6 Machine-Readable Dictionaries

ROBERT A. AMSLER
SRI International

INTRODUCTION

A common solution to a variety of natural-language processing problems has occurred independently to a number of researchers. They have recognized that published dictionaries contain a wealth of information that can be used to provide computational solutions to problems of speech generation, spelling correction, automatic hyphenation, content analysis, parsing, machine translation (MT), computer-assisted instruction (CAI), the development of readability measures, and stylistic analysis and can even furnish insights into deeper questions regarding the nature of lexical meaning and other epistemological issues. This common solution entails the use of suitably selected extracts from the machine-readable versions of published dictionaries that have been converted into databases of lexical information.

Explicit knowledge of the tens of thousands of lexical elements of the language and their properties will be required before artificial intelligence (AI) and computational linguistics systems can compute over the entire domain of a natural language. Machine-readable dictionaries (MRDs) thus constitute a research frontier for both empirically based investigations into the nature of language and the creation of advanced natural-language processing systems.

This chapter sketches the history and background of work on MRDs and provides a literature review of current applications and research. It is restricted to documents written in English and largely about English. The chapter does

The preparation of this chapter was supported in parts by Grants IST-8208578, IST-8209346, and IST-8300940 from the National Science Foundation and Grant LM 03611 from the National Library of Medicine.

Annual Review of Information Science and Technology (*ARIST*), Volume 19, 1984
Martha E. Williams, Editor
Published for the American Society for Information Science (ASIS)
by Knowledge Industry Publications, Inc.

not deal explicitly with computational lexicography,[1] which is the use of computers to assist in the creation of dictionaries, except where such work is relevant to an understanding of the resulting MRDs or their applications. The emphasis is thus on events that occur after an MRD comes into existence and efforts directed toward developing machine-readable lexical resources from which one can derive information that is relevant to computation using words.

The chapter is organized into four main sections: History, Applied Research, Commercial Applications, and Theoretical Research in Lexicology. The last three sections may appear to be oddly ordered, but this sequence reflects the chronological development of the field. Work on MRDs began at academic institutions and nonprofit research centers as researchers attempted to apply such dictionaries to problems in their area specialties. This early stage was followed by a transition to commercial exploitation and then to the development of dictionary applications by publishers and software vendors. Finally, with recent progress in AI and computational linguistics and with the availability of sufficiently large and powerful computers a new basis for theoretical work on the lexicon is establishing the foundations of a science of "computational lexicology." The nature of this new science is discussed at some length in the final section.

What is a Machine-Readable Dictionary?

Before describing what an MRD is, it is necessary to indicate what it is not. The term "dictionary" is used in at least two areas of computation with which we have little or no concern in this chapter.

A "data dictionary" is used in database management to refer to a software package by means of which the attributes of a database are stored together with descriptions of their possible values. In effect it is a form of database documentation that is kept in dictionary format. That is not what this chapter is about.

Second, there is a well-established use of the term "dictionary" to refer to what should properly be called a "word list." Although some work on building word lists has led to the development of dictionaries and is thus relevant to the intent of this chapter, there is one use of "dictionary"—meaning "word list" in the computer science literature—that has little to do with the study of natural languages. In particular, sophisticated algorithms for efficient "hash coding," storage, and retrieval of lexical elements from a computer memory are an ongoing product of computer science. Such work addresses both general database-management storage techniques and the use of such lexicons in compilers. Although any MRD involves a large number of lexical elements, such uses of word lists are not really relevant to this chapter, and they are not discussed in great depth here.

The concept of an MRD that is addressed here is quite simple. We are concerned with the digital encoding of symbols in a standard computer character

[1] An excellent review of the computational lexicography literature can be found in KIPFER and need not be repeated here.

code (e.g., ASCII,[2] EBCDIC[3]) from which the information in printed dictionaries can be derived via computer typesetting. Specifically excluded are analog or even digitized images of dictionary pages that can only be redisplayed as a video image. A typical example of an MRD would be a nine-track, 1,600-bpi,[4] $\frac{1}{2}$-inch, magnetic, computer-phototypesetting tape used to produce the actual pages of a published dictionary. A slightly more sophisticated MRD would be the same medium containing database records written specifically to facilitate manipulation of the contents. The most sophisticated MRD would be a database loader stream or an online database itself, which exists on disc in a database management system (DBMS) and allows Boolean retrieval of selected definitions or parts of definitions in a variety of formats. Such a system might also be capable of interactive updates or modifications.

The common element that all these media and representations share is that the information in the dictionary can not only be displayed but can also be computationally manipulated down to the character level. This latter characteristic makes the dictionary an MRD.

HISTORY

SDC Lexicographic Project

The first commonly available MRDs, the *Merriam-Webster Seventh New Collegiate Dictionary* (commonly abbreviated as *W7*) and the *Merriam-Webster New Pocket Dictionary* (commonly abbreviated as *MPD*) were produced under the grant, "An Investigation of the Structure of an English Lexicon," awarded to Olney and Ziff at System Development Corp. (SDC) in 1966–1968 (SYSTEM DEVELOPMENT CORPORATION). The dictionaries were rekeyboarded from the printed books via Flexowriter paper tape since the original compositor's tapes were not available. A format derived from that of M. KAY & ZIEHE was used to translate the several fonts of the published dictionaries into the limited character set of keypunched cards. URDANG (1984) describes similar early efforts to create an MRD at the Random House publishing company, and although the processing steps seem trivial today, they were heroic efforts in the 1960s.

Programs written during the SDC Lexicographic Project provided: 1) parsing capability for portions of the dictionary entries, 2) a concordance index of words occurring in definitions, 3) a suffix routine, 4) a program to generate inflected forms of words, and 5) a program to subset the dictionary entries for further analysis (OLNEY, 1968; REICHERT ET AL.; SCHOENE). The MRDs and these programs were widely distributed to researchers in computational linguistics, information science, and other fields[5] (OLNEY, 1983;

[2] The American Standard Code for Information Interchange of 128 symbols used by most non-IBM computer systems.

[3] The Extended Binary Coded Decimal Interchange Code of 256 symbols developed by IBM for IBM computing equipment.

[4] Bits per inch (per track).

[5] The official list of recipients is maintained by John Olney, System Development Corp., 2500 Colorado Ave., Santa Monica, CA 90406.

OLNEY & PARIS) with the explicit receipt of signed agreements guaranteeing that no additional copies would be made or distributed and that the existing copy would be used only for noncommercial research.

A followup grant from 1970–1972, "Semantic Foundations for Question-Answering," awarded to Olney and Borko by the National Institutes of Health (NIH), began a partial disambiguation of the definitions in the *W7* and *MPD* and continued work on "steps toward developing a metaphysically and epistemologically adequate conceptual scheme within which linguistically motivated semantic components for English could be defined." However, many goals of these projects (OLNEY ET AL., 1967a; OLNEY & RAMSEY) were not realized. Specifically, efforts to produce a set of semantic primitives for the *W7*, to provide a hierarchical organization of the definitions, and to chart the semantic fields of the entire lexicon were only initially explored. The following sections elaborate on the more recent efforts to study these and other topics in these dictionaries.

The SDC Lexicographic Project was a pioneering effort that was strongly limited by the level of computer technology available. It left behind a rich legacy of research reports such as those by GIVON (1967a; 1967b), GRUBER, MACHINA (1967), REVARD (1967; 1968a; 1968b; 1969) SPARCK JONES (1967), and ZIFF. Despite the inability to investigate the lexicon as completely as Olney and others had wanted, the project established several entirely new lines of research involving the use of MRDs in areas such as spelling correction, hyphenation, parsing, speech research, and computational semantics.

The *W7* and *MPD* continue to this date as the most frequently used MRDs. A recent newsletter from OLNEY (1983) indicated that requests for copies of the *W7* and *MPD* MRDs were still appearing at an average rate of one every other month during 1981–1983. Recent recipients have described efforts to apply the *W7* to the problems of student essay analysis, text understanding, parsing definitions, generation of derivative definitions, experimentation in information retrieval, transcription of English to Braille, speech recognition and understanding, research on syllable structure and phonotactics, development of algorithms for spelling correction and hyphenation, readability studies, and the development of prosthetic devices for the verbally handicapped.

Dictionary Archives

A dictionary is compiled by lexicographers who perform essentially a cluster analysis of "citations" (i.e., occurrences of words, in context, recorded on separate slips of paper), which are extracted from texts, sorted into stacks on the basis of their similarity of meaning, and then used as the basis for writing definitions. In the United States, the G.&C. Merriam Co. possesses the largest collection of such citations, recently estimated at 13 million (MISH, 1983a). This collection is truly a national archive, and efforts by G.&C. Merriam to computerize the files have begun (MISH, 1983b).

KRASKE provides an elementary introduction to how dictionaries are created but also gives useful information for anyone interested in a behind-the-scenes look at a dictionary publisher. Although written for young readers, Kraske's book contains much interesting information about commercial lexicography. The photographs alone are invaluable and show a portion of G.&C. Merriam's immense file rooms. NORLING-CHRISTENSEN provides a more contemporary view of how computers are used in lexicography from the perspective of a commercial publisher.

The development of computer-generated concordances naturally led to the realization by publishers that these listings could be used in place of the painstakingly gathered and manually typed file cards containing citations. During 1963–1966, Kučera and Francis at Brown University produced what has become the most famous machine-readable word-frequency list of the English language, the *Standard Corpus of Present-Day Edited American English*, based on a collection of five hundred 2,000-word samples of text extracted from the published American literature of the year 1961 (FRANCIS, 1964; 1965; 1980; FRANCIS & KUČERA; KUČERA & FRANCIS). This million-word corpus, frequently called the "Kučera-Francis," "Brown-Kučera," or simply the "Brown" corpus, was used by the Houghton Mifflin Co. in producing the popular *American Heritage Dictionary of the English Language* (*AHD*) (HOUGHTON MIFFLIN COMPANY, 1982; KUČERA, 1969; MORRIS, 1974). So successful was the *AHD* that Houghton Mifflin embarked on an even more ambitious statistical analysis of a five-million-word corpus, the *American Heritage Intermediate Corpus* (JOHN B. CARROLL; JOHN B. CARROLL ET AL.) of "textual samples from published materials to which students are exposed in grades 3 through 9." From this corpus, the *American Heritage School Dictionary* (HOUGHTON MIFFLIN COMPANY, 1972) was produced. Houghton Mifflin has become the first dictionary publisher to commercially market its computerized lexical databases (HOUGHTON MIFFLIN COMPANY, 1981).

In the years following the introduction of MRDs and the beginnings of commercial computerized lexicography, the *MPD* and *W7* dictionaries were extensively studied and reformatted by researchers several times for new applications. In part, such reformatting reflected the changing role of computers: from punched-card processing through magnetic-tape data processing to online interactive computing. Work on reformatting the original punched-card formats distributed by Olney took place at the University of Texas (AMSLER & WHITE; J. L. PETERSON, 1982a), the IBM T. J. Watson Research Center, and Stanford University (SHERMAN, 1974a; 1974b).

Sherman also began an effort to create a Computer Archive of Language Materials (Project CALM) at Stanford, which tried to assemble, in one machine-readable archive, all existing MRDs as well as other types of machine-readable language materials for scholarly access (SHERMAN, 1979a; 1979b; 1982). He designed a standardized format for dictionary entries based on the Library of Congress (LC) MARC (machine-readable cataloging) format. This format did not gain wide acceptance, however, and eventually Project CALM collapsed from lack of funding. Computer users were switching from batch magnetic-tape data processing of MRDs to online disc access. Disc storage, always at a premium on crowded interactive computer systems, dictated that the dictionary's format reflect this need to be both compact and easily readable. This transition can be seen in the formats developed by Amsler, Slocum, Alberga, and Peterson (AMSLER & WHITE; OLNEY, 1983; J. L. PETERSON, 1982b).

SIMMONS & AMSLER began investigating the *W7* and *MPD* MRDs at the University of Texas in Austin. In the mid-1970s, Amsler became interested in the construction of a complete lexical classification of the language based on the contents of an MRD, and while working at the Linguistics Research Center with White, he supervised the complete semantic disambiguation of all the head nouns and head verbs in the definitions of nouns and verbs from the *MPD* (AMSLER & WHITE). This work was performed by manually disambiguating specially created concordance output containing all the definition texts indexed by their nouns and verbs. The information on nouns and verbs was then assembled into two gigantic data structures, using the LISP programming language to connect the disambiguated head words, to form two taxonomic-like semantic networks. So immense were these resulting networks (27,000 noun concepts and 12,000 verb concepts) that address-space limitations prohibited interactive access to the data. AMSLER (1980; 1981) describes these steps and provides a preliminary survey of the resulting data structures. AMSLER (1984) describes the concept of a "lexical knowledge base" that embodies these networks and other forms of lexically related information.

While the format in which MRDs were stored was changing in the hands of academic and industrial users and the computational analysis of the *MPD* and *W7* continued, the commercial dictionary field started catching up. Longman produced the *Longman Dictionary of Contemporary English*, or *LDOCE* (PROCTER, 1978), and the *Longman Dictionary of English Idioms*, or *LDOEI* (LONG). The machine-readable tapes of these dictionaries have been made available at nominal cost to a number of researchers under strict contractual agreements with Longman regarding their controlled research use. The *LDOCE* tape was created from a database within which the phototypesetting information was embedded rather than the converse. It contains explicit syntactic and semantic codes for its lexicon that show significant promise for computational applications (MICHIELS, 1981; MICHIELS ET AL., 1980a; 1980b; 1980c; 1982; MICHIELS & NOËL, 1982a; 1982b; WALKER & AMSLER).

Computers and Dictionary Production

While preparing this chapter, I surveyed the dictionary publishers to determine how many of them actually had MRDs available. Nearly all of those who responded are indeed producing MRDs as a byproduct of their computer phototypesetting operations. Most publishers are either using these tapes to revise their dictionaries or contemplate doing so shortly. Few publishers have released copies of their tapes for reasons other than business, and equally few have entered into the more advanced stages of MRD use, such as database publishing or database access to their dictionaries.

One reason for this guarded attitude is that a dictionary is a difficult work to protect. Computational processing of dictionary information to extract subsets renders proof of the original source exceedingly hard to establish. In addition, many computer programmers either are oblivious of the copyrighted

nature of MRDs or do not realize that such electronic copying is a violation of the law. Merriam-Webster has prosecuted personal computer software vendors who have attempted to sell *W7* or *MPD* information, some even brazenly advertising their source as "Webster's Dictionary." What apparently is not well understood by computer professionals is that although the original Noah Webster's dictionary as well as most books from the 19th century are indeed now in the public domain and although the term "Webster's" is not a trademark, dictionary publishers have invested tens of millions of dollars in new editions of their dictionaries, and these works are very much covered by the 1976 Copyright Act. Unauthorized electronic reuse of a copyrighted MRD is dangerously close to outright plagiarism. Sale of any derivative products is illegal.

The whole issue of electronic copyright is a serious one, especially in terms of how to protect reference works, but a full discussion of it goes beyond the scope of this chapter. KEPLINGER discusses copyright in the context of information technology applications in general. NEAL & SLOWINSKI provide a reasonable introduction to the legal aspects of the situation, and HOLMES describes a concept of "reasonable use" for machine-readable data. OLSSON discusses the problem with specific reference to lexicography.

Given that most contemporary dictionary publishers in fact do have MRD versions of their published dictionaries available, anyone who wishes to do research in this area can only be advised to contact the dictionary publishers directly and inquire about possible contractual and financial arrangements to secure a copy of any dictionary in which they are interested. However, the most authoritative unabridged dictionaries, such as the *Merriam-Webster Third New International* (*W3*) (GOVE, 1961) and the 1933 *Oxford English Dictionary* (*OED*) (MURRAY ET AL.) are not available in machine-readable form because they were created before computer phototypesetting came into widespread use. Oxford University Press announced recently that they intend to computerize the *OED* (LAMB; HORWELL & WINCHESTER; SEROY).

Although somewhat out of date, the 1977 *Dictionary Buying Guide* (KISTER) is still an excellent survey of the parameters appropriate for evaluating dictionaries. It contains a wealth of information about the types and content of dictionaries that could prove useful for those seeking access to an MRD. ARNOLD (1980-1981) presents an assessment of dictionaries based on Kister's selected set of test words.

APPLIED RESEARCH

Applied research involving MRDs has been undertaken in as many fields as there are disciplines concerned with language and the lexicon. In this area I have tried to focus on work that is specific to MRDs and related word lists rather than on work that is just concerned about the lexicon in general. Thus, instead of dealing with the use of the dictionary in education, I have considered only the small amount of work concerned with the use of MRDs as an educational resource. Likewise, work in sociolinguistics includes only the few studies that examine the dictionary, its definitions, and their sociolinguistic import. When we turn to information science, the focus is even narrower, and only studies that used MRDs or accessed them for use are considered. The section on library science discusses the future of MRDs, the broader context of their relationship to other machine-readable texts, and the potential technology for their access. Word lists are considered as a separate subject since they are generated and used by so many different disciplines that it would not be appropriate to place them under any one classification. Computer science is mentioned in the context of the issues of MRD storage and access. Lexicography appears both in the section discussing applications, as well as later in the section discussing commercial developments, because some dictionaries are produced solely for scholarly research and are not mass-market products.

Education

The role of dictionaries in education was given a big push in 1978 when Texas Instruments marketed its "Speak and Spell" word game (*ELECTRONICS*). This device gave educators hope that new computer microprocessor technology was going to make possible additional advanced English-language training aids. In 1978-1979, the National Institute of Education sponsored two conferences that considered the future and potential of automated dictionaries (G. A. MILLER, 1979a; 1979b). A report was commissioned from Carnegie-Mellon University to assess the state of the art in microprocessor design relative to the requirements for storing and accessing an entire dictionary in a portable computer device similar to "Speak and Spell" (FOX ET AL.). The conclusion of the report was that the price of a computer-accessible portable dictionary of 30,000 entries (averaging 2,800 bits per entry) would drop logarithmically from more than $5,000 in 1982 to less than $500 by 1988.

A. KAY & GOLDBERG, at the Xerox Palo Alto Research Center, had described a similar device much earlier, but the specific use of their "dynamic book" or "Dynabook" for retrieval of dictionary entries was not discussed (see also the section on library science). In an unrelated hardware evaluation, MURRAY & KLINGENSTEIN considered the computer architecture requirements for an electronic book based on a VLSI (very large scale integration) design, although they gave no price estimates.

