AD-A008 459

USER'S GUIDE TO THE SOLAR BIBLIOGRAPHY FILE

Timothy Diller

System Development Corporation

Prepared for:

Advanced Research Projects Agency

30 December 1974

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

TM-5292/000/02

USER'S GUIDE TO THE SOLAR BIBLIOGRAPHY FILE

30 DECEMBER 1974

DR. TIMOTHY DILLER

THIS RESEARCH IS SUPPORTED BY THE ADVANCED RESEARCH PROJECTS AGENCY OF THE DEPARTMENT OF DEFENSE UNDER CONTRACT NO. DAHC15-73-C-0080.

THE VIEWS AND CONCLUSIONS CONTAINED HEREIN ARE THOSE OF THE AUTHOR AND SHOULD NOT BE INTERPRETED AS NECESSARILY REPRESENTING THE OFFICIAL POLICIES, EITHER EXPRESSED OR IMPLIED, OF THE ADVANCED RESEARCH PROJECTS AGENCY OF THE U. S. GOVERNMENT.

SYSTEM DEVELOPMENT CORPORATION
2500 COLORADO AVENUE • SANTA MONICA, CALIF. 90406

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) User's Guide to the SOLAR Bibliography File		5. TYPE OF REPORT & PERIOD COVERED special technical
		6. PERFORMING ORG. REPORT NUMBER TM-5292/000/02
7. AUTHOR(s) Dr. Timothy Diller		8. CONTRACT OR GRANT NUMBER(s) DAHC15-73-C-0080
9. PERFORMING ORGANIZATION NAME AND ADDRESS System Development Corporation '00 Colorado Avenue Santa Monica, California 90406		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ARPA Order 2254 Program Code 5D30
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 30 December 1974
		13. NUMBER OF PAGES 28
14. MONITORING AC_ICY NAME & ADDRESS(if different from Controlling Office) ARPA/Information Processing Techniques 1400 Wilson Boulevard Arlington, Virginia 22209		15. SECURITY CLASS. (of this report) 0
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

English semantics
speech understanding research

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This document contains a general explanation of the bibliographic citation file of SOLAR (a Semantically Oriented Lexical ARchive). It is intended as an introduction and reference manual for the on-line user, the casual reader, or the data collector. The document indicates the design concepts, retrieval procedures, and data collection procedures. A complete list of SOLAR documentation is given in the introduction to this document. This document is a reissue of TM-5292/000/01 and supersedes that document.

DD FORM 1 JAN 7- 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

This document contains a general explanation of the bibliographic citation file of SOLAR (a Semantically-Oriented Lexical Archive). It is intended as an introduction and reference manual for the on-line user, the casual reader, or the data collector. The document indicates the design concepts, the resulting file structure, the intended file content, retrieval procedures, and data collection procedures. A complete list of SOLAR documentation is given in the introduction to this document. This document is a reissue of TM-5292/000/01 and supersedes that document.

FOREWORD

This document is one of a series provided by System Development Corporation as a guide to the SOLAR system. Users are encouraged to advise us by phone or in writing of errors, ambiguities, or other deficiencies and difficulties arising in the use of this document and/or the SOLAR system. Communicate with:

Dr. Timothy C. Diller
System Development Corporation
2500 Colorado Avenue
Santa Monica, Cal. 90406
Phone: 213-829-7511

Table of Contents

1. INTRODUCTION

SOLAR OVERVIEW

This section serves as a common preface to each of the user's guides describing the SOLAR files. It outlines the goals of SOLAR and the relationship of each file to those goals. It ends with a list of the documents describing SOLAR.

SOLAR is intended to provide easy access to a large variety of semantic data pertaining to a selected set of English words. Data have been collected to date on about 2,000 SUR words, i.e., words found in the lexicons of the Speech Understanding Research groups being sponsored by ARPA.(1) Each of the eight principal SOLAR files contains semantic data of a different type. Two supplementary files facilitate use of the archive: a word index and a bibliography.(2)

(1) The file of semantic analyses consists of formal descriptions of word meanings, primarily those descriptions given in papers written by linguists, philosophers, and computer scientists. Whatever information the author's presents on such topics as predicate-argument relations, semantic components, presuppositions, and/or entailments is abstracted. In addition, qualifications and informal explanations by the author are included as are criticisms of his description by other writers and/or by us.

---

(1)Although the words for which data is currently being collected all come from the lexicons being used by the SUR projects at Carnegie-Mellon University, Bolt Beranek and Newman, and System Development Corporation, we are willing to extract and collect data on other word sets also.

(2)I wish to acknowledge John Olney's contributions to the archive he was largely responsible for the original design of SOLAR as set forth in Miller and Olney (1973) and continues to be responsible for the preparation of integrative summaries of conceptual analyses.

(2) A second file provides a concise digest of the theoretical background of each semantic analysis. The author's theoretical orientation, his assumptions, and his notational conventions are discussed.

(3) Explanatory notes for the semantic components used in the semantic analyses are entered into a third file. These notes explain as precisely as possible the conceptual content each author evidently intends his component(s) to have. Included in the file are any comments on the author's use of components that the SOLAR builders have deemed appropriate.

(4) A file of conceptual analyses contains integrative summaries of the best analyses found in the recent literature of analytic philosophy and artificial intelligence for particular notions, primarily those coinciding with or underlying the semantic components entered in the third file.

(5) A collocational feature file contains, for SUR words, the definitions from Webster's Seventh New Collegiate Dictionary (W7) in which a subject label, a parenthetic phrase, a usage note, or a verbal illustration appears. Each of these elements supplies some indication of the words or word classes permissible in the immediate context of a given SUR word.

