NATURAL LANGUAGE PROCESSING

Thomas C. Rindflesch

INTRODUCTION

Work in computational linguistics began very soon after the development of the first computers (Booth, Brandwood and Cleave 1958), yet in the intervening four decades there has been a pervasive feeling that progress in computer understanding of natural language has not been commensurate with progress in other computer applications. Recently, a number of prominent researchers in natural language processing met to assess the state of the discipline and discuss future directions (Bates and Weischedel 1993). The consensus of this meeting was that increased attention to large amounts of lexical and domain knowledge was essential for significant progress, and current research efforts in the field reflect this point of view.

The traditional approach in computational linguistics (Allen 1987) included a prominent concentration on the formal mechanisms available for processing language, especially as these applied to syntactic processing and, somewhat less so, to semantic interpretation. In recent efforts, work in these areas continues, but there has been a marked trend toward enhancing these core resources with statistical knowledge acquisition techniques. There is considerable research aimed at using online resources for assembling large knowledge bases, drawing on both natural language corpora and dictionaries and other structured resources. Recent research in lexical semantics reflects an interest in the proper structuring of this information to support linguistic processing. Furthermore, the availability of large amounts of machine-readable text naturally supports continued work in analysis of connected discourse. In other trends the use of statistical techniques are being used as part of the parsing process, for automatic part of speech assignment, and for word-sense disambiguation.

An indication of the development of natural language processing systems is that they are increasingly being used in support of other computer programs. This trend is particularly noticeable with regard to information management applications. Natural language processing provides a potential means of gaining access to the information inherent in the large amount of text made available through the Internet. In the following survey I look in further detail at the recent trends in research in natural language processing and conclude with a discussion of some applications of this research to the solution of information management problems.

CORPUS LINGUISTICS

The papers in Zampolli, Calzolari and Palmer 1994 (Section 3) discuss issues relevant to the design, acquisition, and use of corpora for computational linguistics, with an emphasis on design and acquisition. The papers in Oostdijk and de Haan 1994 and in Fries, Tottie and Schneider 1994 also discuss the design and acquisition of large corpora, but in both volumes (especially Fries, Tottie and Schneider 1994) there are a number of papers describing research that uses these texts. In Oostdijk and de Haan 1994 the emphasis is on computational linguistics (e.g. automatic tagging and parsing), while in Fries, Tottie and Schneider 1994 there are a number of studies which use evidence from corpora to support research of interest to the general linguist, including distributional characteristics of words for stylistics. Marcus, Santorini and Marcinkiewicz (1993) discuss the Penn Treebank, which is of particular interest because it is a large corpus (4.5 million words) with annotations. All words have been tagged with part-of-speech labels, and more than half of the text has been analyzed with partial syntactic structure (labelled brackets).

THE LEXICON

In computational linguistics the lexicon supplies paradigmatic information about words, including part of speech labels, irregular plurals, and subcategorization information for verbs. Traditionally, lexicons were quite small and were constructed largely by hand. There is a growing realization that effective natural language processing requires increased amounts of lexical (especially semantic) information. A recent trend has been the use of automatic techniques applied to large corpora for the purpose of acquiring lexical information from text (Zernik 1991). Statistical techniques are an important aspect of automatically mining lexical information. Manning (1993) uses such techniques to gather subcategorization information for verbs. Brent (1993) also discovers subcategorization information; in addition he attempts to automatically discover verbs in the text. Liu and Soo (1993) describe a method for mining information about thematic roles.

The additional information being added to the lexicon increases the complexity of the lexicon. This added complexity requires that attention be paid to the organization of the lexicon: Zernik 1991 (Part III) and Pustejovsky 1993 (Part III) both contain several papers which address this issue. McCray, Srinivasan and Browne(1993) discuss the structure of a large (more than 60,000 base forms) lexicon designed and implemented to support syntactic processing.

AUTOMATIC TAGGING

Automatically disambiguating part-of-speech labels in text is an important research area since such ambiguity is particularly prevalent in English. Programs resolving part-of-speech labels (often called automatic taggers) typically are around 95% accurate. Taggers can serve as preprocessors for syntactic parsers and contribute significantly to efficiency. There have been two main approaches to automatic tagging: probabilistic and rule-based. Merialdo (1994) and Dermatos and Kokkinakis (1995) review several approaches to probabilistic tagging and then offer new proposals. Typically, probabilistic taggers are trained on disambiguated text and vary as to how much training text is needed and how much human effort is required in the training process. (See

Schütze 1993 for a tagger that requires very little human intervention.) Further variation concerns knowing what to do about unknown words and the ability to deal with large numbers of tags.