Today we have seen the birth and rapid success of portable computers (WINER & WINER) that are not much bigger than a book (e.g., Radio Shack TRS-80 Model 100, Gavilan, Epson HX-20, Sharp PC 5000, GRID Systems Compass, TI CC-40, Teleram T-3000) and the appearance of a computer magazine devoted to such portables (*PORTABLE COMPUTER*). Cleverness in compacting lexical information into ever smaller amounts of computer memory (e.g., WILLIAMS & MEYER) and compacting ever larger amounts of computer memory into portable computers make the portable computer dictionary a strong prospect in less than a decade. In fact, I would be surprised if it did not appear within the next five years.

In addition to the development of specific computer hardware for MRDs, MICHIELS (1981) discusses the exploitation of the *LDOCE* for teaching English vocabulary. In this context, an MRD can serve as a source of lexical and grammatical information for a program that generates appropriate examples for students.

Library Science

MRDs would obviously be a great asset if they could be converted into personal electronic books for individual use. However, libraries are the traditional repository for all except the lightest and least expensive dictionaries. It is appropriate therefore to consider how an MRD would fit into the existing structure of library science as an alternative nonprint media.

Traditional librarianship continues to have considerable difficulty in dealing with machine-readable data, especially if the use is intended to be similar to what we currently consider normal for printed books, phonograph records, cassettes, and even videotapes. One recent article that favors providing patrons with direct access to such data nevertheless assumes the traditional view of machine-readable data as consisting of numeric information (ISAACSON). ROWE and BERESFORD & POMERANCE likewise discuss the availability of machine-readable government documents but only for statistical information. This situation is changing rapidly however, with the arrival of machine-readable encyclopedias (P.R. COOK; FLAGG; HARTER & KISTER) and research on software especially designed to access such texts interactively (LENAT ET AL.; WEYER).

Today, however, most publishers still tend to regard the machine-readable texts of their books as virtual trade secrets, and the thought of a library having machine-readable texts of books for computer manipulation or of providing free online access to their contents within the library itself on an in-house computer must seem akin to giving away the printing plates. This situation could change in as little as five to ten years. The technology to manufacture low-cost digital representations of books, such as machine-readable encyclopedias and dictionaries, and to disseminate them via a publication-like system, already exists in the form of digitally encoded videodiscs (*ELECTRONIC LIBRARY; MONITOR*). The computer resources to view such materials are still rather expensive (several thousand dollars per unit), but their prices can be expected to decline by a factor of ten in less than a decade.

The concept of a "library of the future" that purchases videodiscs containing the digital texts of reference books from publishers and provides patrons access to them via interactive color graphic display systems is technologically attainable. Such a system would permit modes of access that are not possible today, allowing one to search through, sort, and computationally process those texts and copy portions onto personal computer diskettes for home use, much the way one can photocopy portions of existing reference books at a library. Visionaries would claim that even this is too modest a goal (NELSON, 1981a; 1981b). It would offer an attractive alternative to present remote access to online databases, which are too expensive for free public library use. Making electronic copies of electronic publications in libraries may also be more likely if telecommunications costs do not drop (and there is little reason that they should in the near future). Database vendors may find it impossible to extend customer service to tens of millions of people every day, using existing phone, satellite, and computer systems and still remain competitive with individualized videodisc access on personal machines.

Sociolinguistics

Dictionaries have a unique role in our formalization of knowledge because they purport to tell us what the words of our language mean. In doing so, they can also be conscious or unconscious vehicles of our cultural prejudices. DAVIS discovered twice as many negative connotations for "black" as for "white" in an unspecified edition of Rogets *Thesaurus of the English Language*, and DUNCAN argued that the *Random House Dictionary of the English Language* (*RHD*) (STEIN) showed racist tendencies. GERSHUNY (1973; 1974) used example sentences from definitions in the *RHD* to examine the incidence of sex-role stereotypes and found a considerable bias against women. During the 1970s, dictionary publishers made efforts to eliminate such biases. KRASKE notes that the *American Heritage School Dictionary* (HOUGHTON MIFFLIN COMPANY, 1972) was the first dictionary to avoid sexism. The cover of the 1978 pocket-sized *Random House Dictionary* (STEIN & SU) proclaims that it is also "the first nonsexist dictionary in history," and inside it notes that, "Throughout this book we have tried to free our definitions of sexism, racism, and other prejudices. We believe we have done so more thoroughly than ever before in lexicographic history."

In his book, *1984*, George Orwell coined the term Newspeak for a language controlled by the state to promote the concepts it wanted. Control of the meanings of words and the entries in the official Newspeak dictionary was one mechanism for this control. Given our society's position—i.e., on the verge of complete computerization of the lexicon—it is doubly relevant for us today to be concerned that we do not bias our electronic dictionaries of the future. We can only hope that the ability to scan the contents of an MRD and detect stereotypical uses of the language will lead to their elimination and not to their systematic manipulation.

Information Science

Automated indexing. Indexing by computer has usually been considered to depend solely on the statistical properties of the words and phrases in a text. The MRD provides a new option for automated indexing—namely, that one may be able to computationally "understand" the meanings of the phrases and words in a text and generate appropriate index terms that may not be in the text. BELL & JONES suggest that an AI approach to indexing is needed. A comprehensive review of automated indexing was provided as part of the chapter by TRAVIS & FIDEL.

Content analysis. Computerized content analysis has always depended on large lexicons for its success. WALKER & AMSLER have shown how a program can use subject codes from the machine-readable tape of the *LDOCE* to perform content analysis of text. They extracted some 17,958 subject codes from this MRD and have tested the method on an entire day's worth of stories from the New York Times News Service (NYTNS). The most promising aspect of the result is that the methodology could be applied to foreign language material as well since it is lexically based rather than dependent on syntactic analysis.

Thesaurus construction. The thesaurus is a proven tool in information science (TOWNLEY & GEE). It also is an accepted alternative display of some of the information in a conventional dictionary. AMSLER & WHITE and AMSLER (1980) describe a technique for building a thesaurus-like structure from the definitions in an ordinary MRD (see also MICHIELS & NÖEL, 1982a). This technique permits one to check the definitions to determine that they indeed express the same view of meanings as the thesaurus. It also offers the possibility of inverting the process and building a dictionary from a thesaurus. The recently published new edition of *Roget's Thesaurus of English Words and Phrases* (LLOYD) used a computer to check all cross-references and to generate lists of cross-references at the end of each section. *The Historical Thesaurus of English* (COLLIER & C. J. KAY; C. J. KAY) is an attempt to organize hundreds of thousands of citations taken from the *OED* into a semantic thesaurus of English. Efforts to computerize this work are in progress. The ties between dictionary and thesaurus thus appear to be growing ever stronger as the computer provides a convenient means of realizing their correspondences while preserving their distinctly different formats.

Machine Translation

Machine translation (MT) of languages may be the hardest task imaginable for computational linguistics. Outside the United States, the need for translation is recognized as critical by many countries. Consequently, large collections of lexical data have been amassed for this purpose. While a complete MRD would include definitions of the terms it contains, there are many multilingual books called dictionaries that list only word equivalents in two or more languages, possibly with parts of speech. These and linguistically more complete dictionaries are vital to MT efforts today (KNOWLES). Such databases are known as terminology banks, but in their most complete instantiations they come very close to being complete dictionaries. TUCKER & NIRENBURG discusses the structure of MRDs for MT systems.

A terminology bank is typically a very large (tens or hundreds of thousands of entries) database of lexical information (ARTHERN). Terminology banks in Japan are discussed by NAGAO ET AL. and NOMURA, who also mention two machine-readable Japanese translation dictionaries. MCNAUGHT (1981; 1982) describes the British Linguistic Data Bank. LIPPMANN examines interactive access to a terminology bank. MCNAUGHT (1983) presents a technique for the automatic generation of new entries for inclusion in an existing terminology bank. GOETSCHLALCKX & ROLLING consider several lexicographic perspectives on terminology banks in Europe. CIGNONI ET AL. present the results of a European Science Foundation survey of lexicographic projects, including terminology banks. ALLEN (1977b; 1980; 1981a; 1981b) describes the development of the language bank concept in Sweden.

Word Lists

Word lists derived from machine-readable dictionaries. MRDs are often sought as a source for a large word list to be used in a spelling-correction or hyphenation program, for speech recognition or generation, or for use in designing or solving word games, puzzles, and contests. Here we encounter the need for careful attention to what a dictionary offers. Dictionaries do not list all forms of every word. Inflected forms of regular verbs or normal plurals of nouns do not appear as dictionary main entries.[6] Adjectives and adverbs derived from nouns and verbs may not have separate entries in a dictionary. Variants of words or irregular forms of a word may appear in a printed dictionary only as suffixes to be changed or added to listed uninflected forms. The algorithm for adding such forms may depend on phonologic or other data that are not easily computed. Since the dictionary does not explicitly contain a list of all the forms of a word, it is important for anyone who turns to an MRD for lexical information to be aware of what they will obtain.

The word list from a slightly damaged version of the *Merriam-Webster Second International Dictionary*[7] (published in 1934) has existed online at Carnegie-Mellon University for several years now. Containing some 232,404 words, it is probably the largest machine-readable word list commonly available to researchers. The *Normal and Reverse English Word List* (AUGUSTUS F. BROWN) from the University of Pennsylvania and *The English Word Speculum* (DOLBY & RESNIKOFF; DOLBY ET AL.) produced in the 1960s at the Lockheed Missiles & Space Co. (now DIALOG Information Services, Inc.) were even larger, but no known machine-readable copy has surfaced in many years.

Word lists derived from machine-readable text. The alternative to an MRD as the source of a word list is a frequency list of the words derived from a corpus of the language. The most popular list has been the *Standard Corpus of Edited Present-Day American English* produced two decades ago at Brown University (FRANCIS, 1965; FRANCIS & KUČERA; KUČERA & FRANCIS). Once again, however, the word list obtained does not include a list of all the word forms in the language. The Brown Corpus consists of a one-million word corpus of published text fragments taken from a statistically balanced sampling of American publications in 1961. This number of words often sounds more than adequate to a novice who is starting on an English-language computation, but it is really quite small, and the Brown Corpus contains only 50,406 *different* spelling forms (KUČERA, 1969), no more than a pocket dictionary. Because it is based on actual text, it also faithfully replicates the spelling aberrations of that original text. Words are often present *only* in their commonest inflected forms—for example, "abandon, abandoning, and abandoned" occur in the Brown Corpus, but "abondons" does not; "asterisks" occurs, but "asterisk" does not.

WALKER & AMSLER at SRI International have been building a large word list based on data from NYTNS, and at last count, the lexicon of some eight million words of text had been analyzed as part of this project. This word list contains some interesting information that was not previously recorded by other researchers. Amsler developed a formal definition of a text

[6] They *may* occur within the text of the dictionary's definitions, but in this regard a dictionary is being considered as though it were just another machine-readable text.

[7] The current version is the Third International (1961).

symbol ("sequence of non-blank characters") and text word ("a symbol with preceding and trailing punctuation removed") and applied these throughout the frequency analysis. The corpus also preserves information on the upper/lower case of word occurrences and on their surrounding punctuation. This information was lost in both the original publications of word lists from the Brown Corpus (KUČERA & FRANCIS) and *The American Heritage Word Frequency Book* (JOHN B. CARROLL ET AL.). The newer "tagged" Brown Corpus frequency list (FRANCIS & KUČERA), so named because it identifies the syntactic classes of the words, also provides the correct upper/lower case distinctions and some information on accompanying punctuation. The processing of textual information by computer makes such information useful because in a digital computer many algorithms depend on the precise distinction of symbols expressed in ASCII, where upper/lower case alphabetics and punctuation characteristics of the language may be required knowledge for text formatters, word processors, parsers, or other computational linguistic programs.

Neither the Brown Corpus nor the NYTNS corpus nor any word list readily available contains all the word forms of the English language. To see why this is so, consider what happens to a noun, such as ice cream, or a proper noun, such as New York City, when it is found in a text. As analyzed, these open compounds would be segmented into separate words—i.e., "ice" and "cream," and "New," "York," and "City." Statistically they would contribute to the frequencies of each separate word of which they are composed, and no record of their existence as a single lexeme would exist in derived word lists.

This situation creates problems for the computer-system developer. Although human beings are aware of these elements as lexical units, computer programs will need an explicit enumeration of them for correct processing. MRDs have some of the ordinary compounds, but often systematically exclude many classes of proper nouns. Some special-purpose dictionaries list specific types of proper noun compounds, such as geographic terms, biographical names, corporation names, and trademarks, but no one dictionary lists the commonest of these in all categories, as would be encountered in general newspaper text. WALKER & AMSLER compared the word types in their NYTNS corpus with the main entries of the *W7* dictionary and found that only 23% of these word types were common to both word lists; 36% of the word types were only in the *W7*, and 41% were only in the NYTNS corpus. Of the word types appearing in the NYTNS corpus that didn't appear in the *W7*, approximately one-fourth were inflected forms, another one-fourth were proper nouns, one-sixth were hyphenated forms, one-twelfth were misspellings, and the remainder could not be categorized without checking their contexts. Thus, somewhat sadly, one must admit that there does not exist a word list of English that enumerates the true lexical units of the language based on their representative frequencies.

This discussion points out that the term "word" is highly ambiguous. One researcher may use it simply to refer to sequences of characters separated by blanks or punctuation (still begging the question as to what and when punctuation can be inside such "words"); another may use it to refer to so-called lemmatized words, which are the uninflected dictionary entries of the original spelling forms found in text, (e.g., "girl" is the lemmatized form of "girls" and "girl's"); yet still another researcher may mean the linguistic concept of "word" in which "ice cream" and "New York City" are "words." Thus the apparently simple request for a "word list" of English cannot be readily fulfilled without clarification of what is meant by a "word" and what aspect of English the "list" is intended to represent.

Word list bibliographies. ALAN BROWN cataloged some 172 documents dealing with the properties of "verbal material (letters, consonant-vowel-consonant combinations, words, etc.)" in an effort to produce a "reference manual for researchers in all areas involving verbal processes, including learning, memory, linguistics, psycholinguistics, mental assessment, etc." It includes sources for information about sounds, punctuation, letters, numbers, *n*-grams (i.e., digrams, trigrams, etc.), syllables, *n*-syllable combinations, strings, words by concepts, parts of speech and other special classes such as homographs and homophones. VENEZKY (1979) provides a survey of "orthographic regularities in English words," and WAIBEL reviews large lexicons specifically useful in speech research.

Computer Science

Dictionary database design. FREDERICKSEN outlines a database structure for dictionary definitions that permits the user convenient access to all words that share the same sense, all senses of a given word, or all phrases in which a given word occurs. NIX and FRAENKEL & MOR discuss compact storage of word forms, which of course is related to hash coding techniques in general. CERCONE ET AL. discuss near-optimal hash functions in the context of natural-language work.

Lexicography

Lexical data banks. When a database of lexical information is compiled with the aid of a computer, it is often referred to as a data bank (BAHR) rather than a terminology bank. The primary distinction between the two, apart from the intended applications of the data, is in the contents of the database. Terminology banks often do not include "ordinary language" but are more closely related to "sublanguages" (KITTREDGE; KITTREDGE & LEHRBERGER)—i.e., areas in which there is specialized terminology (which, in translation, will often not be understood by the translator without access to a special dictionary). Additionally, compilers of data banks will more often collect usage examples from texts to assist in understanding the justification for the entries in the database. VENEZKY (1969; 1973a; 1973b) and VENEZKY ET AL. (1976a; 1976b; 1977a; 1977b) examine many aspects of providing an integrated computer system for building a lexical data or terminology bank. A number of substantial foreign lexical data banks are being developed (see CIGNONI ET AL.).

PAIKEDAY (1983) notes rightly that the existence of online computer access to the full text of many newspapers and journals has forever changed the nature of lexicography. Now dictionary publishers need not invest directly in the compilation or maintenance of corpora of English. They need only connect through a commercial online service to the appropriate databases and ask to see all occurrences of a given term in context. Authentication of the first date of appearance of a new sense or word can be documented whether or not the lexicographers spot it when it enters the language. Online access is not subject to the faults of human concentration, which can miss a word in reading a text. RAOUL SMITH (1972) suggests that interaction with informants should be built into any updating system for a lexical database. Anthropological linguists have a well-developed role for human informants in the gathering of lexical information. WERNER proposes a synthetic (i.e., computational) informant to test language understanding.

Medical Lexicology

Medical terminology has interested lexicographers and lexicologists for many years. This fascination is partly due to the emphasis that the medical sciences place on morphology as the basis for understanding the meaning of medical terminology. PACAK, PACAK ET AL. (1973; 1976; 1980), PRATT ET AL., and PRATT & PACAK (1969a; 1969b) have worked extensively on the morphology of medical terminology at NIH.

COMMERCIAL APPLICATIONS

Spelling Correction

Correction of spelling errors is one of the most common applications for large word lists. These lists can be used either in the design and testing of spelling-correction algorithms or as the basis of an algorithm that uses word-list lookup during its operation.

There are various techniques possible for spelling correction. One method is to approach the problem of detecting spelling errors probabilistically, through the likelihood of bigrams (two-letter sequences) or trigrams (three-letter sequences) occurring in the language. One can readily build a bigram or trigram transition table based on a word list or running text. The words encountered in a text are then analyzed as though they were composed of a series of bigrams or trigrams, and a probability that the word is correctly spelled is determined. If the probability is below a certain threshold, the word is submitted to the user as unlikely to be correctly spelled, and a correction is accepted if appropriate. This technique is used in the UNIX operating system spelling corrector, TYPO.

Hardware produced by MICROMARK (CHICHELLI & IOBST) performs spelling correction using "hash coding with allowable errors," a technique in which a hash code is calculated for each word and checked against an array to see if this hash code could correspond to a correctly spelled word. Using a 30,000-word lexicon MICROMARK claims a chance of only 1 in 4,000 of mistaking an incorrectly spelled word for a correctly spelled one.

The final technique often used is to store a complete lexicon of correctly spelled (and even incorrectly spelled) terms and perform a lookup of each term in this lexicon, allowing for morphological variants such as plurals, tenses of verbs, affixes such as non- and -ness. GORIN developed one of the first such programs for a Digital Equipment Corp. DEC-10 computer at

Stanford University. Since then, improved versions of this system have been installed on virtually every DEC-10 and DEC-20 computer system. An added advantage of this type of spelling corrector is that additional uses for the lexicon may be found, such as in automatic hyphenation or text compression.

