Don McTavish, Thomas Pötter, Adam Kilgarriff, Bonnie Dorr, Lisa Rau, Marti Hearst, and anonymous reviewers for their discussions and comments on issues

relating to this paper and its initial draft.

9 References

Allen, J. (1995). *Natural language understanding* (2nd). Redwood City, CA: The Benjamin/Cummings Publishing Company, Inc.

Barrière, C., & Popowich, F. (1996). Concept clustering and knowledge integration from a children's dictionary. COLING96.

Brachman, R., & Schmolze, J. (1985). An overview of the KL-ONE knowledge representation language. *Cognitive Science*, 171-216.

Burstein, J., Kaplan, R., Wolff, S., & Lu, C. (1996, June). 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.

CL Research. (1997 - in preparation). *DIMAP-3 users manual*. Gaithersburg, MD.

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.

Dorr, B. (in press). Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Journal of Machine Translation*.

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.

Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language*, *39*, 170-210.

Laffal, J. (1995, October). A concept analysis of Jonathan Swift's *A Tale of a Tub* and *Gulliver's Travels*. *Computers and the Humanities*, pp. 339-361.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: The University of Chicago Press.

Liddy, E. D., Paik, W., & Yu, E. S. (1993, June 22). Document filtering using semantic information from a machine readable dictionary. Workshop on Very Large Corpora: Academic and Industrial Perspectives. Columbus, OH: Association for Computational Linguistics.

Lindberg, D. A., B., Humphreys, B. L., & McCray, A., T. (1993). The Unified Medical Language System. *Methods of Information in Medicine*, *32*, 281-291.

Longman Dictionary of Contemporary English (P. Proctor, Ed.). (1978). Harlow, Essex, England: Longman Group.

Macleod, C., & Grishman, R. (1994). *COMLEX syntax reference manual*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

McRoy, S. W. (1992). Using multiple knowledge sources for word sense discrimination. *Computational Linguistics*, *18*(1), 1-30.

McTavish, D. G., Litkowski, K. C., & Schrader, S. (1995, September). A computer content analysis approach to measuring social distance in residential organizations for older people. Society for Content Analysis by Computer. Mannheim, Germany.

McTavish, D. G., Litkowski, K. C., & Schrader, S. (1997). A computer content analysis approach to measuring social distance in residential organizations for older people. *Social Science Computer Review*, in press.

McTavish, D. G., & Pirro, E. B. (1990). Contextual content analysis. *Quality & Quantity*, *24*, 245-265.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235-244.

Nida, E. A. (1975). *Componential analysis of meaning*. The Hague: Mouton.

Pentheroudakis, J., & Vanderwende, L. (1993). Automatically identifying morphological relations in machine readable dictionaries (pp. 114-131). Oxford, England.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: The MIT Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Lawrence Erlbaum.

UMLS knowledge sources [7th Experimental Edition]. (1996). Bethesda, MD: National Library of Medicine.

Wolff, S. R., Macleod, C., & Meyers, A. (1995). *COMLEX word classes*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.