One drawback to stochastic taggers is that they are very large programs requiring considerable computational resources. Brill (1992) describes a rule-based tagger which is as accurate as stochastic taggers, but with a much smaller program. The program is slower than stochastic taggers, however. Building on Brill's approach, Roche and Schabes (1995) propose a rule-based, finite-state tagger which is much smaller and faster than stochastic implementations. Accuracy and other characteristics remain comparable.

PARSING

The traditional approach to natural language processing takes as its basic assumption that a system must assign a complete constituent analysis to every sentence it encounters. The methods used to attempt this are drawn from mathematics, with context-free grammars playing a large role in assigning syntactic constituent structure. Partee, ter Meulen and Wall (1993) provide an accessible introduction to the theoretical constructs underlying this approach, including set theory, logic, formal language theory, and automata theory, along with the application of these mechanisms to the syntax and semantics of natural language.

The program described in Alshawi 1992 is a very good example of a complete system built on these principles. For syntax, it uses a unification-based implementation of a generalized phrase structure grammar (Gazdar *et al.* 1985) and handles an impressive number of syntactic structures which might be expected to appear in "interactive dialogues with information systems...although of course there is still a large residue even of this variety of English that the system fails to analyze properly." (Alshawi 1992:61).

In continuing research in this tradition, context-free grammars have been extended in various ways. The so-called "mildly context sensitive grammars," such as tree adjoining grammars, have had considerable influence on recent work concerned with the formal aspects of parsing natural language (e.g. Satta 1994, Schabes and Shieber 1994).

Several recent papers pursue nontraditional approaches to syntactic analysis. One such technique is partial, or underspecified, analysis. For many applications such an analysis is entirely sufficient and can often be more reliably produced than a fully specified structure. Chen and Chen (1994), for example, employ statistical methods combined with a finite state mechanism to impose an analysis which consists only of noun phrase boundaries, without specifying their complete internal structure or their exact place in a complete tree structure. Agarwal and Boggess (1992) successfully rely on semantic features in a partially specified syntactic representation for the identification of coordinate structures. In an innovative application of dependency grammar and dynamic programming techniques, Kurohashi and Nagao (1994) address the problem of analyzing very complicated coordinate structures in Japanese.

A recent innovation in syntactic processing has been investigation into the use of statistical techniques. (See Charniak 1993 for an overview of this and other statistical applications.) In probabilistic parsing, probabilities are extracted from a parsed corpus for the purpose of choosing the most likely rule when more than one rule can apply during the course of a parse (Magerman and Weir 1992). In another application of probabilistic parsing the goal is to choose the (semantically) best analysis from a number of syntactically correct analyses for a given input (Briscoe and Carroll 1993, Black, Garside and Leech 1993).

A more ambitious application of statistical methodologies to the parsing process is grammar induction where the rules themselves are automatically inferred from a bracketed text; however, results in the general case are still preliminary. Pereira and Schabes (1992) discuss inferring a grammar from bracketed text relying heavily on statistical techniques, while Brill (1993) uses only modest statistics in his rule-based method.

WORD-SENSE DISAMBIGUATION

Automatic word-sense disambiguation depends on the linguistic context encountered during processing. McRoy (1992) appeals to a variety of cues while parsing, including morphology, collocations, semantic context, and discourse. Her approach is not based on statistical methods, but rather is symbolic and knowledge intensive. Statistical methods exploit the distributional characteristics of words in large texts and require training, which can come from several sources, including human intervention. Gale, Church and Yarowsky (1992) give an overview of several statistical techniques they have used for word-sense disambiguation and discuss research on evaluating results for their systems and others. They have used two training techniques, one based on a bilingual corpus, and another on *Roget's Thesaurus*. Justeson and Katz (1995) use both rule-based and statistical methods. The attractiveness of their method is that the rules they use provide linguistic motivation.