Combinations of these techniques are clearly possible. J. L. PETERSON (1980a; 1980b) has examined spelling correction as a problem in program methodology. AMSLER (1982) has noted deficiencies in the one-letter error-correction methodology, which ignores the near impossibility of predicting one-letter errors in three- or four-letter words while unreasonably assuming that looking for one-letter errors in very long words is the best reasonable effort. As a word's length increases, the likelihood of the word's being recognizable as a misspelling of another, even when two or three letters are wrong, is considerably higher.

Much work on spelling correction has been undertaken by researchers working in connection with large bibliographic databases where the problems of finding documents by titles, keywords, or author's names are multiplied by the lack of consistent spelling. LERNER ET AL. describe the use of spelling-correction techniques in commercial electronic publishing. POLLOCK & ZAMORA and ZAMORA have worked on the problem in the context of the databases of the American Chemical Society (ACS).

The Institute for Scientific Information (ISI) maintains a number of databases in addition to its bibliographic files. One of these, the Unique Word Dictionary (GARFIELD, 1981a), is "a machine-readable master list of correctly spelled terms which have been certified as 'real' words." In 1980, the Unique Word Dictionary contained 450,000 different terms taken from the titles of articles in journals. The growth rate of the lexicon is slowing, from an increase of 130,000 new terms over the years 1977–1979 to just over 20,000 terms in the two years 1979 and 1980. Nevertheless, ISI estimates that they capture some 500 to 1,000 newly coined terms every year. The important feature to note is that a fundamentally lexicographic function is being carried out by a database vendor as a by-product of its bibliographic indexing work. This project suggests that the future of machine-readable dictionaries may lie more with those companies that process the most information than with those that continue to follow standard lexicographic traditions manually. PAIKEDAY (1983) called attention to this state of affairs, suggesting that the handwriting is on the wall for dictionary publishers.

Automatic Hyphenation

Closely allied with spelling correction is automatic hyphenation of words. Typesetters have long recognized the need for hyphenation in typesetting justified text. Aesthetic standards dictate that even when variable-width spacing between characters and symbols is available, it will occasionally be necessary to break longer words into two segments across a line or page boundary. As computer typesetting has advanced, this need has been perceived as more important and deserving of scholarly attention.

LIANG presents a historical introduction to the problems of automatic hyphenation. In particular, he describes the computer hyphenation algorithm and methodology used in the TeX typesetting program (KNUTH, 1979; 1984).

Liang used the *MPD* and a published hyphenation dictionary to develop a highly efficient algorithm using a special machine-readable file of hyphenation patterns. It permits TeX to avoid incorrect hyphenation of words without explicitly requiring them all to be stored in the computer.

Grammar and Style Correction

While spelling-correction techniques may find random errors or those caused by mishaps during keyboarding, they do not work at all on some errors of human judgment, such as the choice of an incorrect homophone (e.g., their vs. there). Human errors in text may thus also be grammatical or stylistic.

Two efforts to address these types of grammatical and stylistic errors are the Writer's Workbench developed at Bell Labs (CHERRY, 1978; 1981; CHERRY & MACDONALD; CHERRY & VESTERMAN; CHERRY & WALLACE; MACDONALD ET AL., 1980; 1982; MACDONALD & POLLER) and EPISTLE (Evaluation, Preparation, and Interpretation System for Text and Language Entities), under development at the IBM T. J. Watson Research Center in Yorktown Heights, NY (HEIDORN ET AL.; L. A. MILLER; L. A. MILLER ET AL.; WAY). R. SMITH (1981) provides an assessment of these two computerized aids to writing.

The Writer's Workbench consists of a series of programs that run under the UNIX operating system and perform relatively simple pattern-matching and text-processing operations on a text. PARTS determines the parts of speech of the words in the text. It is similar to a number of other programs that do this without resorting to explicit parsing of the text (CHAPIN & NORTON; EARL, 1967; HAYS, 1967; M. KAY & MARTINS; KLEIN & SIMMONS; LAMB ET AL.; PACAK ET AL., 1976; WINOGRAD, 1971). STYLE provides measures of readability based on text statistics, such as sentence length and type (simple, complex, compound, compound-complex), word usage, and sentence openers. PROSE compares the statistics that are output by STYLE with those of "good" documents and suggests improvements. REWRITE capitalizes poor prose and outputs the document with the questionable text thus highlighted. TOPIC locates frequent noun phrases for potential use as keywords or index entries. SPLITINF finds split infinitives. DICTION prints sentences in the text that contain undesirable phrases (e.g., slang, verbose writing, cliches), and SUGGEST offers alternatives to these offending expressions via interactive access to a thesaurus. DOUBLE looks for occurrences of two identical words in sequence, as might occur when the writer repeats a preposition or article at the end of one line and at the beginning of the next. PUNCT checks for some classes of punctuation errors.

EPISTLE, which is still in the applied research stage, follows a more sophisticated approach. Using computational linguistic techniques it attempts to parse the writer's text and actually determine from the parse tree it constructs not only what the writer is doing, but what may have been done incorrectly. EPISTLE thus requires a parsing system that can accept ill-formed input and nevertheless reach some successful interpretation. JENSEN & HEIDORN thus describe the EPISTLE parsing algorithm as a "fitted parse." EPISTLE currently deals with grammatical errors, such as a lack of number agreement between subject and verb (i.e., singular vs. plural), and with stylistic errors, such as the use of overly complex sentences.

Recreational Applications of Dictionaries

Word lists are often sought for use in the design or solution of word games. Home computer enthusiasts now have a variety of word-based games available for their amusement and possible education.

Crossword puzzles. Crossword-puzzle generation may seem to be an unlikely topic for scholarly consideration, but in fact game-playing programs have a respectable background in AI research. The algorithmic solution to the generation of crossword puzzles is nontrivial and has been investigated by a number of researchers. COX, FEGER, MAZLACK (1972; 1973a; 1973b; 1976), P. D. SMITH, P. D. SMITH & STEEN, and WILLIAMS & WOODHEAD have worked on the problem.

Word games. The growth of personal computers and their recreational use has also contributed to new word-game software (SANDBERG-DIMENT; STAPLES), software books (MAU), and products such as a portable Scrabble player (SUGARMAN; *WALL STREET JOURNAL*).

THEORETICAL RESEARCH IN LEXICOLOGY

Two primary areas of current research are considered here, corresponding to two goals of those who use computers to work with the lexicon. These goals are: 1) to gain an understanding of the lexicon itself, using computers as aids to lexical analysis and 2) to build programs that exhibit human-level behavior when dealing with lexical problems such as disambiguation or understanding the meaning of compound nouns. Properly speaking the first area might be termed computational lexicology and the second artificial intelligence, but both share many common goals and are interdependent.

Computational Lexicology

A number of researchers have approached the problem of computational lexicology from the perspective of attempting to deal with either the entire lexicon or a considerable portion of it at one time. The original compilers of unabridged dictionaries certainly would have understood such an ambitious goal. The classic 1852 *Thesaurus of English Words and Phrases* by ROGET is perhaps the most widely known effort to characterize the entire lexicon semantically. SEDELOW prepared a machine-readable version of the thesaurus to make it available for computational use and study, but it has not been widely distributed or used.

Synthetic languages. In the 20th century, some additional efforts to produce major lexical statements for the whole language have been produced as well. OGDEN (1934), OGDEN ET AL., and RICHARDS promoted a scaled-down version of the English language in their system of "Basic English." Basic English claims to provide all the expressive power of ordinary English in a vocabulary of only 850 "basic words" with only 18 verbs. The actual number of words in the Basic English vocabulary is somewhat higher, due to the use of hyphenated forms that are not counted as separate words, but still only about 1,000 word forms are involved. Entire texts were translated into Basic English (see RICHARDS), and a Basic English dictionary (OGDEN, 1942) exists that offers Basic English definitions (in the restricted vocabulary) for the words of ordinary English.

Basic English might seem to be the answer to a computational linguist's dream—a small vocabulary with the expressive power of ordinary English—but it appears to be a step in the wrong direction. One difficulty is that although Basic English words are few in number, they have *more* meanings on the average than ordinary English words. Thus, a vital semantic trap underlies the use of Basic English. Basic English also does not really solve any major computational linguistic problems. As computational linguists will readily agree, the real problem is not the many different nouns and verbs in ordinary English but the difficulty of mastering the first few hundred most frequent words including the prepositions, pronouns, and those very difficult major verbs—all of which are in Basic English with no reduction in complexity.

The task of creating synthetic languages will probably always attract some people, and variations on this theme have appeared periodically starting soon after the turn of the century. Esperanto (BAKER; CLARK; CONNOR). Loglan (J. C. BROWN), and other such languages are relatively useless for computational linguistic purposes because they do not really solve any problems of understanding and representation; they offer only simplifications in terms of the number of word forms used, usually at the expense of added complexity in semantics. Also, because the major problem is not to generate English but to understand English that has already been produced, these languages are not very useful to the computational linguist or the computational lexicologist.

Semantic features. Somewhat removed from these attempts to replace the English language are efforts to assign to all the concepts in English some limited set of features or properties that can explain how these concepts express meaning. OSGOOD ET AL. and SIDNER & OSGOOD describe a meaning representation called the "semantic differential," which comes very close to capturing the essence of connotative word meaning along three axes closely tied to the English adjective pairs of good-bad, active-passive, and strong-weak. Other dimensions were found for the semantic differential but were not validated. AMSLER (1969) explored the use of the semantic differential in a language-understanding system, and while many subtleties were evidenced (e.g., associating colors with nouns, such as "white" with "winter"), denotative meaning was totally lost. HEISE presents a set of standard scores for 1,000 words on the semantic differential, and although it does not seem that these scores will assist in an understanding of the denotative meaning of language, they may prove crucial to representing the connotative impact of language for computational purposes. STONE ET AL. used semantic differential scores in the lexicon of the General Inquirer, a much publicized content-analysis package introduced in the late 1960s, as part of the basis of that program's capability of finding relationships between concepts mentioned in texts.

LAFFAL (1973) created a set of 118 conceptual categories to characterize all the words in a 23,500-entry lexicon. By using these codes in pairs, it is possible to represent ordinary English in a small set of primitive meanings. Laffal's original purpose was to use these codes to analyze psychiatric transcripts, but they seem to have more general applicability to other areas of content analysis, such as in the analysis of political or historical documents.

Restricted vocabulary. The *LDOCE* (PROCTER, 1978) used only about 2,000 words in its "defining vocabulary," i.e., the set of words used in definitions of main entries in the dictionary. This was possible because of the complete computerization of the dictionary's text prior to its publication. Since *LDOCE* is an "advanced learner's" dictionary, intended to be used by those who are studying English as a foreign language, this controlled defining vocabulary was helpful in minimizing the number of concepts with which a beginner must be familiar. For computational linguistic purposes the impact of this controlled defining vocabulary is not as well understood. MICHIELS & NOËL (1982a) reflect on the effect it has had on the taxonomic structure of definitions compared with that of a conventional dictionary such as the *MPD*.

Semantic primitives. There are two components of efforts to simplify or explain the lexicon in terms of primitives. One is the set of semantic primitives being defined, the other is the structure of the network that relates the concepts of the language to each other. The lexicographic tradition is to explain the lexicon wholly within the lexicon itself. Thus, Roget's *Thesaurus* uses English concepts as the basis of its classification of the words of the lexicon. Basic English is another effort to describe English in terms of existing concepts.

Within linguistics, psychology, and AI, efforts to characterize the language have often resulted in the introduction of metaconcepts or metalinguistic semantic primitives. The semantic differential, while having axes that are related to adjective meanings, in fact is represented by a spatial concept of meaning derived from factor analysis and not truly realizable as ordinary English. LAFFAL (1973) uses primitive concepts that are approximations to English concepts. KATZ & FODOR postulate a system of features to describe meanings, in effect posing the task of constructing hundreds or thousands of polarized features to describe all of English lexical meaning. SCHANK presents 14 primitive events as the basis for representing all English verbs. WILKS describes semantic primitives and discuss their role in AI. BARR & FEIGENBAUM (1981a) describe further details of the history of semantic primitives in AI.

These two techniques are not necessarily at odds with one another. Several researchers have proposed that the lexicographic tradition of defining the language within itself, although it necessitates circularity in definitions (AMSLER, 1980; CALZOLARI, 1977; SPARCK JONES, 1967), does not preclude the existence of primitives. A circular set of definitions identifies a semantic primitive, and giving this primitive a synthetic name is merely a convenience for reference.

If we turn from the effort to characterize the concepts of the language in terms of a newly coined set of primitives and look at the structural properties of the concepts of the lexicon, we find a number of studies displaying sizeable portions of a lexicon as various taxonomic and taxonomic-like structures.

NIDA (1975a; 1975b) offers a taxonomic characterization of the major verb and noun classes of English based on a multilingual analysis of meanings. BERLIN ET AL. and HUNN describe sizeable taxonomies of Tzeltal plant and animal names, respectively, illustrating the ethnosemantic technique of taxonomic enumeration as a means of studying a language. BALLMER & BRENNENSTUHL provide not only a detailed analysis of "speech act" verbs (see SEARLE) but discuss the philosophical and methodological justification for the use of enumerative lexical analysis.

Computational-semantic epistemology. The original motivation for working on MRDs included interest in the dictionary as the basis for an empirically based epistemology. FRAWLEY restates this sentiment in proposing the dictionary as a means of studying the philosophy of science. SEDELOW & SEDELOW (1978; 1979b) proposed a concept of "formalized historiography," which Frawley renames "computational-semantic epistemology."

The basic premise is that a dictionary contains all the concepts needed to describe the world, together with definitions that relate statements of their meanings to other words. By pruning away the terms shown to be definable in terms of ever more primitive concepts and determining the naturally occurring "defining formulae" used to create definitions (OLNEY ET AL., 1967b; OLNEY & RAMSEY), one can arrive at an empirically based description of a minimal set of concepts necessary to define the language. LITKOWSKI (1975; 1976; 1978) has used directed graphs to identify and study the primitive verb concepts of dictionary definitions in the *W3*.

Semantic relations between concepts. The issue of semantic primitives, however, is only one of the questions being explored by computational lexicologists. In addition to investigating what words mean as concepts, it is necessary to explore what primitive semantic relationships relate these concepts to each other. Word associations show an uncanny similarity among language users, and it is quite natural to ask what forms the basic set of relations among concepts. EVENS ET AL. (1980) establish solid evidence for super/subordinate and part/whole relationships. BRACHMAN continues the discussion of the nature of the super/subordinate relationship, often called the "ISA" relationship (also called "IS-A") in AI (see next section). Case grammar offers a series of relationships between a verb and its sentential arguments (FILLMORE; NILSEN). Morphology provides characteristic relationships between some verbs and many of these case arguments—e.g., "to pay" is accompanied by the "payer" and "payee." Other case arguments are represented by specific names that are syntactically related to the verb—e.g., "pilot" is both the verb and the name of its agent. MEL ČUK (1981) describes the "meaning-text model" and its underlying assumptions. Through the use of lexical functions such as *Magn* (meaning "very"—an intensifier), *Anti* (meaning antonym), *Son* (meaning "typical sound"), etc., mappings can proceed from existing words into new word forms that express the effect of the lexical function. Thus, *Magn*(belief) = staunch; *Anti*(victory) = defeat; *Son*(cow) = to low, to moo.

Bibliographic databases. The linguistics literature is filled with detailed analyses of individual words or sets of words. However, it is often difficult to locate articles by those authors that have discussed the syntax and semantics of these words. In an effort to make this information more easily accessible, DILLER & OLNEY and BYE ET AL. created SOLAR (Semantically-Oriented Lexical Archive), which describes as completely as possible the contents of lexical-semantic publications.

Use of the Lexicon in Artificial Intelligence Systems

Paralleling the growth of precise descriptions of the lexicon in computational lexicology has been the effort to use the lexicon computationally in programs that perform human-level understanding and reasoning tasks. Although natural-language processing systems have long used some form of taxonomic information, only recently have there been investigations of the potential utility of very large taxonomic structures and the processing techniques needed to handle them. Large taxonomies are important knowledge-base components of expert systems such as MYCIN (SHORTLIFFE) and PROSPECTOR (DUDA ET AL.).

FAHLMAN (1977; 1979) and FAHLMAN ET AL. have discussed the potential viability of specific computer hardware for accessing very large taxonomies. Fahlman points out that existing problem-solving systems based on deductive search are extremely slow, even with the use of complex search-optimizing algorithms. This is of concern not only because of the necessity of writing new search-optimizing algorithms for each new knowledge domain, but because this type of search is trivially easy for humans. Thus, something is wrong with our existing computer hardware that makes it prohibitively expensive to search large taxonomic data structures computationally. This in turn results in programming efforts that avoid the use of very large taxonomies as an integral component of AI processing systems. By designing hardware that makes exhaustive taxonomic search efficient, new types of AI software using such data could be developed.

Word disambiguation. Although examples such as the classic sentence of BAR-HILLEL, "The pig is in the pen," were deemed to be beyond the capabilities of computers to understand in 1964, today the disambiguation of "pen" as "an enclosure for animals" is merely a matter of methodological debate in AI rather than a feasibility counterexample. An active area of research concerns how information contained in such facts as: "pen" is a "writing instrument" or an "enclosure," should be represented in networks of ISA relationships forming "inheritance" or "type" hierarchies (BRACHMAN; SOWA).

Several computational mechanisms for performing word disambiguation have been proposed. In one mechanism we can represent experiences in frames and scripts (see BARR & FEIGENBAUM, 1981b) and then cleverly probe our way through such information to deduce the correct interpretation of the word in context (GRANGER, 1977; 1982). This strategy has a major liability in that, apart from a few examples constructed to illustrate the techniques involved, there is no ready source of such remembered experiences in a form suitable for entry into a computer.

An alternative to this clever search methodology is a technique dating from QUILLIAN (1968) called "marker passing." Marker passing originated in the context of dictionary definitions, the original subject matter of Quillian's research, but it has been resurrected today in much more carefully elaborated designs (CHARNIAK; FAHLMAN, 1977 and 1979; HIRST; HIRST & CHARNIAK) and with the suggestion that parallel computer hardware be specifically created to render these designs more efficient than existing search techniques on conventional computer systems (FAHLMAN ET AL.).

Basically, in marker passing, a trail of "markers" is left at the nodes of a semantic network of concepts such that one can determine whether one has reached some node that has been previously reached via another route. Markers are disseminated to the nodes in a network according to the connections between these nodes and the conditions on traversing these paths. FAHLMAN (1977; 1979) proposed an elaborate marker-passing mechanism for his system and accompanying hardware designs of enormous numbers of microprocessor chips to make such designs extremely efficient. These designs are still several years from actual experimentation, but they could become a veritable sixth-generation computer system if they are successful. Because the fifth-generation is already dedicated to systems that will execute AI software efficiently, I use the term "sixth generation" to refer to systems that will use AI techniques in their hardware design.

HAYES (1975; 1977) and AMSLER (1980) have addressed some of the processing algorithms that can be applied to a very large taxonomic lexicon to perform automatic disambiguation. However, many methodological issues remain to be resolved. One is "default reasoning," which, in the case of an inheritance hierarchy, asks how to achieve a network of connections between concepts so that the first path completed by a program that is exploring this network will correctly reflect our perception of the world (ETHERINGTON & REITER; RICH).

Noun compounds. One of the first applications of Fahlman's NETL system was to the task of understanding nominal compounds (MCDONALD, 1981). This task may be ideally suited to the use of MRDs since linguistic descriptions of nominal compounding and word formation involve hundreds or even thousands of special cases that depend on particular idiosyncratic properties of words. For example, "snowman," meaning "a likeness of a man made of snow," bears no relationship to "iceman," meaning "a man who delivers ice." It is hard to escape the conclusion that social chance determined how these two compounds were defined, and thus the need for explicit enumeration of their meanings becomes essential. The classic linguistic treatments of lexical compounds as in LEES (1960; 1970) and in MARCHAND, have led to a recent interest in the study of the syntax of words (BAUER, 1983; SELKIRK). JOHN M. CARROLL (1981b) has explored the syntax of a related problem in understanding proper-noun phrases.

Computationally, QUILLIAN (1969) was among the first to demonstrate elements of compound-word understanding in his Teachable Language Comprehender. RUSSELL, RHYNE, and MCDONALD (1982) have all completed Ph.D. dissertations on the topic without the ability to claim more than a partial solution for some cases. It would appear that the lexical analysis methodology proposed by BALLMER & BRENNENSTUHL, in which the enumeration of several hundred lexical forms is used to specify the nature of the semantic phenomena that underlie lexical meaning, may also be required for noun compounds and proper nouns.

Generation of dictionary entries. In 1972, Gove could find "no one who says that a machine may someday be able in some way to define" (GOVE, 1972). The automatic generation of dictionary entries has received surprisingly little attention from computational linguists, considering the many possible ways in which a computer could help lexicographers. VENEZKY (1969;

1973a) and VENEZKY ET AL. (1976a; 1976b; 1977a; 1977b) discuss a system in which "textual material may be stored, edited, concorded, and lemmatized to assist the researcher," but these are really just automated bookkeeping tasks and not creative construction of definitions. PAIKEDAY (1983) notes how a simple personal computer can today perform the basic tasks necessary to produce a dictionary. DAMERAU (1972) addressed the task of generating biographical dictionaries from textual information. However, it is not until GRANGER (1977), KEIRSEY, and MCNAUGHT (1983) that we find the beginnings of a response to Gove's original request in the description of computational linguistic techniques for learning the meaning of words from textual contexts and generating definitions automatically. Granger shows how a program that is analyzing a text can attempt to determine the meaning of a word from its context. Keirsey presents evidence that an ISA hierarchy can be used to learn new words from their textual contexts. McNaught describes the use of a terminology thesaurus to generate definitions in an online environment.

Parsing dictionary definitions. Dictionary definitions may be considered as a specialized form of text and may be parsed in the same way as the ordinary sentences of the language. AMSLER (1980) offers a grammar for verb definitions in the *MPD*, noting that it is primarily a grammar of "and's and or's," which reflects their extensive use in definitions. AHLSWEDE (1983a; 1983b) developed a string grammar for the adjective definitions of the *W7* and used it to parse such definitions with the Linguistic String Parser (LSP) (SAGER, 1981).

CONCLUSION

This chapter has presented a survey of machine-readable dictionaries over the past two decades. The dominant theme has been the impact of the changes in computer hardware on the work on MRDs. Storage for MRDs has evolved from punched cards and paper tape through magnetic tape and computer disc storage to the edge of new technologies such as videodisc and computer VLSI chips, which can be used for the dictionaries of the future. The computing environment has also progressed from batch processing through interactive computing to the edge of distributed computing systems and worldwide network access to data. If anything can be concluded about these changes, it is that we are very naive about how transitory our current hardware problems really are when they are taken in a historical perspective.

Tasks undertaken in the earliest days of MRD research reflected efforts to amass data, tabulate them, and present the hard-won computational outputs for others to digest. As computational linguistics began to unfold the layers of problems in making computers understand natural language, the scope of dictionary research began to diminish and an increased emphasis on parsing and generation of natural language became dominant. Today, having discovered that there are a multitude of possible techniques for treating natural-language problems on a small scale, there seems again to be a recognition that language is not going to be manipulated intelligently by computational means without comprehensive attention to the details of its description on a large scale.

If current trends continue, we can expect to find the MRD of the future to be a fusion of the needs of humans and computers. Information in MRDs will become better organized, more explicit, and much more accessible than ever before as computers are used to create, check, and present the contents of electronic archives of lexical information. As more and more information becomes computationally accessible, MRDs will grow and their contents will become easier to process. I fully expect that at some point in the future a computer will be able to explain to me what a word means. The only question is how distant is that future. The caution of this chapter's history is that the hardware of today still preoccupies us and clouds our ability to perceive the capabilities of the computers of the future in dealing with MRDs.

BIBLIOGRAPHY

AHLSWEDE, THOMAS E. 1983a. A Linguistic String Parser for Adjective Definitions. In: Williams, Stephanie, ed. Humans and Machines: The Interface through Language. Ablex; 1983.

AHLSWEDE, THOMAS E. 1983b. A Linguistic String Grammar of Adjective Definitions from *Webster's Seventh Collegiate Dictionary*. Chicago, IL 60616: Illinois Institute of Technology; 1983. 33p. (Master's thesis). Available from: University Microfilms, Ann Arbor, MI.

AITCHISON, JEAN; GILCHRIST, ALAN. 1979. Thesaurus Construction: A Practical Manual. London, England: Aslib; 1979. 95p. SBN: 85142-042-7.

ALFORD, M. H. T. 1973. The Computer and Lexicography. Bulletin of the Association for Literary and Linguistic Computing. 1973; I(3): 8–9.

ALLEN, JONATHAN. 1976a. Synthesis of Speech from Unrestricted Text. Proceedings of the IEEE. 1976; LXIV(4): 433–442. ISSN: 0018-9219.

ALLÉN, STURE. 1970a. Vocabulary Data Processing. In: Benediktsson, H., ed. The Nordic Languages and Modern Linguistics. Reykjavik, Iceland: Visindafelag Islendinga; 1970.

ALLÉN, STURE. 1976. On Phraseology in Lexicology. Cahiers de Lexicologie. 1976; 29: 83–90.

ALLÉN, STURE. 1977a. Text-Based Lexicography and Algorithmic Text Analysis. Association for Literary and Linguistic Computing Bulletin. 1977; 5(2): 126-131.

ALLÉN, STURE. 1977b. The Swedish Logotheque. A Computer-Based Test and Word Bank. In: Van Sterkenburg, P. G. J., ed. Lexicologie. Een bundel opstellen voor F. de Tollenaere. Groningen, The Netherlands: Groningen; 1977.

ALLÉN, STURE. 1978. The Role and Organization of a Corpus in the Investigation of a Language. Association for Literary and Linguistic Computing Bulletin. 1978; 6(1): 10–14.

ALLÉN, STURE. 1980. The Language Bank Concept. In: Raben, J.; Marks, G., eds. Data Bases in the Humanities and Social Sciences. New York, NY: Oxford; 1980.

ALLÉN, STURE. 1981a. The Lemma-Lexeme Model of the Swedish Lexical Data Base. In: Rieger, B., ed. Empirical Semantics 1-2. A Collection of New Approaches in the Field. Bochum; 1981.

ALLÉN, STURE. 1981b. Sprakdata Lexibase System: An Integrated View of a Lexical Project. In: Proceedings of the European Science Foundation Workshop on the Possibilities and Limits of the Computer in Producing and Publishing Dictionaries. Pisa, Italy: European Science Foundation; 1981.

ALTERMAN, RICHARD. 1983. A Dictionary Based on Concept Coherence. (Technical Report). Austin, TX: University of Texas at Austin, Computer Sciences Dept; 1983 April.

AMSLER, ROBERT A. 1969. Semantic Space: A Computer Technique for Modeling Word Meanings. New York, NY: New York University, Courant Institute of Mathematical Sciences; 1969 February. 132p. (Master's thesis).

AMSLER, ROBERT A. 1980. The Structure of the Merriam-Webster Pocket Dictionary. Austin, TX: University of Texas; 1980. (PhD thesis). Available from: University Microfilms, Ann Arbor, MI.

AMSLER, ROBERT A. 1981. A Taxonomy for English Nouns and Verbs. In: Proceedings of the Association for Computational Linguistics (ACL). 19th Annual Meeting ACL; 1981. 133-138. Available from: the author, Bell Communications Research, 445 South Street, Morristown, NJ 07960.

AMSLER, ROBERT A. 1982. Computational Lexicology: A Research Program. In: Proceedings of the American Federation for Information Processing Societies (AFIPS) National Computer Conference; AFIPS, 1982. 657-663.

AMSLER, ROBERT A. 1984. Lexical Knowledge Bases: Abstract of talk given at Panel Session on Machine-Readable Dictionaries. In: Proceedings of COLING84: 10th International Conference on Computational Linguistics (ACL); 1984 July 2-6; Stanford, CA. ACL; 1984. 458-459. Available from: D. E. Walker, Bell Communications Research, 435 South Street, Morristown, NJ 07960.

AMSLER, ROBERT A.; WHITE, JOHN S. 1979. Development of a Computational Methodology for Deriving Natural Language Semantic Structures via Analysis of Machine-Readable Dictionaries. Austin, TX: University of Texas at Austin, Linguistics Research Center; 1979 July. (Final Report on NSF Project MCS77-01315).

APRESYAN, YU. D. 1973. Synonymy and Synonyms. In: Kiefer, F., ed. Trends in Soviet Theoretical Linguistics. Dordrecht, The Netherlands: D. Reidel Publishing Co.; 1973. 173-199. ISBN: 90-277-0274-8. (Also see paper by Mel'čuk).

APRESYAN, YU. D.; MEL'ČUK, I. A.; ŽOLKOVSKY, A. D. 1970. Semantics and Lexicography: Toward a New Type of Unilingual Dictionary. In: Kiefer, F., ed. Studies in Syntax and Semantics: Foundations of Language. Dordrecht, The Netherlands: D. Reidel Publishing Co.; 1970.

ARCHBOLD, ARMAR A.; GROSZ, BARBARA; SAGALOWICZ, DANIEL. 1981. A TEAM User's Guide. Menlo Park, CA: SRI International; 1981 December. (Technical Note 254).

ARNOLD, DONNA I. 1979. Synonyms and the College-Level Dictionary. Dictionaries. 1979; (1): 103-112.

ARNOLD, DONNA I. 1980-1981. College-Level Dictionaries and Freshman Composition. Dictionaries. 1980-1981; (3): 69-79.

ARTHERN, PETER J. 1979. Machine Translation and Computerized Terminology Systems—A Translator's Viewpoint. In: Snell, Barbara M., ed. Translating and the Computer: Proceedings of a Seminar; 1978 November 14; London, England. Amsterdam, The Netherlands: North-Holland; 1979. 77-108. ISBN: 0-444-85302-2.

ASSOCIATION INTERNATIONALE DE LINGUISTIQUE APPLIQUÉE (AILA). 1981. Papers Presented at the Applied Linguistics Congress; 1981 August; Lund, Sweden. AILA; 1981.

BAHR, J. 1978. Reflections on the Project of a Lexical Data Bank. Cahiers de Lexicologie. 1978; 32(1): 55-64.

BAILEY, RICHARD W. 1969. Research Dictionaries. American Speech. 1969; 44: 166-172.

BAKER, ARTHUR, comp. and ed. 1908. The American Esperanto Book: A Compendium of the International Language. 2nd edition. Chicago, IL: American Esperanto Co.; 1908.

BAKIS, R. 1964. Dictionary Storage. IBM Technical Disclosures Bulletin. 1964; I(6): 55-56.

BALLMER, TH.; BRENNENSTUHL, W. 1981. Speech Act Classification: A Study of the Lexical Analysis of English Speech Activity Verbs. New York, NY: Springer-Verlag; 1981. 274p. ISBN: 0-387-10294-9.

BAR-HILLEL, YEHOSHUA. 1964. Language and Information. Reading, MA: Addison-Wesley, Inc.; 1964.

BARR, AVRON; FEIGENBAUM, EDWARD A., eds. 1981a. Semantic Primitives. In: The Handbook of Artificial Intelligence: Volume 1. Los Altos, CA: William Kaufmann, Inc.; 1981. 207-215.

BARR, AVRON; FEIGENBAUM, EDWARD A., eds. 1981b. Frames and Scripts. In: The Handbook of Artificial Intelligence: Volume 1. Los Altos, CA: William Kaufmann, Inc.; 1981. 216-222.

BAUER, LAURIE. 1979. On the Need for Pragmatics in the Study of Nominal Compounding. Journal of Pragmatics. 1979; 3: 45-50.

BAUER, LAURIE. 1983. English Word-Formation. Cambridge, England: Cambridge University Press; 1983. ISBN: 0-521-24167-7 (hard cover); ISBN: 0-521-28492-9 (paperback).

BELL, COLIN; JONES, KEVIN P. 1979. Back-of-the-Book Indexing: A Case for the Application of Artificial Intelligence. In: MacCaffery, Maxine; Gray, Kathleen, eds. Informatics 5: The Analysis of Meaning: Proceedings of a conference held by the Aslib Informatics Group and the BCS Information Retrieval Specialist Group; 1979 March 26-28; Oxford, England. London, England: Aslib; 1979. 155-161. ISBN: 0-85142-125-3.

BERESFORD, JOHN C.; POMERANCE, DEBORAH S. 1981. Federal Policy about Machine-Readable Statistical Documents. Government Publications Review. 1981; 81A: 221-229.

BERLIN, BRENT; BREEDLOVE, DENNIS E.; RAVEN, PETER H. 1974. Principles of Tzeltal Plant Classification. New York, NY: Academic Press; 1974.

BIERWISCH, M.; HEIDOLPH, K. E., eds. 1970. Progress in Linguistics. The Hague, The Netherlands: Mouton; 1970.

BOLINGER, DWIGHT. 1965. The Atomization of Meaning. Language. 1965; 41(4): 555-573.

BOLINGER, DWIGHT. 1975. Aspects of Language. 2nd edition. New York, NY: Harcourt, Brace, Jovanovich; 1975.

BORKO, HAROLD, ed. 1967. Automated Language Processing. New York, NY: John Wiley and Sons, Inc.; 1967. 386p.

BRACHMAN, RONALD. 1982. What 'ISA' Is and Isn't. In: Proceedings of the Canadian Society for Computational Studies of Intelligence [Meeting]; 1982 May 17-19; Saskatoon, Saskatchewan, Canada. 1982. 212-220.

BRATLEY, P.; LUSIGNAN, S. 1976. Information Processing in Dictionary Making: Some Technical Difficulties. Computers and the Humanities. 1976; 10: 133-143.

BRAUN, STEPHAN; SCHWIND, CAMILLA. 1976. Automatic, Semantic-Based Indexing of Natural Language Texts for Information Retrieval Systems. Information Processing and Management. 1976; 12: 147-153.

BROWN, ALAN S. 1976. Catalog of Scaled Verbal Material. Memory and Cognition. 1976; 4(1B): 1S-45S.

BROWN, AUGUSTUS F. 1963. Normal and Reverse English Word List. Philadelphia, PA: University of Pennsylvania; 1963. 8 volumes.

BROWN, J. C. 1960. Loglan. Scientific American. 1960 June; 202: 53-63. ISSN: 0036-8733.

BYE, TOM; DILLER, TIMOTHY; OLNEY, JOHN. 1975. User's Guide to the SOLAR Semantic Analysis File. Santa Monica, CA: System Development Corp. (SDC); 1975 April 31. (Technical Report TM-5292/001/00).

BYRD, ROY J. · 1983. Word Formation in Natural Language Processing Systems. In: Proceedings of the 8th International Joint Conference on Artificial Intelligence, 1983 August 8-12; Karlsruhe, West Germany. International Joint Conferences on Artificial Intelligence, Inc.; 1983. 704-706. ISBN: 0-86576-064-0.

CALZOLARI, NICOLETTA. 1977. An Empirical Approach to Circularity in Dictionary Definitions. Cahiers de Lexicologie. 1977; 31(2): 118-128.

CALZOLARI, NICOLETTA. 1983. Lexical Definitions in a Computerized Dictionary. Computers and Artificial Intelligence. 1983; 3.

CALZOLARI, NICOLETTA; CECCOTTI, M. L. 1981. Organizing a Large Scale Lexical Database. In: Actes du Congrès International Informatique et Sciences Humaines; 1981 November 18-21; Liege, Belgium.

CARROLL, JOHN B. 1971. Behind the Scenes in the Making of a Corpus-Based Dictionary and a Word Frequency Book. Paper presented at: National Council of Teachers of English (NCTE) Conference; 1971 November.

CARROLL, JOHN B.; DAVIES, PETER; RICHMAN, BARRY. 1971. The American Heritage Word Frequency Book. Boston, MA: Houghton Mifflin; 1971.

CARROLL, JOHN M. 1981a. Nameheads. Yorktown Heights, NY 10598: IBM Thomas J. Watson Research Center; 1981 September 9. (Research Report RC 9018 (No. 39499). Available from: the Center.

CARROLL, JOHN M. 1981b. Towards an Integrated Study of Creative Naming. Yorktown Heights, NY 10598: IBM Thomas J. Watson Research Center; 1981 August 31. 22p. (Research Report RC 9016 (No. 39483). Available from: the Center.

CARROLL, JOHN M. 1982. The Name Game: Creative and Practical Names for a Building. Yorktown Heights, NY 10598: IBM Thomas J. Watson Research Center; 1982 July 2. 27p. (Technical Report RC 9450). Available from: the Center.

CASSIDY, FREDERIC G. 1977a. Computer-Aided Usage 'Labeling' in a Dictionary. Computers and the Humanities. 1977; 11: 89-99.