SEMANTICS

Formal semantics is rooted in the philosophy of language and has as its goal a complete and rigorous description of the meaning of sentences in natural language. It concentrates on the structural aspects of meaning. Chierchia and McConnell-Ginet (1990) provide a good introduction to formal semantics. The papers in Rosner and Johnson 1992 discuss various aspects of the use of formal semantics in computational linguistics and focus on Montague grammar (Montague 1974), although Wilks (1992) dissents from the prevailing view. King (1992) provides an overview of the relation between formal semantics and computational linguistics. Several papers in Rosner and Johnson discuss research in the situation semantics paradigm (Barwise and Perry 1983), which has recently had wide influence in computational linguistics, especially in discourse processing. See Alshawi 1992 for a good example of an implemented (and eclectic) approach to semantic interpretation.

Lexical semantics (Cruse 1986) has recently become increasingly important in natural language processing. This approach to semantics is concerned with psychological facts associated with the meaning of words. Levin (1993) analyzes verb classes within this framework, while the papers in Levin and Pinker 1991 explore additional phenomena, including the semantics of events and verb argument structure. A very interesting application of lexical semantics is WordNet

(Miller 1990), which is a lexical database that attempts to model cognitive processes. The articles in Saint-Dizier and Viegas 1995 discuss psychological and foundational issues in lexical semantics as well as a number of aspects of using lexical semantics in computational linguistics.

Another approach to language analysis based on psychological considerations is cognitive grammar (Langacker 1988). Olivier and Tsujii (1994) deal with spatial prepositions in this framework, while Davenport and Heinze (1995) discuss more general aspects of semantic processing based on cognitive grammar.

DISCOURSE ANALYSIS

Discourse analysis is concerned with coherent processing of text segments larger than the sentence and assumes that this requires something more than just the interpretation of the individual sentences. Grosz, Joshi and Weinstein (1995) provide a broad-based discussion of the nature of discourse, clarifying what is involved beyond the sentence level, and how the syntax and semantics of the sentences support the structure of the discourse. In their analysis, discourse contains linguistic structure (syntax, semantics), attentional structure (focus of attention), and intentional structure (plan of participants) and is structured into coherent segments. During discourse processing one important task for the hearer is to identify the referents of noun phrases. Inferencing is required for this identification. A coherent discourse lessens the amount of inferencing required of the hearer for comprehension. Throughout a discourse the particular way that the speaker maintains “focus of attention” or “centering” through choice of linguistic structures for referring expressions is particularly relevant to discourse coherence.

Other work in computational approaches to discourse analysis has focused on particular aspects of processing coherent text. Hajicova, Skoumalova and Sgall (1995) distinguish topic (old information) from focus (new information) within a sentence. Information of this sort is relevant to tracking focus of attention. Lappin and Leass (1994) are primarily concerned with intrasentential anaphora resolution, which relies on syntactic, rather than discourse, cues. However, they also address intersentential anaphora, and this relies on several discourse cues, such as saliency of an NP, which is straightforwardly determined by such things as grammatical role, frequency of mention, proximity, and sentence recency. Huls, Bos and Claasen (1995) use a similar notion of saliency for anaphora resolution and resolve deictic expressions with the same principles. Passonneau and Litman (1993) study the nature of discourse segments and the linguistic structures which cue them. Sonderland and Lehnert (1994) investigate machine learning techniques for discovering discourse-level semantic structure.

Several recent papers investigate those aspects of discourse processing having to do with the psychological state of the participants in a discourse, including, goals, intentions, and beliefs: Asher and Lascarides (1994) investigate a formal model for representing the intentions of the participants in a discourse and the interaction of such intentions with discourse structure and semantic content. Traum and Allen (1994) appeal to the notion of social obligation to shed light on the behavior of discourse. Wiebe (1994) investigates psychological point of view in third person narrative and provides an insightful algorithm for tracking this phenomenon in text. The point of view of each sentence is either that of the narrator or any one of the characters in the narrative.

Wiebe discusses the importance of determining point of view for a complete understanding of a text, and discusses how this interacts with other aspects of discourse structure.

APPLICATIONS

As natural language processing technology matures it is increasingly being used to support other computer applications. Such use naturally falls into two areas, one in which linguistic analysis merely serves as an interface to the primary program, and another in which natural language considerations are central to the application.

Natural language interfaces to data base management systems (e.g. Bates 1989) translate users' input into a request in a formal data base query language, and the program then proceeds as it would without the use of natural language processing techniques. It is normally the case that the domain is constrained and the language of the input consists of comparatively short sentences with a constrained set of syntactic structures.

The design of question answering systems is similar to that for interfaces to data base management systems. One difference, however, is that the knowledge base supporting the question answering system does not have the structure of a data base. See, for example Kupiec 1993, where the underlying knowledge base is an on-line encyclopedia. Processing in this system not only requires a linguistic description for users' requests, but it is also necessary to provide a representation for the encyclopedia itself. As with the interface to a DBMS, the requests are likely to be short and have a constrained syntactic structure. Lauer, Peacock and Graesser (1992) provide some general considerations concerning question answering systems and describe several applications.

In message understanding systems, a fairly complete linguistic analysis may be required, but the messages are relatively short and the domain is often limited. Davenport and Heinze (1995) describe such a system in a military domain. See Chinchor, Hirschman and Lewis 1993 for an overview of some recent message understanding systems.

In three closely related applications (information filtering, text categorization, and automatic abstracting) no constraints on the linguistic structure of the documents being processed can be assumed. One mitigating factor, however, is that effective processing may not require a complete analysis. For all of these applications there are also statistically based systems based on frequency distributions of words. These systems work fairly well, but most people feel that for further improvements, and for extensions, some sort of understanding of the texts, such as that provided by linguistic analysis, is required.

Information filtering and text categorization are concerned with comparing one document to another. In both applications, natural language processing imposes a linguistic representation on each document being considered. In text categorization a collection of documents is inspected and all documents are grouped into several categories based on the characteristics of the linguistic representations of the documents. Blosseville *et al.* (1992) describe an interesting system which combines natural language processing, statistics, and an expert system. In information filtering, documents satisfying some criterion are singled out from a collection. Jacobs and Rau (1990) discuss a program which imposes a quite sophisticated semantic representation for this purpose.

In automatic abstracting, a summary of each document is sought, rather than a classification of a collection. The underlying technology is similar to that used for information filtering and text categorization: the use of some sort of linguistic representation of the documents. Of the two major approaches, one (e.g. McKeown and Radev 1995) puts more emphasis on semantic analysis for this representation and the other (e.g. Paice and Jones 1993), less.

Information retrieval systems typically allow a user to retrieve documents from a large bibliographic database. During the information retrieval process a user expresses an information need through a query. The system then attempts to match this query to those documents in the database which satisfy the user's information need. In systems which use natural language processing, both query and documents are transformed into some sort of a linguistic structure, and this forms the basis of the matching. Several recent information retrieval systems employ varying levels of linguistic representation for this purpose. Sembok and van Rijsbergen (1990) base their experimental system on formal semantic structures, while Myaeng, Khoo and Li (1994) construct lexical semantic structures for document representations. Strzalkowski (1994) combines syntactic processing and statistical techniques to enhance the accuracy of representation of the documents. In an innovative approach to document representation for information retrieval, Liddy *et al.* *(*1995) use several levels of linguistic structure, including lexical, syntactic, semantic, and discourse.

ANNOTATED BIBLIOGRAPHY

Allen, J. 1987. *Natural language understanding.* Menlo Park, CA: The Benjamin/Cummings Publishing Company, Inc.

This is a very useful and accessible introduction to natural language processing. The book covers a wide range of the components involved in writing computer programs for understanding natural language, including the major approaches to parsing, semantic interpretation, discourse analysis, and representation of domain knowledge. For parsing, for example, such core topics as the top-down and bottom-up methods as well as unification are covered in considerable detail. An admirable characteristic of the book is that the phenomena of natural language which motivate particular computational mechanisms are given ample treatment. In the section on semantic interpretation, there is a separate chapter, for example, dealing with scoping phenomena, nominal modification, and tense and aspect. Each chapter includes exercises and there is an extensive index and bibliography as well as an appendix providing background material on symbolic computation and logic and rules of inference.

Bates, M. and R. M. Weischedel (eds.) 1993. *Challenges in natural language processing.* Cambridge: Cambridge University Press.

The work in this volume represents the results of a conference attended by prominent researchers in computational linguistics. The goal of the meeting was to assess the current state of the discipline and plan future directions. The consensus of the conference was that the most pressing concerns center around the need for large amounts of linguistic and world knowledge to support natural language processing systems. Reflecting this concern, half of the papers deal with lexical issues and knowledge representation. (There are also articles on discourse processing and speech technology.) Of particular interest are the first and last papers, which discuss critical challenges for natural language processing and outline a plan for the future. The book is useful not only for the insight into the current state of the discipline but also for the substantive research reported.

Rosner, M. and R. Johnson (eds.) 1992. *Computational linguistics and formal semantics*. Cambridge: Cambridge University Press.