CERCONE, NICK; KRAUSE, MAX; BOATES, JOHN. 1983. Minimal and Almost Perfect Hash Function Search with Application to Natural Language Lexicon Design. In: Cercone, Nick, ed. Computational Linguistics. Oxford, England: Pergamon Press; 1983. p.215-231. ISBN: 0-08-030 253-X.

CHAPIN, PAUL G.; NORTON, L. M. 1968. A Procedure for Morphological Analysis. The Mitre Corp.; 1968. (Technical Report MTP-101).

CHAPKIS, RAOUL. 1970. Exercises in Computational Linguistics. Amsterdam, The Netherlands: Mathematisch Centrum; 1970.

CHARNIAK, EUGENE. 1983. Passing Markers: A Theory of Contextual Influence in Language Comprehension. Cognitive Science. 1983; 7: 171-190.

CHERRY, LORINDA L. 1978. PARTS—A System for Assigning Word Classes to English Text. Bell Laboratories; 1978. (Computing Science Technical Report 81).

CHERRY, LORINDA L. 1981. A Toolbox for Writers and Editors. In: Digest of the Office Automation Conference; 1981, Houston, TX. AFIPS Press; 1981. 221-227.

CHERRY, LORINDA L.; MACDONALD, NINA H. 1983. The *UNIX* Writer's Workbench Software. BYTE. 1983; 8(10): 241-246. ISSN: 0360-5280.

CHERRY, LORINDA L.; VESTERMAN, W. 1981. Writing Tools—The *STYLE* and *DICTION* Programs. Bell Laboratories; 1981. (Computing Science Technical Report 91).

CHICHELLI, RICHARD J.; IOBST, JOHN W. 1977. MICROMARK. American Newspaper Publishers Association/Research Institute (ANPA/RI); 1977 September 1. (Technical Report CADTR-7702).

CIGNONI, L.; PETERS, C.; ROSSI, S., eds. 1983. European Science Foundation: Survey of Lexicographic Projects. Pisa, Italy: Istituto di Linguistica Computazionale; 1983.

CLARK, WALTER JOHN. 1907. International Language, Past, Present and Future, with Specimens of Esperanto and Grammar. J. M. Dent and Co.; 1907.

COLLIER, L. W.; KAY, CHRISTIAN J. 1980-1981. The Historical Thesaurus of English. Dictionaries. 1980-1981; (2-3): 80-89.

COLLIN, A. J. T. 1960. The Automatic Construction of a Glossary. Information and Control. 1960; IX(3): 211-230.

CONNOR, GEORGE ALAN, comp. 1966. Esperanto: The World Interlanguage. 1966. 249p.

COOK, MICHAEL GARNET. 1980. Machine-Readable Archives. In: Archives and the Computer. Boston, MA: Butterworths; 1980. 107-145.

COOK, PETER R. 1984. Electronic Encyclopedias. BYTE. 1984 July; 9(7): 151-170. ISSN: 0360-5280.

COX, J. 1979. Crossword Compiling Puzzles the Programmer. Computer Weekly. 1979 August 30: 17.

CRAWFORD, T. D. 1978. Wanted—Computer Readable Dictionaries. Association for Literary and Linguistic Computing Bulletin. 1978; 6(2): 177-179.

DAMERAU, FREDERICK J. 1964. A Technique for Computer Detection and Correction of Spelling Errors. Communications of the ACM. 1964; 7(3): 171-176. ISSN: 0001-0782.

DAMERAU, FREDERICK J. 1970. Automatic Parsing for Content Analysis. Communications of the ACM. 1970; 13(6): 356-360. ISSN: 0001-0782.

DAMERAU, FREDERICK J. 1972. Automatic Compilation of Biographical Dictionaries. Information Storage and Retrieval. 1972 December; 8(6): 315-327. ISSN: 0020-0271.

DAVIS, OSSIE. 1969. The Language of Racism: The English Language is My Enemy. In: Postman, Neil; Weingartner, Charles; Morean, Terence P., eds. Language in America. New York, NY: Pegasus; 1969. 74-75.

DILLER; T.; OLNEY, JOHN. 1974. SOLAR (A Semantically-Oriented Lexical Archive): Current Status and Plans. Computers and the Humanities. 1974; 8: 301-311. ISSN: 0010-4817.

DOLBY, J. L.; RESNIKOFF, H. L., comps. 1964. The English Word Speculum. Sunnyvale, CA: Lockheed Missiles & Space Company; 1964. 6 volumes.

DOLBY, J. L.; RESNIKOFF, H. L.; MACMURRAY, E. 1963. A Tape Dictionary for Linguistic Experiments. In: Proceedings of the American Federation of Information Processing Societies (AFIPS): Fall Joint Computer Conference; Volume 24; 1963. Baltimore, MD: Spartan Books; 1963. 419-423.

DOYLE, LAUREN B. 1961. Semantic Road Maps for Literature Searchers. Journal of the Association for Computing Machinery. October 1961; 8(4): 553-578. ISSN: 0004-5411.

DUDA, RICHARD O.; HART, PETER E.; NILSSON, NILS J.; REBOH, RENE; SLOCUM, JONATHAN; SUTHERLAND, GEORGIA L. 1977. Development of a Computer-Based Consultant for Mineral Exploration. Menlo Park, CA: SRI International; 1977. (Technical Report).

DUNCAN, WILLIAM WALTER. 1970. How White is your Dictionary? ETC: A Review of General Semantics. 1970 March; 27(1): 89-90.

EARL, LOIS L. 1967. Automatic Determination of Parts of Speech of English Words. Mechanical Translation. 1967 September/December; 10(3/4): 53-67. NTIS: AD 687 629.

EARL, LOIS L. 1973. Use of Word Government in Resolving Syntactic and Semantic Ambiguities. Information Storage and Retrieval. 1973; 9(12): 639-664.

ELECTRONIC LIBRARY. 1983. New Videodisc Supplier. The Electronic Library. July 1983; 1(3): 165-167. ISSN: 0260-6658.

ELECTRONICS. 1978. Single Silicon Chip Synthesizes Speech in $50 Learning Aid. Electronics. 1978 June 22; 51(13): 39-40. ISSN: 0013-5070.

ETHERINGTON, DAVID W.; REITER, RAYMOND. 1983. On Inheritance Hierarchies with Exceptions. In: Proceedings of the National Conference on Artificial Intelligence. 95 First St., Los Altos, CA 94022: William Kaufmann, Inc.; 1983. 104-108. ISBN: 0-86576-065-9.

EVENS, MARTHA W.; LITOWITZ, BONNIE E.; MARKOWITZ, JUDITH A.; SMITH, RAOUL N.; WERNER, OSWALD. 1980. Lexical-Semantic Relations: A Comparative Survey. Current Inquiry into Language and Linguistics. Volume 34. P.O. Box 5677, Station 'L', Edmonton, Alberta, Canada T6C 4G1: Linguistic Research, Inc.; 1980. ISBN: 0-88783-036-6.

EVENS, MARTHA W.; SMITH, RAOUL N. 1978. A Lexicon for a Computer Question-Answering System. American Journal of Computational Linguistics. 1978; XV (4 (Microfiches 81, 83)): 1-101. Also available from: Technical Dept. of Computer Sciences, Illinois Inst. of Technology. Sept., 1977, Report no. 77-14.

EVENS, MARTHA W.; VANDENDORPE, JAMES; WANG, YIH-CHEN. 1983. Lexical-Semantic Relations in Information Retrieval. Chicago, IL: Illinois Institute of Technology; 1983. (Technical Report).

FAHLMAN, SCOTT E. 1977. A System for Representing and Using Real-World Knowledge. Cambridge, MA: Massachusetts Institute of Technology; 1977. (PhD thesis).

FAHLMAN, SCOTT E. 1979. NETL: A System for Representing and Using Real-World Knowledge. Cambridge, MA: M.I.T. Press; 1979. 278p. ISBN: 0-262-0609-8.

FAHLMAN, SCOTT E.; HINTON, GEOFFREY E.; SEJNOWSKI, TERRENCE J. 1983. Massively Parallel Architectures for AI: NETL, Thistle, and Boltzmann Machines. In: Proceedings of the National Conference on Artificial Intelligence. 95 First St., Los Altos, CA 94022: William Kaufmann, Inc.; 1983. 109-113. ISBN: 0-86576-065-9.

FEGER, O. 1975. A Program for the Construction of Crossword Puzzles. Angewandte Informatik. 1976; 17(5): 189-195. ISSN: 0013-5704.

FERGUSON, CHARLES A. 1977. Constructing a Textual and Lexicographic Archive for Language Research. Stanford, CA: Linguistics Department, Stanford University; 1977.

FILLMORE, CHARLES J. 1968. The Case for Case. In: Bach, E.; Harms, R., eds. Universals of Linguistic Theory. New York, NY: Holt, Rinehart, and Winston; 1968. 1-90.

FINKENSTAEDT, THOMAS; LEISI, ERNEST; WOLFF, DIETER. 1970. A Chronological English Dictionary: Listing 80,000 Words in Order of their Earliest Known Occurrences. Heidelberg, West Germany: Carl Winter; 1970.

FINKENSTAEDT, THOMAS; WOLFF, DIETER; NEUHAUS, H. JOACHIM; HERGERT, WINFRED. 1973. Ordered Profusion: Studies in Dictionaries and the English Lexicon. Heidelberg, West Germany: Carl Winter; 1973. 166p. ISBN: 3-533-02253-6 Kt.

FLAGG, GORDON. 1983. Online Encyclopedias: Are they Ready for Libraries? Are Libraries Ready for Them? American Libraries. 1983 March; 134-135. ISSN: 0002-9769.

FOX, MARK S.; BEBEL, DONALD J.; PARKER, ALICE C. 1980. The Automated Dictionary. Computer. 1980 July; 35-48. ISSN: 0018-9162.

FRAENKEL, AVIEZRI S.; MOR, MOSHE. 1984. Combinatorial Compression and Partitioning of Large Dictionaries. The Computer Journal. November 1984; 26(4): 336-343. ISSN: 0010-4620.

FRANCIS, WINTHROP NELSON. 1964. Manual of Information to Accompany a Standard Sample of Present-Day Edited American English, for Use with Digital Computers. Providence, RI: Brown University, Department of Linguistics; 1964.

FRANCIS, WINTHROP NELSON. 1965. A Standard Corpus of Edited Present-Day American English. College English. 1965; 26: 267-273.

FRANCIS, WINTHROP NELSON. 1980. A Tagged Corpus—Problems and Prospects. In: Greenbaum, S.; Leech, G.; Svartvik, J., eds. Studies in English Linguistics, for Randolph Quirk. New York, NY: Longman; 1980.

FRANCIS, WINTHROP NELSON; KUČERA, HENRY. 1982. Frequency Analysis of English Usage: Lexicon and Grammar. Boston, MA: Houghton Mifflin Co.; 1982. ISBN: 0-395-32250-2.

FRAWLEY, WILLIAM. 1980-1981. Lexicography and the Philosophy of Science. Dictionaries. 1980-1981; 2/3 18-27.

FREDERICKSEN, DICK H. 1975. An Organization for a Dictionary of Senses. Yorktown Heights, NY: IBM Thomas J. Watson Research Center; 1975 July 31. 14p. (Technical Report RC 5548 (No. 23775).
FRIEDMAN, JOYCE. 1971. A Computer Model of Transformational Grammar. New York, NY: American Elsevier; 1971.
FRIES, CHARLES C.; TRAVER, A. AILEEN. 1940. English Word Lists: A Study of Their Adaptability for Instruction. Washington, DC: American Council on Education; 1940.
GALLI, ENRICO J.; YAMADA, H. 1967. An Automatic Dictionary and the Verification of Machine-Readable Text. IBM Systems Journal. 1967; 6(3): 192-207.
GALLI, ENRICO J.; YAMADA, HISAO, M. 1968. Experimental Studies in Computer-Assisted Correction of Unorthographic Text. IEEE Transactions on Engineering Writing and Speech. 1968; EWS-11(2): 75-84.
GARFIELD, EUGENE. 1981a. Current Comments: Another Look at ISI's Master Dictionary—Aiding Scientific Etymology and Reflecting Changes in Science. Current Contents. 1981 October 26: 5-12.
GARFIELD, EUGENE. 1981b. Current Comments: What's in a Surname? Current Contents. February 16, 1981; 5-9.
GELLERSTAM, MARTIN, ed. 1982. Sprakdata Research Report 1982. Goteborg, Sweden: Dept. of Computational Linguistics, Institutionen for sprakvetenskaplig databehandling, Göteborgs Universitet; 1982. (Technical Report). ISSN: 0347-9218.
GERBNER, GEORGE; HOLSTI, OLE R.; KRIPPENDORFF, KLAUS; PAISLEY, WILLIAM J.; STONE, PHILIP J., eds. 1969. The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques. New York, NY: John Wiley & Sons, Inc.; 1969. 597p. LC: 74-78476; SBN: 471 29660.
GERSHUNY, H. LEE. 1973. Public Doublespeak: The Dictionary. College English. 1973: 938-942.
GERSHUNY, H. LEE. 1974. Sexist Semantics in the Dictionary. ETC: A Review of General Semantics. 1974 June; 31(2): 159-169.
GIVON, TALMY. 1967a. Transformations of Ellipsis, Sense Development and Rules of Lexical Derivation. Santa Monica, CA 90406: System Development Corp. (SDC) 2500 Colorado Ave.; 1967 July 22. 62p. (Technical Report SP-2896).
GIVON, TALMY. 1967b. Some Noun-to-Noun Derivational Affixes. Santa Monica, CA 90406: System Development Corp. (SDC), 2500 Colorado Ave.; 1967 July 20. 47p. (Technical Report SP-2893).
GOETSCHALCKX, J.; ROLLING, L., eds. 1982. Lexicography in the Electronic Age. Amsterdam, The Netherlands: North-Holland Publishing; 1982. 276p. ISBN: 0-444-86404-0.
GORIN, RALPH E. 1971. *SPELL:* Spelling Check and Correction Program. Online documentation in file DOC: SPELL.DOC at CMU-CS-A.ARPA. Pittsburgh, PA: Carnegie Mellon University, Dept. of Computer Science; 1971.
GOVE, PHILIP B., ed. 1961. Webster's Third New International Dictionary of the English Language. Springfield, MA: G. & C. Merriam Company; 1961. 2662p. ISBN: 0-87779-10X-X.
GOVE, PHILIP B. 1972. English Dictionaries of the Future. In: Weinbrot, Howard D., ed. New Aspects of Lexicography: Literary Criticism, Intellectual History and Social Change. Carbondale, IL: Southern Illinois University Press; 1972.

GRANGER, RICHARD H. 1977. FOUL-UP: A Program that Figures Out Meanings of Words from Context. In: Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI). Pittsburgh, PA 15213: Carnegie-Mellon University, Department of Computer Science; 1977. 172-178. (Also see paper by Cercone).
GRANGER, RICHARD H. 1982. Scruffy Text Understanding: Design and Implementation of 'Tolerant' Understanding. In: Proceedings of the Association for Computational Linguistics 20th Annual Meeting; 1982 June 16-18. Toronto, Ontario, Canada. Association for Computational Linguistics; 1982. 157-160.
GRUBER, J. S. 1967. Functions of the Lexicon in Formal Descriptive Grammars. Santa Monica, CA: System Development Corp. (SDC); 1967. 148p. (Technical Report TM-3770).
HARALICK, ROBERT M.; RIPKEN, KNUT. 1975. An Associative-Categorical Model of Word Meaning. Artificial Intelligence. 1975; 6: 75-99.
HARTER, STEPHEN P.; KISTER, KENNETH F. 1981 Online Encyclopedias: The Potential. Library Journal. 1981 September: 1600-1602. ISSN: 0000-0027.
HAYES, PHILIP J. 1975. A Process to Implement Some Word-Sense Disambiguation. Rochester, NY: University of Rochester Computer Science Department; 1975. (Technical Report 6).
HAYES, PHILIP J. 1977. Some Association-Based Techniques for Lexical Disambiguation by Machine. University of Rochester; 1977. (PhD thesis).
HAYS, DAVID G. 1967. Introduction to Computational Linguistics. New York, NY: Elsevier; 1967.
HEIDORN, GEORGE E.; JENSEN, KAREN; MILLER, LANCE A.; BYRD, ROY J.; CHODOROW, MARTIN S. 1982. The EPISTLE Text-Critiquing System. IBM System Journal. 1982; 21(3): 305-326.
HEISE, DAVID R. 1965. Semantic Differential Profiles for 1000 Most Frequent English Words. Psychological Monographs. Washington, DC: American Psychological Association; 1965.
HELLBERG, STAFFAN. 1972. Computerized Lemmatization without the Use of a Dictionary: A Case Study from Swedish Lexicology. Computers and the Humanities. 1972; 6: 209-212. ISSN: 0010-4817.
HIRSCHMAN, L.; GRISHMAN, R.; SAGER, N. 1975. Grammatically-Based Automatic Word Class Formation. Information Processing and Management. 1975; 11: 39-57. ISSN: 0306-4573.
HIRSCHMAN, L.; SAGER, N. 1982. Automatic Information Formatting of a Medical Sublanguage. In: Kittredge, Richard I.; Lehrberger, John, eds. Sublanguage: Studies of Language in Restricted Semantic Domains. New York, NY: W. de Gruyter; 1982.
HIRST, GRAEME. 1983. Semantic Interpretation Against Ambiguity. Brown University; 1983 December. (PhD thesis). Also available as: Brown University, Department of Computer Science, Technical Report 83-25.
HIRST, GRAEME; CHARNIAK, EUGENE. 1982. Word Sense and Case Slot Disambiguration. In: Proceedings of the National Conference on Artificial Intelligence; 1982 August 18-20; Pittsburgh, PA. Los Altos, CA: William Kaufmann, Inc.; 1982. 95-98. ISBN: 0-86576-043-8.
HOLMES, P. L. 1980. The Re-Use of Machine-Readable Data and Copyright—A Pragmatic Approach to the Problems. In: [Proceedings of the]

4th International Online Information Meeting; 1980, December 9-11; London, England. Oxford, England: Learned Information; 1980. 1-13.

HORWELL, VERONICA; WINCHESTER, SIMON. 1984. Flower of English Yanked into a Computer. Sunday Times. 1984 February 5. p5.

HOUGHTON MIFFLIN COMPANY. 1972. The American Heritage School Dictionary. Boston, MA: Houghton Mifflin Company; 1972. 992p. ISBN: 0-395-13850-7.

HOUGHTON MIFFLIN COMPANY. 1981. WORDS Lexical Database. Advertising Literature available from: David R. Replogle, Senior Vice President, Houghton Mifflin Company, Two Park Street, Boston, MA 02107.