The ten papers contained in this volume present a wide range of the practical and theoretical concerns that underpin current thinking on semantic interpretation in computational linguistics. After an introduction which clearly summarizes the remaining work, the bulk of the book (3 papers) deals with issues in Montague semantics, which continues to have substantial influence in the field. However, situation semantics, which is of particular interest to researchers in discourse processing, is also covered. There are papers on mechanisms for implementing semantic interpretation and an insightful discussion of the relation between form and content in semantics. The final paper is a very useful discussion of the relation between computational linguistics and formal semantics. The references from all papers are included in a comprehensive bibliography at the end of the volume. There is no index.

Saint-Dizier, P. and E. Viegas (eds.) 1995. *Computational lexical semantics*. Cambridge: Cambridge University Press.

This collection of papers (20 articles) provides a broad treatment of knowledge representation as it pertains to lexical information, a topic that is of central concern in current research on natural language processing. There is a valuable introduction which gives an overview of the field and serves as the background for the remaining articles in the volume. The papers are written by researchers working in a range of disciplines. Issues related to the representation and use of lexical information are discussed in detail from several points of view, including the cognitive and linguistic foundations of lexical representation. The particular concerns of the lexicon are related to broader issues in computer science, such as knowledge representation, artificial intelligence, and applications development. The book includes a subject and author index.

Wiebe, J. M. 1994. Tracking point of view in narrative. *Computational Linguistics* 20.2.233-287.

This substantial article provides an excellent example of recent work in computational linguistics on discourse analysis. Fictional narrative discourse contains objective sentences, which describe the fictional world, and subjective sentences, which describe a character's inner state and take that character's psychological point of view. In order to adequately understand fiction it is necessary to distinguish objective from subjective sentences, and for subjective sentences, it is necessary to determine which character's point of view is being taken. Drawing on computer science, linguistics, and literary theory and the extensive use of example text from more than ten novels this article satisfyingly illuminates an intricate phenomenon in discourse. The detailed algorithm for identifying subjective sentences and determining whose inner state is being described is admirably lucid and accessible.

UNANNOTATED BIBLIOGRAPHY

Agarwal, R. and L. Boggess. 1992. A simple but useful approach to conjunct identification. In *Proceedings of the 30th annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 15-21.

Alshawi, H. (ed.) 1992. *The core language engine*. Cambridge, MA: The MIT Press.

Asher, N. and A. Lascarides. 1994. Intentions and information in discourse. In *Proceedings of the 32nd annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 34-41.

Barwise, J. and J. Perry. 1983. *Situations and attitudes*. Cambridge, MA: The MIT Press.

Bates, M. 1989. Rapid porting of the Parlance Natural Language Interface. In *Proceedings of the speech and natural language workshop*. San Mateo, CA: Morgan Kaufmann Publishers. 83-88.

Black, E., R. Garside and G. Leech (eds.) 1993. *Statistically-driven computer grammars of English: The IBM/Lancaster approach*. Amsterdam: Editions Rodopi.

Blosseville, M.J., *et al*. 1992. Automatic document classification: Natural language processing, statistical analysis, and expert system techniques used together. In N. Belkin, P. Ingwesen and A. M. Pejtersen (eds.) *Proceedings of the 15th annual international ACM SIGIR conference on research and development in information retrieval.* NewYork: Association for Computing Machinery. 51-58.

Booth, A. D., L Brandwood and J. P. Cleave. 1958. *Mechanical resolution of linguistic problems.* London: Butterworths Scientific Publications.

Brent, M. R. 1993. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics* 19.2.243-262.

Brill, E. 1992. A simple rule-based part of speech tagger. In *Proceedings of the third conference on applied natural language processing.* Trento, Italy. San Francisco: Morgan Kaufmann Publishers. 152-155.

______ 1993. Automatic grammar induction and parsing free text: A transformation-based approach. In *Proceedings of the 31st annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 259-265.

Briscoe, T. and J. Carroll. 1993. Generalized probabilistic LR parsing of natural language (corpora) with unification-based grammars. *Computational Linguistics* 19.1.25-59.

Charniak, E. 1993. *Statistical language learning.* Cambridge, MA: The MIT Press.

Chen, K.-H. and H.-H. Chen. 1994. Extracting noun phrases from large-scale texts: A hybrid approach and its automatic evaluation. In *Proceedings of the 32nd annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 234-241.

Chierchia, G. and S. McConnell-Ginet. 1990. *Meaning and grammar: An introduction to semantics.* Cambridge, MA: The MIT Press.