HOUGHTON MIFFLIN COMPANY. 1982. The American Heritage Dictionary of the English Language. 2nd College edition. Boston, MA: Houghton Mifflin Company; 1982.

HUNN, EUGENE S. 1977. Tzeltal Folk Zoology. New York, NY: Academic Press; 1977.

ISAACSON, KATHY. 1982. Machine-Readable Information in the Library. RQ: Reference Quarterly. 1982 Winter: 64-70. ISSN: 0033-7072.

JACOBSON, SVEN. 1978. Semantic Representations with Added Pragmatic Information. In: Gregersen, Kirsten; Basboll, Hans; Mey, Jacob, eds. Papers from: 4th Scandinavian Conference of Linguistics, 1978 January, 6-8; Hindsgavl, Denmark. Odense University Press; 1978. 295-301. (Also see paper by Bauer).

JENSEN, KAREN; HEIDORN, GEORGE E. 1982. The Fitted Parse: 100% Parsing Capability in a Syntactic Grammar of English. Yorktown Heights, NY: IBM T. J. Watson Research Center; 1982 December 6. 18p. (Research Report Number RC 729 (No. 42958)).

JOHANSSON, STIG. 1980. Some Thoughts on the Use of Computers in Linguistic Research. Humanistiske Data. 1980; 1: 31-39.

JOHANSSON, STIG, ed. 1982. Computer Corpora in English Language Research. Bergen, Norway: Norwegian Computing Centre for the Humanities; 1982.

KAPLAN, RONALD; BRESNAN, JOAN. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In: The Mental Representation of Grammatical Relations. Cambridge, MA: M.I.T. Press; 1982.

KATZ, JERROLD J.; FODOR, JERRY A. 1964. The Structure of Language: Readings in the Philosophy of Language. Englewood Cliffs, NJ: Prentice-Hall; 1964. 612p.

KAY, ALAN; GOLDBERG, ADELE. 1977. Personal Dynamic Media. Computer. 1977 March; 10(3): 31-41. (An expanded version of this paper was produced as 'Personal Dynamic Media,' Xerox Palo Alto Research Center Technical Report SSL 76-1, March, 1976).

KAY, CHRISTIAN J. 1979. The Historical Thesaurus of English. Old English Newsletter. 1979 Fall; 13(1).

KAY, MARTIN; MARTINS, G. R. 1970. The *MIND* Systems: The Morphological-Analysis Program. Santa Monica, CA: The RAND Corp.; 1970. (Technical Report RM-626512-PR).

KAY, MARTIN; ZIEHE, T. 1965. Natural Language in Computer Form. Santa Monica, CA: The RAND Corp.; 1965 February. (Technical Report RM-4390).

KEIRSEY, DAVID M. 1982. Word Learning with Hierarchy-Guided Inference. In: Proceedings of the National Conference on Artificial Intelligence, 1982 August 18-20. Pittsburgh, PA. Los Altos, CA: William Kaufmann, Inc.; 1982. 99-102. ISBN: 0-86576-043-8.

KELLY, EDWARD FRANCIS; STONE, PHILIP J. 1975. Computer Recognition of English Word Senses. Amsterdam, The Netherlands: North-Holland; 1975.

KEPLINGER, MICHAEL S. 1980. Copyright and Information Technology. In: Williams, Martha E., ed. Annual Review of Information Science and Technology: Volume 15. White Plains, NY: Knowledge Industry Publications, Inc. for the American Society for Information Science; 1980. ISSN: 0066-4200; CODEN: ARISBc.

KIPFER, BARBARA. 1983. Bibliography of Computer Applications in Lexicography and Lexicology. Dictionaries. 1983.

KISTER, KENNETH F. 1977. Dictionary Buying Guide. New York, NY: R. R. Bowker Co.; 1977. 358p. ISBN: 0-8352-1038-3.

KITTREDGE, RICHARD I. 1983. Semantic Processing of Texts in Restricted Sublanguages. In: Cercone, Nick, ed. Computational Linguistics. Oxford, England: Pergamon Press; 1983. 45-58. (Also see paper by Cercone, Krause & Boates; Also available in: Computers and Mathematics with Applications, 9(1): special issue). ISBN: 0-08-030253-X.

KITTREDGE, RICHARD I.; LEHRBERGER, JOHN. 1982. Sublanguage: Studies of Language in Restricted Semantic Domains. Berlin, West Germany: Walter de Gruyter; 1982. 240p. ISBN: 3-11-008244-6.

KLEIN, SHELDON; SIMMONS, ROBERT F. 1963. A Computational Approach to Grammatical Coding of English Words. Journal of Association for Computing Machinery. 1963; 10: 334-347. ISSN: 0004-5411.

KNOWLES, FRANCIS E. 1982. The Pivotal Role of the Various Dictionaries in an MT System. In: Lawson, Veronica, ed. Practical Experience of Machine Translation: Proceedings of a Conference; 1981 November 5-6; London, England. Amsterdam, The Netherlands: North-Holland; 1982. 149-162. ISBN: 0-444-86381-8.

KNUTH, DONALD E. 1979. *TeX* and *METAFONT*: New Directions in Typesetting. Bedford, MA: American Mathematical Society & Digital Press; 1979. 256p. ISBN: 0-932376-02-9.

KNUTH, DONALD E. 1984. The *TeX*Book. Reading, MA: Addison-Wesley; 1984. ISBN: 0-201-13448-9.

KRASKE, ROBERT. 1975. The Story of the Dictionary. New York, NY: Harcourt, Brace, Jovanovich; 1975. ISBN: 0-15-280850-7.

KUCERA, HENRY; FRANCIS, WINTHROP NELSON. 1967. Computational Analysis of Present-Day American English. Providence, RI: Brown University Press; 1967. 424p.

KUČERA, HENRY. 1969. Computers in Language Analysis and Lexicography. In: American Heritage Dictionary of the English Language. New York, NY: Houghton Mifflin, NY: 1969. XXXVII-XL. ISBN: 395-09064-4.

KUČERA, HENRY. 1975. Computers in Linguistics and in Literary Studies. Providence, RI: Brown University, Department of Linguistics; 1975. 150p.

KUČERA, HENRY. 1982. The Mathematics of Language. In: The American Heritage Dictionary. 2nd College edition. 1982. 37-41.

KUČERA, HENRY; FRANCIS, WINTHROP NELSON. 1967. Computational Analysis of Present-Day American English. Providence, RI: Brown University Press; 1967. 424p.

LAFFAL, JULIUS. 1970. Toward a Conceptual Grammar and Lexicon. Computers and the Humanities. 1970 January; 4(3): 173-186. ISSN: 0010-4817.

LAFFAL, JULIUS. 1973. A Concept Dictionary of English. Essex, CT 06426: Gallery Press, North Main Street; 1973. 305p. ISBN: 0-470-51160-5.

LAMB, JOHN. 1984. Bringing the OED On-Line. Datamation. 1984 February; 30(2): 88,91. ISSN: 0011-6963.

LAMB, SYDNEY M.; GOULD, LAURA; JACOBSEN, WILLIAM H., JR. 1966. A High-Speed Large-Capacity Dictionary System. In: Hays, David G., ed. Readings in Automatic Language Processing. New York, NY: Elsevier; 1966.

LANDAU, SIDNEY I. 1981. Computer Use in Medical Lexicography: IDMB. New York, NY: John Wiley and Sons, Inc.; 1981.

LEES, R. B. 1960. The Grammar of English Nominalizations. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics, 1960. Publication no 12.

LEES, R. B. 1970. Problems in Grammatical Analysis of English Nominal Compounds. In: Bierwirsch, M.; Herdolph, K., eds. Progress in Linguistics. The Hague, The Netherlands: Mouton; 1970. 174-186.

LEHRER, ADRIENNE. 1971. Semantics: An Overview. The Linguistic Reporter. 1971 Fall; 13(4): 13-23.

LENAT, DOUGLAS B.; BORNING, ALAN; MCDONALD, DAVID; TAYLOR, CRAIG; WEYER, STEPHEN. 1983. Knoesphere: Building Expert Systems with Encyclopedic Knowledge. In: IJCAI-83: Proceedings of the 8th International Joint Conference on Artificial Intelligence; 1983 August 8-12; Karlsruhe, West Germany. Los Altos, CA: William Kaufmann, Inc.; 1983. 167-169. ISBN: 0-86576-064-0.

LERNER, RITA G.; METAXAS, TED; SCOTT, JOHN T.; ADAMS, PETER D.; JUDD, PEGGY. 1983. Primary Publication Systems and Scientific Text Processing. In: Williams, Martha E., ed. Annual Review of Information Science and Technology: Volume 18. White Plains, NY: Knowledge Industry Publications, Inc. for the American Society for Information Science; 1983. ISSN: 0066-4200; CODEN ARISBc.

LIANG, FRANKLIN MARK. 1983. Word Hy-phen-a-tion by Com-put-er. Stanford, CA: Stanford University Department of Computer Science; 1983 August. (PhD dissertation). Available from: the Dept as Stanford Computer Science Research Report no. STAN-CS-83-977.

LIPPMANN, E. O. 1975. On-Line Generation of Terminological Digests in Language Translation: An Aid in Terminology Processing. IEEE Transactions on Professional Communication. 1975 December; PC-18(4): 309-319. Also available from IBM as IBM Technical Report RC-5624.

LITKOWSKI, KENNETH C. 1975. Toward Semantic Universals. University of Delaware; 1975 November. (Delaware Working Papers in Language Studies 18).

LITKOWSKI, KENNETH C. 1976. On Dictionaries and Definitions. University of Delaware; 1976 March. (Delaware Working Papers in Language Studies 17).

LITKOWSKI, KENNETH C. 1978. Models of the Semantic Structure of Dictionaries. American Journal of Computational Linguistics. 1978.

LJUNG, MAGNUS. 1974. A Frequency Dictionary of English Morphemes. Stockholm, Sweden: AWE/Gebers; 1974.

LLOYD, SUSAN M., ed. 1982. Roget's Thesaurus of English Words and Phrases. London, England: Longman; 1982. 1247p. ISBN: 0-582-55635-X (standard edition).

LONG, THOMAS HILL, ed. 1979. Longman Dictionary of English Idioms. (LDOEI). London, England: Longman; 1979.

MACDONALD, NINA H.; FRASE, LAWRENCE T.; GINGRICH, PATRICIA S.; KEENAN, STACEY A. 1982. The Writer's Workbench: Computer Aids for Text Analysis. IEEE Transactions on Communication. 1982. ISSN: 0090-6778.

MACDONALD, NINA H.; FRASE, LAWRENCE T.; KEENAN, STACEY A. 1980. Writer's Workbench: Computer Programs for Text Editing and Assessment. Murray Hill NJ: Bell Laboratories; 1980. (Technical Report).

MACDONALD, NINA H.; POLLER, MERLE P. 1982. Language Processing Tools: The Writer's Workbench and Extensions. In: OAC '82: Digest of papers from the Office Automation Conference; 1982 April 5-7; San Francisco, CA. AFIPS Press; 1982. 667-673. ISBN: 0-88283-036-8; ISSN: 0272-4855.

MACHINA, K. 1967. 'Having', 'Characterized by' and 'Marked by' in *Webster's Seventh New Collegiate Dictionary.* Santa Monica, CA: System Development Corp. (SDC); 1967 November 6. 28p. (SDC Technical Report TM-3722).

MARCHAND, H. 1969. The Categories and Types of Present-Day English Word Formation. Munich, West Germany: C. H. Beck; 1969.

MAU, ERNEST E. 1982. Create Word Puzzles with Your Microcomputer. Rochelle Park, NJ: Hayden; 1982. 304p. ISBN: 0-8104-6251-6.

MAZLACK, LAWRENCE J. 1972. An Investigation of Computer Generated Crossword Puzzles. St. Louis, MO: Washington University; 1972 December. (PhD dissertation).

MAZLACK, LAWRENCE J. 1973a. Data Structures Required for Crossword Puzzle Construction. In: Proceedings of the American Society for Information Science (ASIS) Meeting: Volume 10; 1973 October 21-25. Westport, CT: Greenwood Press; 1973. 141-142.

MAZLACK, LAWRENCE J. 1973b. Single and Double Rooted Structures to Hold Very Large Dictionaries. Guelph, Ontario, Canada: University of Guelph, Computer and Information Science Department; 1973. 9p.

MAZLACK, LAWRENCE J. 1976. Computer Construction of Crossword Puzzles Using Precedence Relationships. Artificial Intelligence. 1976; 7(1): 1-19. ISSN: 0004-3702.

MCCALLUM, DOUGLAS R.; PETERSON, JAMES L. 1982. Computer-Based Readability Indexes. In: Proceedings of the Association for Computing Machinery (ACM) Conference; 1982 October 25-27; Dallas, TX. ACM; 1982. 44-48. ISBN: 0-89791-085-0.

MCDAVID, RAVEN I.; DUCKERT, AUDREY R., eds. 1973. Lexicography in English. New York, NY: New York Academy of Sciences; 1973.

MCDONALD, DAVID B. 1981. Compound: A Program that Understands Noun Compounds. In: Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI); 1981 August 24-28; Vancouver, B. C., Canada. The IJCAI; The Canadian Society for Computational Studies of Intelligence; The American Association for Artificial Intelligence; 1981. 1061. (Also see paper by Small).

MCDONALD, DAVID B. 1982. Understanding Noun Compounds. Pittsburgh, PA: Carnegie-Mellon University; 1982 January. 249p. (PhD dissertation). Available from: The University as Technical Report CMU-CS-82-102.

MCNAUGHT, JOHN. 1981. Terminology Data Banks: A Model for a British Linguistic Data Bank (LDB). In: Proceedings of an Aslib Conference. London, England: Aslib; 1981. 297-308.

MCNAUGHT, JOHN. 1982. Specialised Lexicography in the Context of a British Linguistic Data Bank. See reference: Goetschalckx, J.; Rolling, L., eds.

MCNAUGHT, JOHN. 1983. The Generation of Term Definitions from an On-Line Terminology Thesaurus. In: [Proceedings of the] Association for Computational Linguistics European Chapter 1st Conference; Pisa, Italy. Association for Computational Linguistics; 1983. 90-95.

MEAD DATA CENTRAL. 1980. NEXIS, A Primer and NEXIS Quick Reference Manual. New York, NY: Mead Data Central; 1980.

MEL'CUK, IGOR A. 1981. Meaning-Text Models: A Recent Trend in Soviet Linguistics. Annual Review of Anthropology. 1981; 10: 27-62.

MICHELL, GILLIAN, ed. 1982. Papers of the Dictionary Society of North America 1979. Ontario, Canada: The University of Western Ontario School of Library and Information Science; 1982. (See particularly paper by Cameron).

MICHIELS, ARCHIBAL. 1980. Syntactic Manipulation of Frames. In: Dominicy, Mark; Peeters, C., eds. Linguistics in Belgium. Brussels, Belgium: Didier-Hatier; 1980.

MICHIELS, ARCHIBAL. 1981. Exploiting a Large Dictionary Data Base. Liege, Belgium: University of Liege; 1981. (PhD dissertation).

MICHIELS, ARCHIBAL. 1983. Automatic Analysis of Texts. In: Jones, Kevin, ed. Informatics 7. Cambridge, England: Association of Specialized Libraries; 1983.

MICHIELS, ARCHIBAL; MOULIN, A.; NOËL, JACQUES. 1980a. Working with LDOCE. ABLA Papers. 1980; 4: 119-165.

MICHIELS, ARCHIBAL; MULLENDERS, JACQUES; NOËL, JACQUES. 1980b. An English Dictionary and Data Base by Longman as a Step Towards an English-French Data Base for Computer-Aided Translation. Tijdschrift van de Vrije Universiteit Brussel. 1980; XXI(2): 38-59.

MICHIELS, ARCHIBAL; MULLENDERS, JACQUES; NOËL, JACQUES. 1980c. Exploiting a Large Data Base by Longman. In: Proceedings of the 8th International Conference on Computational Linguistics (COLING); 1980; Tokyo, Japan. International Conference on Computational Linguistics/COLING 80; 1980. 374-382. (Also see papers by Kučera; Nagao, Tsujii, Ueda & Takiyama; and Ralph).

MICHIELS, ARCHIBAL; MULLENDERS, JACQUES; NOËL, JACQUES. 1982. The Longman-Liege Project. See reference: Goetschalckx, J.; Rolling, L., eds.

MICHIELS, ARCHIBAL; NOËL, JACQUES. 1982a. Approaches to Thesaurus Production. In: COLING 82: Proceedings of the 9th International Conference on Computational Linguistics; 1982; Prague, Czechoslovakia. 1982. (Also see paper by Calzolari).

MICHIELS, ARCHIBAL, NOËL, JACQUES. 1982b. Retrieving and Improving Collocability Information in 'LDOCE': An Interactive Treatment. Paper presented at Association for Literary and Linguistics Computing (ALLC) Meeting; 1982 June. Pisa, Italy. (Also see paper by Calzolari).

MILLER, GEORGE A. 1979a. Automated Dictionaries. In: Testing, Teaching and Learning: Report of a Conference on Research on Testing. 1978 August 17-26. Washington, DC: U.S. Department of Health, Education, and Welfare, National Institute of Education; 1979. 180-184.

MILLER, GEORGE A., ed. 1979b. Automated Dictionaries, Reading and Writing: Chairman's Report of a Conference on Educational Uses of Word Processors with Dictionaries; 1979 June 14-15. Washington, DC: U. S. Department of Health, Education, and Welfare, National Institute of Education, 1979.

MILLER, LANCE A. 1980. Project *EPISTLE:* A System for the Automatic Analysis of Business Correspondence. In: Proceedings of the 1st Annual National Conference on Artificial Intelligence; 1980; Stanford, CA. Menlo Park, CA 94025: American Association for Artificial Intelligence (AAAI), 445 Burgess Dr., 1980. 280-282.

MILLER, LANCE A.; HEIDORN, GEORGE E.; JENSEN, KAREN. 1981. Text-Critiquing with the *EPISTLE* System: An Author's Aid to Better Syntax. In: Proceedings of the American Federation of Information Processing Societies (AFIPS) National Computer Conference: Volume 50; 1981; Chicago, IL. AFIPS; 1981. 649-655. Also available from: IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598. Research Report RC 8601 (No. 37554), 1980. 15p.

MINSKY, MARVIN, ed. 1968. Semantic Information Processing. Cambridge, MA: M.I.T. Press; 1968.