Chinchor, N., L. Hirschman and D. D. Lewis. 1993. Evaluating message understanding systems: An analysis of the Third Message Understanding Conference (MUC-3). *Computational Linguistics* 19.3.409-450.

Cruse, D. A. 1986. *Lexical semantics.* Cambridge: Cambridge University Press.

Davenport, D. M. and D. T. Heinze. 1995. Crisis action message analyzer - EDM. *Proceedings of the 5th annual dual-use technologies and applications conference*. SUNY Institute of Technology at Utica/Rome, NY. 284-289.

Dermatas, E. and G. Kokkinakis. 1995. Automatic stochastic tagging of natural language texts. *Computational Linguistics* 21.2.137-163.

Fries, U., G. Tottie and P. Schneider (eds.) 1994. *Creating and using English language corpora: Papers from the fourteenth international conference on English language research on computerized corpora, Zurich 1993.* Amsterdam: Editions Rodopi.

Gale, W., K. W. Church and D. Yarowsky. 1992. Estimating upper and lower bounds on performance of word-sense disambiguation programs. In *Proceedings of the 30th annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 249-256.

Gazdar, G., *et al.* 1985. *Generalized phrase structure grammar.* Oxford: Blackwell Publishing and Cambridge, MA: Harvard University Press.

Grosz, B. J., A. K. Joshi and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21.2.203-225.

Hajicova, E., H. Skoumalova and P. Sgall. 1995. An automatic procedure for topic-focus identification. *Computational Linguistics* 21.1.81-94.

Huls, C., E. Bos and W. Claasen. 1995. Automatic referent resolution of deictic and anaphoric expressions. *Computational Linguistics* 21.1.59-79.

Jacobs, P. S. and L. F. Rau. 1990. SCISOR: Extracting information from on-line news. *Communications of the ACM* 33.11.88-97.

Justeson, J. S. and S. M. Katz. 1995. Principled disambiguation: Discriminating adjective senses with modified nouns. *Computational Linguistics* 21.1.1-27.

King, M. 1992. Epilogue: On the relation between computational linguistics and formal semantics. In M. Rosner and R. Johnson (eds.) *Computational linguistics and formal semantics*. Cambridge: Cambridge University Press. 283-299.

Kupiec, J. 1993. MURAX: A robust linguistic approach for question answering using an on-line encyclopedia. In R. Korfhage, E. Rasmussen and P. Willett (eds.) *Proceedings of the 16th* *annual international ACM SIGIR conference on research and development in information retrieval.* New York: Association for Computing Machinery. 181-190.

Kurohashi, S. and M. Nagao. 1994. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. *Computational Linguistics* 20.4.507-534.

Langacker, R. W. 1988. An overview of cognitive grammar. In B. Rudzka-Ostyn (ed.) *Topics in cognitive linguistics.* Amsterdam/Philadelphia: John Benjamins Publishing Company. 3-48.

Lappin, S. and H. J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics* 20.4.535-561.

Lauer, T. W., E. Peacock and A. C. Graesser (eds.) 1992. *Questions and information systems.* Hillsdale, NJ: Lawrence Erlbaum.

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

_______ and S. Pinker (eds.) 1991. *Lexical and conceptual semantics.* Oxford: Blackwell Publishing.

Liddy, E. D., *et al.* 1994. Document retrieval using linguistic knowledge." *RIAO 94 conference proceedings.* 106-114. [Organized by the Center for Advanced Study of Information systems, Inc. and Centre de Hautes Etudes Internationales d'Informatique Documentaire.]

Liu, R.-L. and V. W. Soo. 1993. An empirical study on thematic knowledge acquisition based on syntactic clues and heuristics. In *Proceedings of the 31st annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 243-250.

Magerman, D. M. and C. Weir. 1992. Efficiency, robustness and accuracy in *P*icky chart parsing. In *Proceedings of the 30th annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 40-47.

Manning, C. D. 1993. Automatic acquisition of a large subcategorization dictionary from corpora. In *Proceedings of the 31st annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 235-242.

Marcus, M. P., B. Santorini and M. A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19.2.313-330.

McCray, A., S. Srinivasan and A. C. Browne. 1994. Lexical methods for managing variation in biomedical terminologies. In J. G. Ozbolt (ed.) *Proceedings of the eighteenth annual symposium on computer applications in medical care.* Philadelphia: Hanley & Belful, Inc. 235-239.