MISH, FREDERICK C., ed. 1983a. Webster's Ninth New Collegiate Dictionary. Springfield, MA: Merriam-Webster; 1983. 1564p. ISBN: 0-87779-508-8.

MISH, FREDERICK C. 1983b. Comments upon Dictionary Work in Progress at G. & C. Merriam Co. In: Walker, Donald E.; Amsler, Robert A., eds. Proceedings of the 1st Workshop on Machine-Readable Dictionaries; 1983 April 11-13; Menlo Park, CA. Washington, DC: National Science Foundation; 1983. (For publication information contact: D. E. Walker or R. A. Amsler, Bell Communications Research, 435 South St., Morristown, NJ 07960).

MONITOR. 1983. LaserData Datadisc Described. Monitor. 1983 May: 11-12. Available from: Learned Information, Ltd., Oxford, England.

MORRIS, ROBERT; CHERRY, LORINDA L. 1975. Computer Detection of Typographic Errors. IEEE Transactions on Professional Communication. March 1975; PC-18(1): 54-56,64.

MORRIS, WILLIAM. 1974. The Making of a Dictionary—1969. In: Estrin, Herman A.; Mehus, Donald V., eds. The American Language in the 70s. San Francisco, CA 94118: Boyd & Fraser Publishing Company, 3627 Sacramento Street; 1974. ISBN: 0-87835-044-6.

MURRAY, JAMES A. H.; BRADLEY, HENRY; CRAIGIE, W. A.; ONIONS, C. T.; BURCHFIELD, R. W., eds. 1933. Oxford English Dictionary. Oxford University Press; 1933. 13 volumes. ISBN: 0-19-861101-3.

MURRAY, JOHN M.; KLINGENSTEIN, KENNETH J. 1982. The Architecture of an Electronic Book. IEEE Transactions on Industrial Electronics. February 1982; IE-29(1): 82-91.

NADEN, TONY. 1977. Words and Meanings. Tamale, Ghana: Institute of Linguistics; 1977.

NAGAO, M.; TSUJII, UEDA, Y.; TAKIYAMA, M. 1982. An Attempt to Computerize Dictionary Data Bases. See reference: Goetschalckx, J.; Rolling, L., eds. Also available in: COLING 80 conference, 534-542.

NASATIR, MARILYN. 1983. Machine-Readable Data Files and Networks. Information Technology and Libraries. 1983 June: 159-164. ISSN: 0730-9295.

NEAL, PEYTON R., JR.; SLOWINSKI, FRANCIS HILL. 1980. The Changing Law on Proprietary Protection for Computer Databases. In: [Proceedings of the] 4th International Online Information Meeting; 1980 December 9-11; London, England. Oxford, England: Learned Information; 1980. 409-420.

NELSON, THEODOR HOLM. 1981a. Literary Machines: A Report on, and of, Project Xanadu, Concerning Word Processing, Electronic Publishing, Hypertext, Thinkertoys, Tomorrow's Intellectual Revolution and Certain Other Topics Including Knowledge, Education, and Freedom. 1981. Available from: T. H. Nelson, Box 128, Swarthmore PA 19081 for $15.

NELSON, THEODOR HOLM. 1981b. Why Not Have It All? In: [Proceedings of the] VIDEOTEX '81 International Conference & Exhibition; 1981; Toronto, Canada. Northwood Hills, Middlesex, England: Online Conferences, Ltd.; 1981. 309-315.

NIDA, EUGENE ALBERT. 1975a. Componential Analysis of Meaning: An Introduction to Semantic Structures. The Hague, The Netherlands: Mouton; 1975. 272p.

NIDA, EUGENE ALBERT. 1975b. Exploring Semantic Structures. Munich, West Germany: Wilhelm Fink Verlag; 1975. 211p.

NILSEN, DON L. 1973. The Instrumental Case in English: Semantic and Syntactic Considerations. The Hague, The Netherlands: Mouton; 1973.

NIX, ROBERT. 1981. Experience with a Space Efficient Way to Store a Dictionary. Communications of the ACM. 1981 May; 24(5): 297-298. ISSN: 0001-0782.

NOËL, JACQUES. 1979. Syntax, Semantics, and Pragmatics in the Automatic Analysis of Texts. In: New Trends in Documentation and Information. London, England: Association of Specialized Libraries; 1979.

NOMURA, HIROSATO. 1982. Terminology Banks and Dictionaries in Japan and Their Computer Processing. In: Aslib '82: Proceedings of an Aslib Conference; London, England. London, England: Aslib; 1982.

NORLING-CHRISTENSEN, O. 1982. Commercial Lexicography on the Threshold of the Electronic Age. See reference: Goetschalckx, J.; Rolling, L., eds.

OCKER, WOLFGANG A. 1971. A Program to Hyphenate English Words. IEEE Transactions on Engineering Writing and Speech. 1971 June; EWS-14(2): 53-59.

OGDEN, CHARLES KAY. 1934. The System of Basic English. New York, NY: Harcourt, Brace & Co.; 1934.

OGDEN, CHARLES KAY. 1942. The General Basic English Dictionary. New York, NY: W. W. Norton & Co., 1942.

OGDEN, CHARLES KAY; HUGON, PAUL D.; LOCKHART, L. W. 1935. Basic English versus the Artificial Languages. London, England: K. Paul, Trench, Trubner and Co.; 1935.

OLNEY, JOHN C. 1968. To: All Interested in the Merriam-Webster Transcripts and Data Derived from Them. Santa Monica, CA 90406: System Development Corp., 2500 Colorado Ave; 1968 October 1. (Technical Report L-13579).

OLNEY, JOHN C. 1983. Newsletter for Research Uses of the Machine-Readable Transcripts of *Webster's Seventh New Collegiate Dictionary* and *The New Merriam-Webster Pocket Dictionary*, and of Data Sets Derived from them: May 1983. Available from: John Olney, System Development Corp., 2500 Colorado Ave., Santa Monica, CA 90406.

OLNEY, JOHN C.; LAM, VAN; YEARWOOD, BRADLEY. 1976. A New Technique for Detecting Patterns of Term Usage in Text Corpora. Information Processing & Management. 1976; 12: 235-250. ISSN: 0306-4573.

OLNEY, JOHN C.; PARIS, JAMES. 1969. Requests Received as of 18 July 1969 for SDC Lexicographic Data. Santa Monica, CA: System Development Corp. (SDC); 1969. 3p. (Technical Report TM-4369).

OLNEY, JOHN C.; RAMSEY, D. 1972. From Machine-Readable Dictionaries to a Lexicon Tester: Progress, Plans, and an Offer. Computer Studies in the Humanities and Verbal Behavior. 1972 November; 3(4): 213-220.

OLNEY, JOHN C.; REVARD, CARTER C.; ZIFF, PAUL. 1967a. Toward the Development of Computational Aids For Obtaining a Formal Semantic Description of English. Santa Monica, CA: System Development Corp. (SDC); 1967 October. (Paper presented at the International Conference on Computational Linguistics, Grenoble, France, 1967 August 23-25.

OLNEY, JOHN C.; REVARD, CARTER C.; ZIFF, PAUL. 1967b. Summary of Some Computational Aids for Obtaining a Formal Semantic Description of English. In: 2eme Conference Internationale sur le Traitement Automatique des Langues. Grenoble, France; 1967. 1-5.

OLNEY, JOHN C.; SCHONFELD, J.; LAM, VAN. 1977. Computer Processing of the Etymologies in a Standard Dictionary: Preliminary Results. Cahiers de Lexicology. 1977; 31(2): 33-62.

OLSSON, A. HENRY. 1982. Copyright Problems and Use of Computers. In: Lexicography in the Electronic Age. New York, NY: North-Holland Publishing Company; 1982. 239-248. ISBN: 0-444-86404-0.

OSGOOD, CHARLES E.; SUCI, GEORGE J.; TANNENBAUM, PERCY H. 1957. The Measurement of Meaning. Urbana, IL: University of Illinois; 1957.

OSHERSON, DANIEL N.; SMITH, EDWARD E. 1981. On the Adequacy of Prototype Theory as a Theory of Concepts. Cognition. 1981; 9: 35-58.

PACAK, M. G. 1967. Computational Morphology. In: Austin, W., ed. Papers in Linguistics in Honor of L. Dostert. The Hague, The Netherlands: Mouton; 1967.

PACAK, M. G.; COUSINEAU, L; WHITE, W. C. 1973. The Segmentation Approach to Dictionary Construction. Journal of Clinical Computing. 1973; 2(3).

PACAK, M. G.; NORTON, L. M.; DUNHAM, G. S. 1980. Morphosemantic Analysis of -ITIS Forms in Medical English. Methods of Information in Medicine—Methodik der Information in der Medizin. 1980 April; 19(2): 99-105. ISSN: 0026-1270.

PACAK, M. G.; PACAK, A. W. 1971. The Function of Semantics in Automated Language Processing. In: Proceedings of the Symposium on Information Storage and Retrieval. College Park, MD: University of Maryland; 1971. 5-18.

PACAK, M. G.; PRATT, A. W.; WHITE, W. C. 1976. Automated Morphosyntactic Analysis of Medical Language. Information Processing & Management. 1976; 12: 71-76. ISSN: 0306-4573.

PAIKEDAY, THOMAS M. 1982. New Dictionaries: Dr. Johnson Meets the Microchip. Baltimore Sunday Sun. 1982 August 8; Sec. 2: 14-15.

PAIKEDAY, THOMAS M. 1983. The Meaning of the Computer for Dictionary Makers. The Wall Street Journal. 1983 November 8.

PETERSON, JAMES L. 1980a. Computer Programs for Spelling Correction: An Experiment in Program Design. New York, NY: Springer-Verlag; 1980. (Lecture Notes in Computer Science. Volume 96). ISBN: 0-387-10259-0.

PETERSON, JAMES L. 1980b. Computer Programs for Detecting and Correcting Spelling Errors. Communications of the ACM. 1980; 23(12): 676-687. ISSN: 0001-0782.

PETERSON, JAMES L. 1982a. Use of *Webster's Seventh Collegiate Dictionary* to Construct a Master Hyphenation List. In: Proceedings of the American Federation for Information Processing Societies (AFIPS) National Computer Conference; 1982 June 7-10; Houston, TX. Arlington, VA 22209: AFIPS Press, 1815 North Lynn Street; 1982.

PETERSON, JAMES L. 1982b. *Webster's Seventh New Collegiate Dictionary:* A Computer-Readable File Format. Austin, TX: Univ. of Texas, Dept of Computer Sciences; 1982 May. (Technical Report TR-196).

PETERSON, NORMAN D. 1969. Textual Processing, a Computerized Aid for the Writer, Editor, and Reader. IEEE Transactions on Engineering Writing and Speech. 1969 May; EWS-12(1): 9-12.

PETERSON, THOMAS H. 1983. Semantic Structure. Journal of Linguistics. 1983 March; 19(1): 79-114.

PICA, P. 1979. Homography or Polysemy: On a Computer-Aided Analysis of Non-Pronomial Neuter Verbs in French. In: Hartmann, Reinhard R. K., ed. Dictionaries and Their Users. Exeter, England: University of Exeter; 1979.

POLLOCK, JOSEPH J.; ZAMORA, ANTONIO. 1982. Automatic Spelling Error Detection and Correction in Textual Databases. In: Petrarca, Anthony E.; Taylor, Celianna, I.; Kohn, Robert S., eds. Information Interaction: Proceedings of the American Society for Information Science (ASIS) 45th Annual Meeting: Volume 19; 1982 October 17-21; Columbus, OH. White Plains, NY: Knowledge Industry Publications, Inc. for ASIS; 1982. 236-238. ISSN: 0044-7870; ISBN: 0-86729-038-2.

PORTABLE COMPUTER. 1984-. Portable Computer. San Francisco, CA 94105: Miller Freeman Publications, Inc., 500 Howard Street; 1984. ISSN: 0738-1220.

PRATT, A. W.; PACAK, M. G. 1969a. Automatic Processing of Medical English. In: International Conference on Computational Linguistcs; Stockholm, Sweden. Washington, DC: Department of Health, Education and Welfare; 1969.

PRATT, A. W.; PACAK, M. G. 1969b. Identification and Transformation of Terminal Morphemes in Medical English. Methods of Information in Medicine—Methodik der Information in der Medizin. 1969 April; 8(2): 84-90. ISSN: 0026-1270.

PRATT, A. W.; PACAK, M. G.; EPSTEIN, M.; DUNHAM, G. 1973. Computers and Natural Language. Journal of Clinical Computing. 1973 September; 3(2): 85-98.

PRINGLE, ALISON M. 1981. Justification with Fewer Hyphens. The Computer Journal. 1981; 24(4): 320-323. ISSN: 0010-4620.

PROCTER, PAUL, ed. 1978. Longman Dictionary of Contemporary English. London, England: Longman; 1978. 1303p.

QUEMADA, BERNARD. 1972. Lexicology and Lexicography. In: Sebeok, Thomas A., ed. Current Trends in Linguistics. The Hague, The Netherlands: Mouton; 1972.

QUILLIAN, M. R. 1968. Semantic Memory. In: Minsky, Marvin, ed. Semantic Information Processing, Cambridge, MA: M.I.T. Press; 1968. 227-270.

QUILLIAN, M. R. 1969. The Teachable Language Comprehender: A Simulation and Theory of Language. Communications of the ACM. 1969; 12(8): 459-476. ISSN: 0001-0782.

QUIRK, RANDOLPH; SVARTVIK, JAN. 1979. A Corpus of Modern English. In: Bergenholtz, Henning; Schaeder, Burkhard, eds. Empirische Textwissenschaft: Aufbau u. Auswertung von Text-Corpora. Scriptor; 1979. ISBN: 3-589-20663-2.

REICHERT, RICHARD; OLNEY, JOHN C.; PARIS, JAMES. 1969. Two Dictionary Transcripts and Programs for Processing Them: Volume I, The Encoding Scheme, PARSENT and CONIX. Santa Monica, CA: System Development Corp. (SDC); 1969 June 15. NTIS: AD 691 098. (Technical Report TM-3978/001/00).

REVARD, CARTER C. 1967. Affixal Derivation, Zero Derivation, and "Semantic Transformations". Santa Monica, CA: System Development Corp. (SDC); 1967 August 21. 2p. (Technical Report SP-2925).

REVARD, CARTER C. 1968a. Affixal Derivation, Zero Derivation, and 'Semantic Transformations.' Santa Monica, CA: System Development Corp. (SDC); 1968 February 5. 24p. (SDC Document TM-3835).

REVARD, CARTER C. 1968b. Notes on the "Conversion" of Nouns into Verbs in English. Santa Monica, CA: System Development Corp. (SDC); 1968 January 15. 41p. (SDC Document TM-2873).

REVARD, CARTER C. 1969. On the Computability of Certain Monsters in Noah's Ark: Using Computers to Study *Webster's Seventh New Collegiate Dictionary* and *The New Merriam-Webster Pocket Dictionary*. Computer Studies in the Humanities and Vebal Behavior. 1969 August; 2(2): 82-90.

RHODE, MARY. 1974. Compilation of the Technical Lexicon. Los Alamitos, CA: Southwest Regional Laboratory for Educational Research and Development; 1974.

RHYNE, JAMES R. 1976. Lexical Rules and Structures in a Computer Model of Nominal Compounding. Austin, TX: University of Texas at Austin; 1976. (PhD dissertation).

RICH, ELAINE. 1983. Default Reasoning as Likelihood Reasoning. In: Proceedings of the National Conference on Artificial Intelligence. 95 First St., Los Altos, CA 94022: William Kaufmann, Inc.; 1983. 348-351. ISBN: 0-86576-065-9.

RICHARDS, I. A. 1943. Basic English and its Uses. W. W. Norton & Co.; 1943.

RIEGER, CHARLES; SMALL, STEVEN. 1981. Toward a Theory of Distributed Word Expert Natural Language Parsing. IEEE Transactions on Systems, Man, and Cybernetics. 1981 January; SMC-11(1): 43-51. ISSN: 0018-9472.

ROBINSON, DAN. 1983. Electric Webster: Much More than a Spelling Checker. Creative Computing. 1983 November; 9(11): 108, 111-112. ISSN: 0097-8140.

ROGET, PETER MARK. 1852. Thesaurus of English Words and Phrases. London, England: Longman, Brown, Green and Longmans; 1852.

ROWE, JUDITH S. 1981. On-Line Sources of Government-Produced Data: Part I-Bibliographic, Reference and Text Data. Government Publications Review. 1981; 81A: 251-257.

RUBENSTEIN, HERBERT; GOODENOUGH, JOHN B. 1965. Contextual Correlates of Synonmy. Communications of the ACM. October 1965; 8(10): 627-633. ISSN: 0001-0782.

RUSSELL, S. 1975. Computer Understanding of Conceptually Complex Phrases. Stanford, CA: Stanford University; 1975. (Ph.D. dissertation).

SAGER, NAOMI. 1981. Natural Language Information Processing: A Computer Grammar of English and its Applications. Reading, MA: Addison-Wesley; 1981. 399p. ISBN: 0-201-06769-2.

SAGER, NAOMI; HIRSCHMAN, LYNETTE; WHITE, CAROLYN; FOSTER, CAROL; WOLFF, SUSANNE; GRAD, ROBERT; FITZPATRICK, EILEEN. 1980. Research into Methods for Automatic Classification and Fact Retrieval in Science Subfields: String Project Report 13. 1980. (Final Report). Available from: New York University Linguistic String Project, 251 Mercer St., New York, NY 10012.

SALTON, G.; LESK, M. E. 1971. Information Analysis and Dictionary Construction. In: Salton, G., ed. The *SMART* System: Experiments in Automatic Document Processing. Englewood Cliffs, NJ: Prentice-Hall; 1971.

SANDBERG-DIMENT, ERIK. 1984. Computer Column. New York Times. 1984 January 3. ISSN: 0362-4331.

SCHANK, ROGER C. 1973. The Fourteen Primitive Actions and Their Inferences. Stanford, CA: Stanford University, AI Laboratory, 1973. (Artificial Intelligence Memo 183).

SCHANK, ROGER C.; NASH-WEBBER, BONNIE L., eds. 1975. Theoretical Issues in Natural Language Processing. New York, NY: Association for Computational Linguistics; 1975. 219p. (See particularly papers by Becker and by Miller).

SCHNEIDER, B. R. 1971. The Prod. ction of Machine-Readable Text: Some of the Variables. Computers and the Humanities. 1971; 6: 39-37. ISSN: 0010-4817.