McKeown, K. and D. R. Radev. 1995. Generating summaries of multiple news articles. In E. A. Fox, P. Ingwersen and R. Fidel (eds.) *Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval.* New York: Association for Computing Machinery. 74-82.

McRoy, S. W. 1992. Using multiple knowledge sources for word sense discrimination. *Computational Linguistics* 18.1.1-30.

Merialdo, B. 1994. Tagging English text with a probabilistic model. *Computational Linguistics* 20.2.155-171.

Miller, G. A. 1990. WordNet: An on-line lexical database. *International Journal of Lexicography* 3.235-312.

Montague, R. 1974. *Formal philosophy: Selected papers of Richard Montague.* [ed. R. H. Thomason.] New Haven: Yale University Press.

Myaeng, S. H., C. Khoo and M. Li. 1994. Linguistic processing of text for a large-scale conceptual information retrieval system. In W. M. Tepfenhart, J. P. Dick and J. F. Sowa (eds.) *Conceptual structures: Current practices*. Berlin: Springer-Verlag. 69-83. [Lecture notes in artificial intelligence 835.]

Olivier, P. and J. Tsujii. 1994. A computational view of the cognitive semantics of spatial prepositions. In *Proceedings of the 32nd annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 303-309.

Oostdijk, N. and P. de Haan (eds.) 1994. *Corpus-based research into language: In honour of Jan Aarts.* Amsterdam: Editions Rodopi.

Paice, C. D. and P. A. Jones. 1993. The identification of important concepts in highly structured technical papers. *Proceedings of the 16th annual international ACM conference on research and development in information retrieval.* 69-78.

Partee, B. H., A. ter Meulen and Robert E. Wall. 1993. *Mathematical methods in linguistics.* Dordrecht: Kluwer Academic Publishers.

Passonneau, R. J. and D. J. Litman. 1993. Intention-based segmentation: Human reliability and correlation with linguistic cues. In *Proceedings of the 31st annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 148-155.

Pereira, F. and Y. Schabes. 1992. Inside-outside reestimation from partially bracketed corpora. In *Proceedings of the 30th annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 128-135.

Pustejovsky, J. (ed.) 1993. *Semantics and the lexicon.* Dordrecht: Kluwer Academic Publishers.

Roche, E. and Y. Schabes. 1995. Deterministic part-of-speech tagging with finite-state transducers.” *Computational Linguistics* 21.2.227-253.

Satta, G. 1994. Tree-adjoining grammar parsing and Boolean matrix multiplication. *Computational Linguistics* 20.2.173-191.

Schabes, Y. and S. M. Shieber. 1994. An alternative conception of tree-adjoining derivations. *Computational Linguistics* 20.1.91-124.

Schütze, H. 1993. Part-of-speech induction from scratch. In *Proceedings of the 31st annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 251-258.

Sembok, T. M. T. and C. J. van Rijsbergen. 1990. “SILOL: A simple logical-linguistic document retrieval system. *Information Processing and Management* 26.1.111-134.

Sonderland, S. and W. Lehnert. 1994. Corpus-driven knowledge acquisition for discourse analysis. *Proceedings of the twelfth national conference on artificial intelligence.* Volume 1. Cambridge, MA: The MIT Press. 827-832.

Strzalkowski, T. 1994. Document indexing and retrieval using natural language processing. *RIAO 94 conference proceedings.* 131-143. [Organized by the Center for Advanced Study of Information systems, Inc. and Centre de Hautes Etudes Internationales d'Informatique Documentaire.]

Traum, D. R. and J. F. Allen. 1994. Discourse obligations in dialogue processing. In *Proceedings of the 32nd annual meeting of the Association for Computational Linguistics.* San Francisco: Morgan Kaufmann Publishers. 1-8.

Wilks, Y. 1992. Form and content in semantics. In M. Rosner and R. Johnson (eds.) *Computational linguistics and formal semantics*. Cambridge: Cambridge University Press. 257-281.

Zampolli, A., N. Calzolari and M. Palmer (eds.) 1994. *Current issues in computational linguistics: In honour of Don Walker.* Dordrecht: Kluwer Academic Publishers and Pisa: Giardini editori e stampatori in Pisa.

Zernik, U. (ed.) 1991. *Lexical acquisition: Exploiting on-line resources to build a lexicon.* Hillsdale, NJ: Lawrence Erlbaum Associates.