SCHOENE, WILLIAM J. 1969. A User's Guide to *QUICKY*. Santa Monica, CA: System Development Corp. (SDC), 2500 Colorado Ave.; 1969 June 15. 13p. (SDC Technical Report TM-4297/000/00).

SCHWAB, WALLACE; ST. DENIS, RICHARD. 1980. A Data Bank Experience on the *PLATO* System. Journal of Computer-Based Instruction. 1980; VII(2): 41-46.

SEARLE, JOHN. 1969. Speech Acts. London, England: Cambridge University Press; 1969.

SEDELOW, SALLY YEATES. 1974. Automated Language Analysis 1973-74. Lawrence, KS: University of Kansas, Departments of Computer Science and Linguistics; 1974.

SEDELOW, WALTER A., JR.; SEDELOW, SALLY YEATES. 1978. Formalized Historiography: The Structure of Scientific and Literary Texts: Part 1. Some Issues Posed by Computational Methodology. Journal of the History of the Behavioral Sciences. 1978; 14: 247-263.

SEDELOW, WALTER A., JR.; SEDELOW, SALLY YEATES, eds. 1979a. Computers in Language Research: Formal Methods. New York, NY: Mouton; 1979. 223p. (Trends in Linguistics: Studies and Monographs: Volume 5).

SEDELOW, WALTER A., JR.; SEDELOW, SALLY YEATES. 1979b. The History of Science as Discourse, the Structure of Scientific and Literary Texts: Part II. Some Issues Posed by Computational Methodology. Journal of the History of the Behavioral Sciences. 1979. 15: 63-72.

SELKIRK, ELISABETH O. 1982. The Syntax of Words. Cambridge, MA: M.I.T. Press; 1982. ISBN: 0-262-19210-1 and 0-262-69079-9 (pbk.).

SEROY, JEFFREY. 1984. Oxford University Press Announces Plans to Computerize the *Oxford English Dictionary*. 1984 May 15. (Press Release). Available from: Oxford University Press, 200 Madison Avenue, New York, NY 10016.

SHERMAN, DONALD. 1974a. A Common Structure for Lexicographic Data. In: Mitchell, John L., ed. [Proceedings of an International Conference on] Computers in the Humanities; 1973; Minneapolis, MN. Edinburgh, Scotland: University of Edinburgh Press; 1974. (Also see paper by Berry-Roche).

SHERMAN, DONALD. 1974b. A New Computer Format for *Webster's Seventh Collegiate Dictionary*. Computers and the Humanities. 1974; 8: 21-26. ISSN: 0010-4817.

SHERMAN, DONALD. 1976. Computer Assistance for the 'Dictionary of Early Modern English Pronunciation' (DEMEP). Stockholm, Sweden: Almqvist and Wiksell International; 1976.

SHERMAN, DONALD. 1978. Special Purpose Dictionaries. Cahiers de Lexicologie. 1978; 32(1): 82-102.

SHERMAN, DONALD. 1979a. Retrieving Lexicographic Citations from a Computer Archive of Language Materials. In: Hartmann, Reinhard R. K., ed. Dictionaries and Their Users. Exeter, England: University of Exeter; 1979.

SHERMAN, DONALD. 1979b. A Semantic Index to Verb Definitions in *Webster's Seventh New Collegiate Dictionary*. Stanford, CA: Stanford University Linguistics Department; 1979. (Computer Archive of Language Materials).

SHERMAN, DONALD. 1982. A Computer Archive of Machine Readable Dictionaries. In: Michell, Gillian, ed. Papers of the Dictionary Society of North America, 1979. Ontario, Canada: The University of Western Ontario School of Library and Information Sciences; 1982.

SHORTLIFFE, E. H. 1976. Computer-Based Medical Consultations: *MYCIN*. New York, NY: American Elsevier; 1976.

SIBUYA, M.; FUJISAKI, T.; TAKAO, Y. 1978. Noun-Phrase Model and Natural Query Language. IBM Journal of Research and Development. 1978 September; 22(5): 533-540. ISSN: 0018-8646.

SIDNER, CANDICE L. 1978. The Use of Focus as a Tool for Disambiguation of Definite Noun Phrases. In: TINLAP-2: Theoretical Issues in Natural Language Processing-2: Proceedings of an Association for Computing Machinery (ACM) SIGART (Special Interest Group in Artificial Intelligence) Conference 1978 July 25-27; Urbana, IL. 1978. 86-95. Available from: ACM.

SIDNER, JAMES G.; OSGOOD, CHARLES E., eds. 1969. Semantic Differential Technique: A Sourcebook. Chicago, IL: Aldine Publishing Co.; 1969. 681p. LC: 68-19874.

SIMMONS, ROBERT F.; AMSLER, ROBERT A. 1975. Modeling Dictionary Data. In: Grishman, Ralph, ed. Directions in Artificial Intelligence, Natural Language Processing. New York, NY: Courant Institute of Mathematical Sciences and New York University Computer Science Department; 1975. (Computer Science Report 7).

SMALL, STEVEN. 1980. Word Expert Parsing: A Theory of Distributed Word-Based Natural Language Understanding. University of Maryland, Department of Computer Science, 1980. 268p. (PhD dissertation). Also available as Technical Report TR-954/NSG-7253, 1980 September.

SMITH, JOAN M. 1977. Computers in Lexicography. In: van Sterkenburg, P. G. J., ed. Lexicologie: een bundel opstellen voor F. de Tollenaere ter gelegenheid van vakgenoten. Groningen, The Netherlands: Wolters-Noordhoff; 1977. 375p.
SMITH, P. D. 1983. *XENO:* Computer-Assisted Compilation of Crossword Puzzles. The Computer Journal. 1983 November; 26(4): 296–302. ISSN: 0010-4620.
SMITH, P. D.; STEEN, S. Y. 1981. A Prototype Crossword Compiler. The Computer Journal. 1981; 24(2): 107–111. ISSN: 0010-4620.
SMITH, RAOUL N. 1972. Interactive Lexical Updating. Computers and the Humanities. 1972 January; 6(3): 137–145. ISSN: 0010-4817.
SMITH, RAOUL N. 1980–1981. On Defining Adjectives: Part III. Dictionaries. 1980–1981; 3: 28–38.
SMITH, RAOUL N. 1981. Computerized Aids to Writing. Paper presented at: 3rd Delaware Symposium on Language Studies: Linguistics and Literacy, Special Session on Computers and Literacy; 1981 October 15–17; University of Delaware. 1981. (GTE Laboratories, Inc. Technical Report).
SOLSO, ROBERT L.; JUEL, CONNIE; RUBIN, DAVID C. 1982. The Frequency and Versatility of Initial and Terminal Letters in English Words. Journal of Verbal Learning and Verbal Behavior. 1982; 21: 220–235.
SOWA, JOHN F. 1983. Conceptual Structures: Information Processing in Mind and Machine. New York, NY: Addison-Wesley; 1983. ISBN: 0-201-14472-7.
SPARCK JONES, KAREN. 1967. Dictionary Circles. Santa Monica, CA 90406: System Development Corp. (SDC), 2500 Colorado Ave. 1967 January. (SDC Technical Report TM-3304).
SPARCK JONES, KAREN. 1983. So What About Parsing Compound Nouns? In: Sparck, Jones, Karen; Wilks, Yorick, ed. Automatic Natural Language Parsing. Chichester, West Sussex, PO19 1EB, England: Ellis Horwood Limited, Cooper Street; 1983. 164–168. ISBN: 0-85312-621-6.
STAPLES, BETSY. 1983. Words, Words, Words: Six Games to Challenge the Verbally Inclined. Creative Computing. September 1983; 9(9): 165, 167–168, 171–172, 174. ISSN: 0097-8140.
STEIN, JESS, ed. 1966. The Random House Dictionary of the English Language. New York, NY: Random House, Inc.; 1966.
STEIN, JESS; SU, P. Y., eds. 1978. The Random House Dictionary. New York, NY: Ballantine Books; 1978. 1070p. ISBN: 0-345-27723-6.
STONE, PHILIP J.; DUNPHY, DEXTER C.; SMITH, MARSHALL S.; OGILVIE, DANIEL M. 1966. The General Inquirer: A Computer Approach to Content Analysis. Cambridge, MA: M.I.T. Press; 1966. 651p. LC: 66-22641.
SUGARMAN, JOSEPH, ed. 1983. Artificial Intelligence: 'Products That Think'. Northbrook, IL 60062: JS&A Group Inc., One JS&A Plaza, 1983. (Catalog).
SYSTEM DEVELOPMENT CORPORATION (SDC). 1966. SDC Proposal 818: Language Processing Research Program. Technical Report, Santa Monica, CA 90406: SDC, 2500 Colorado Ave.; 1966 April.
SZANSER, A. J. 1970. Automatic Error-Correction in Natural Languages. Information Storage and Retrieval. 1970 February; 5(4): 169–174. ISSN: 0020-0271.
TOWNLEY, HELEN M.; GEE, RALPH D. 1980. Thesaurus-Making: Grow Your Own Word-Stock. Andre Deutsch; 1980. 206p. ISBN: 0-233-97225-0.

TRAVIS, IRENE L.; FIDEL, RAYA. 1982. Subject Analysis. In: Williams, Martha E., ed. Annual Review of Information Science and Technology: Volume 15. White Plains, NY: Knowledge Industry Publications, Inc. for the American Society for Information Science; 1982. ISSN: 0066-4200; CODEN: ARISBC.
TUCKER, ALLEN B.; NIRENBURG, SERGEI. 1984. Machine Translation In: Williams, Martha E., ed. Annual Review of Information Science and Technology: Volume 19. White Plains, NY: Knowledge Industry Publications, Inc. for the American Society for Information Science; 1984. ISSN: 0066-4200; CODEN: ARISBc.
TWADDELL, W. F. 1954. A Linguistic Archive as an Indexed Depot. International Journal of American Linguistics (IJAL). 1954; 20: 108–110.
U.S. NATIONAL SCIENCE FOUNDATION (NSF). 1984. Proceedings of the Workshop on Machine-Readable Dictionaries; 1983 April 11–13; SRI International, Menlo Park, CA. Washington, DC: NSF; 1984 (in press). (See particularly papers by Evens; Procter; Simmons; and White). For publication information contact D. E. Walker, Bell Communications Research, 445 South St., Morristown, NJ 07960.
URDANG, LAURENCE. 1966a. The Systems Designs and Devices Used to Process the *Random House Dictionary of the English Language.* Computers and the Humanities. 1966; 1: 31–33. ISSN: 0010-4817.
URDANG, LAURENCE. 1966b. An "Unabridged" Word-Frequency Count of American English: A Proposal for an Integrated System. Word. 1966 April-August-December; 22(1-2-3): 294–302.
URDANG, LAURENCE. 1979. The Human Use of Human Indexers. The Indexer. 1979 April; 11(3): 125–131. ISSN: 0019-4131.
URDANG, LAURENCE. 1984. A Lexicographer's Adventures in Computing. Datamation. 1984 March; 30(3): 185–194. ISSN: 0011-6963.
VENEZKY, RICHARD L. 1969. Storage, Retrieval, and Editing of Information for a Dictionary. American Documentation. 1969; 19: 71–79. Also available from: The University of Wisconsin, Computer Sciences Department, as Technical Report No. 7; 1967 June.
VENEZKY, RICHARD L. 1973a. Computational Aids to Dictionary Compilaton. In: Frank, Roberta; Cameron, Angus Fraser, eds. A Plan for the Dictionary of Old English. Toronto, Canada: University of Toronto Press; 1973. 347p.
VENEZKY, RICHARD L. 1973b. Computer Applications in Lexicography. In: McDavid, Raven I., Jr.; Ducker, Audrey R., eds. Lexicography in English. New York, NY. New York Academy of Sciences; 1973. (Also see papers by Bailey; Barnhart; Chapman; Lehmann; and Urdang).
VENEZKY, RICHARD L. 1979. Orthographic Regularities in English Words. In: Kolers, P. A.; Wrolstad, M. E.; Bouma, H., eds. Processing of Visible Language: Volume 1. New York, NY: Plenum Press; 1979. 283–293.
VENEZKY, RICHARD L.; RELLES, NATHAN; PRICE, LYNNE. 1976a. LEXICO Guides. Madison, WI: University of Wisconsin Computer Sciences Department; 1976.
VENEZKY, RICHARD L.; RELLES, NATHAN; PRICE, LYNNE. 1977a. LEXICO: A System for Lexicographic Processing. Computers and the Humanities. 1977: 127–137. ISSN: 0010-4817.
VENEZKY, RICHARD L.; RELLES, NATHAN; PRICE, LYNNE. 1977b. User Aids in a Lexical Processing System. In: Lusignan, Serge; North, John S., eds. Proceedings of the 3rd International Conference on Computing in the Humanities. University of Waterloo Press; 1977. 365p. (Also see paper by Leavitt & Mitchell).

VENEZKY, RICHARD L.; RELLES, NATHAN; PRICE, LYNNE; BENNETT, JACK. 1976b. Man-Machine Integration in a Lexical Processing System. Computer Sciences Technical Report 280, Madison, WI: University of Wisconsin, Computer Sciences Department; 1976 September.
WAIBEL, A. 1982. Towards Very Large Vocabulary Word Recognition. Pittsburgh, PA: Carnegie-Mellon University, Department of Computer Science; 1982 November 30. 33p. (Technical Report CMU-CS-82-144).
WALKER, DONALD E.; AMSLER, ROBERT A. 1984. The Use of Machine-Readable Dictionaries in Sublanguage Analysis. In: Kittredge, Richard I., ed. [Proceedings of] Workshop on Sublanguage Analysis; 1984; New York, NY. Available from: the authors: at Bell Communications Research, 435 South St., Morristown, NJ 07960.
WALL STREET JOURNAL. 1983. Scrabble Electronic Toy. Wall Street Journal. 1983 January 20: 18.
WAY, T. J. 1981. Studies in Text Processing. IBM Research Highlights. Yorktown Heights, NY: IBM Thomas J. Watson Research Center; 1981 October. Available from: Research Communications, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598.
WERNER, OSWALD. 1977. The Synthetic Informant Model on the Simulation of Large Lexical/Semantic Fields. In: McCormack, William Charles; Wurm, Stefan A., eds. Language and Thought: Anthropological Issues: Papers Prepared for the 9th International Congress of Anthropological and Ethnological Sciences; 1973 Chicago. The Hague, 1977. The Netherlands: Mouton. 525p.
WEST, MICHAEL. 1953. A General Service List of English Words with Semantic Frequencies. London, England: Longmans, Green and Co., Ltd.; 1953.
WEYER, STEPHEN A. 1982. Searching for Information in a Dynamic Book. Technical Report SCG-82-1. Palo Alto, CA: Xerox Palo Alto Research Center; 1982 February. 206p. (Ph.D. dissertation from Stanford University, School of Education).
WIERZBICKA, ANNA. 1972. Semantic Primitives. Frankfurt, West Germany: Athenaeum Verlag; 1972. ISBN: 3-7610-4822-X.
WIGGINS, D. 1971. On Sentence-Sense, Word-Sense, and Difference of Word-Sense. Towards a Philosophical Theory of Dictionaries. In: Steinberg, Danny D.; Jakobovitz, Leon A., eds. Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology. Cambridge, England: Cambridge University Press; 1971. 603p.
WIKBERG, KAY. 1983. Methods in Contrastive Lexicology. Applied Linguistics. 1983 Autumn; 4(3): 213–221.
WILKS, YORICK A. 1977. Methodological Questions about Artificial Intelligence: Approaches to Understanding Natural Language. Journal of Pragmatics. 1977; 1: 69–84.
WILLIAMS, GREGG; MEYER, RICK. 1981. The Panasonic and Quasar Hand-Held Computers: Beginning a New Generation of Consumer Computers. BYTE. 1981 January; 6(1): 34–45. ISSN: 0360-5280.
WILLIAMS, P. W.; WOODHEAD, D. 1979. Computer Assisted Analysis of Cryptic Crosswords. The Computer Journal. 1979; 22(1): 67–70. ISSN: 0010-4620.
WINER, DAVID; WINER, PETER. 1984. Portables—1984 and Beyond: Idea-Processing Software and Portable Computers; The New Wave of Computers Demands New Software. BYTE. 1984 January; 9(1): 243–246. ISSN: 0360-5280.

WINOGRAD, TERRY. 1971. An A. I. Approach to English Morphemic Analysis. Cambridge, MA: Massachusetts Institute of Technology (MIT) Artificial Intelligence Laboratory; 1971. (MIT AI Memo 21).
WINOGRAD, TERRY. 1983. Language as a Cognitive Process: Volume 1, Syntax. Reading MA: Addison-Wesley; 1983. 640p. ISBN: 0-201-08571-2 (v. 1).
WISBEY, ROY. 1971. Publications from an Archive of Computer-Readable Literary Texts. In: Wisbey, Roy, ed. The Computer in Literary and Linguistic Research. Cambridge, England: Cambridge University Press; 1971.
WOLFART, H. CHRISTOPH. 1979. Lexical Documentation. Lingua. 1979; 49(2–3): 137–152.
WSZOLA, STANLEY J. 1983. How to Choose a Portable: Factors to Consider before You Take the Plunge. BYTE. 1983 September; 8(9): 34–47. ISSN: 0360-5280.
ZAMORA, ANTONIO. 1980. Automatic Detection and Correction of Spelling Errors in a Large Data Base. Journal of the American Society for Information Science. 1980 January; 31(1): 51–59. ISSN: 0002-8231.
ZAMPOLLI, ANTONIO; CALZOLARI, NICOLETTA, eds. 1977. Computational and Mathematical Linguistics: Proceedings of the International Conference on Computational Linguistics 1973; Pisa, Italy. Firenze, Italy: L. S. Olschki; 1977. 2 volumes (719p). (Series: Biblioteca dell' Archiuum Romanicum Ser. II, Linguistica; 36; See particularly papers by Calzolari, Pecchia & Zampolli; Paikeday; Smith & Maxwell; Stachowitz; and Wilton).
ZGUSTA, LADISLAV. 1971. Manual of Lexicography. The Hague, The Netherlands: Mouton, 1971.
ZGUSTA, LADISLAV, ed. 1980. Theory and Method in Lexicography. Columbia, SC: Hornbeam Press; 1980. 189p. (See particularly papers by Cassidy and by Makkai).
ZIFF, P. 1966. More on Understanding Understanding Utterances. Santa Monica, CA 90406: System Development Corp. (SDC), 2500 Colorado Avenue; 1966 June 8. 16p. (SDC Report SP-2504).