(6) A semantic field file[3] will [illegible] a series of displays showing most of the other words in the English vocabulary that stand in a morphological, definitional, synonymitive, antonymitive, or thesaural

---

(3)The structure of this file and procedures for creating it have been specified in detail; however, coding has not yet begun on the several programs needed.

relationship to a given word. Such relationships will be machine derived from the W7 transcripts, a partial transcript of Webster's New Dictionary of Synonyms, and a thesaurus transcript (hopefully the transcript of Roget's International Thesaurus being prepared by Sally Sedelow at the University of Kansas).

(7) A file of definitional expansions (4) will indicate the extent and nature of the semantic connectedness among words in a particular lexicon. For each word in a given lexicon, a display will be provided of all the words in that lexicon that can be reached by following W7 definitional links outward to two levels of remoteness from that word.

(8) A key-word-in-context ("KWIC") file(5) will contain all the contexts of a given word's occurrences in the million-word Brown Corpus, the 1.2 million-word corpus of W7 definitions, and dialogues collected by the speech understanding groups.

The first of the supplementary files is a word index, which lists all the words appearing in the speech understanding lexicons, the lexicons they appear in, the parts of speech given for each word in the lexicon together with their corresponding parts of speech in W7, and the types of SOLAR data available for each word.

A bibliography file provides citations to the technical documents in linguistics, philosophy, and computer science that are referenced in other SOLAR files or may be of interest to researchers in natural language processing.

---

(4)Although this file has not yet been produced, its structure has been specified and coding of the programs needed to build it has begun.

(5)This file has been created in part, the Brown Corpus contexts having already been entered.

SOLAR DOCUMENTATION

Archive Overviews

1. Diller, T., & J. Olney. (1973) "SOLAR (A Semantically-Oriented Lexical Archive)" SDC Document SP-3726/000/00

2. Diller, T., & J. Olney. (1974) "SOLAR (A Semantically-Oriented Lexical Archive): Current Status and Plans" Computers and the Humanities 8:301-312.

3. Diller, T. & J. Olney. (forthcoming) "SOLAR: A Comprehensive Source of Semantic Lexical Data" American Journal of Computational Linguistics.

User's Guides

4. Bye, T., T. Diller, & J. Olney. (1975) "User's Guide to the SOLAR Semantic Analysis File" SDC Document TM-5292/001/00

5. Diller, T. (1974) "User's Guide to the SOLAR Bibliography File" SDC Document TM-5292/000/02

6. Diller, T. (in prep.) "User's Guide to the SOLAR Word Index" SDC Document TM-5292/009/00

7. Diller, T., & T. Bye. (in prep.) "User's Guide to the SOLAR Theoretical Backgrounds File" SDC Document TM-5292/002/00

8. Diller, T., T. Bye & J. Olney. (in prep.) "User's Guide to the SOLAR Semantic Component File" SDC Document TM-5292/003/00

9. Diller, T., & F. Heath. (in prep.) "User's Guide to the SOLAR KWIC File" SDC Document TM-5292/008/00

10. Diller, T., & F. Heath. (in prep.) "User's Guide to the SOLAR Collocational Feature File" SDC Document TM-5292/005/00

11. Diller, T., F. Heath, & J. Olney. (in prep.) "User's Guide to the SOLAR Semantic Field File" SDC Document TM-5292/006/00

12. Heath, F., T. Diller, & J. Olney. (in prep.) "User's Guide to the SOLAR Definitional Expansion File" SDC Document TM-5292/007/00

13. Olney, J., F. Delacruz, T. Diller, & N. Uenzoglu. (in prep.) "User's Guide to the SOLAR Conceptual Analysis File" SDC Document TM-5292/004/00

2. FILE DESIGN

The creation of a file of bibliographic citations was a natural corollary activity to the construction of the semantic and conceptual analyses files. As source documents for such analyses were discovered, it was obvious that citation reference within the archive could be simplified by the addition of a citation file. It was also evident that such a file would be useful to both SOLAR staff members and archive users when searching for documents relevant to their current areas of investigation. Bibliographic services such as Language and Language Behavior Abstracts and Language Teaching Abstracts, while perhaps more comprehensive in their source scanning, are deficient in retrieval capabilities, being extremely limited in the search terms they provide. Other services, such as the Modern Language Association's International Bibliography and Bibliographie linguistique, have an additional drawback in the significant time lag existing between review and publication.

Stimulated by these motivating factors, we designed the file of bibliographic citations with the following two criteria uppermost in mind.

First, the file should be of practical assistance to researchers engaged in modeling the understanding of English by computers. Hence, citations center on those documents in linguistics and the philosophy of language that are relevant to the computational processing of natural language. The areas of acoustic phonetics and lexical semantics have received primary attention because of the ties of SOLAR

to automated speech understanding research.

Second, the file should be directly and easily accessible to researchers. The time spent by them in learning the file structure and data management protocols should be minimal. The file accordingly resides in the user-oriented SOLAR data management system, which is accessible via the ARPA Network.

Since the file is to be accessed by researchers with varied interests, we have broadened its appeal and potential usefulness by augmenting its size to nearly 5,000 entries, primarily by obtaining machine-readable bibliographic transcripts from a number of different sources (cf. section 'CITATION ENTRY' for a detailed list of such acquisitions).

Considerable care has been taken in designing this file to (a) allow for a wide range of document types to be entered and (b) allow retrieval on practically every type of information for in a reference while (c) maintaining a simplicity in data entry and (d) a flexibility in composed output. The file accommodates references to seven types of documents: articles published in journals, articles published in books unpublished papers, books, reviews, reports, and journals. Thirty different fields have been set up to accept the variety of information to be entered for these seven document types. They are listed briefly next with an indication of their logical characteristics.

Information about the author of a document is found in two fields. The first contains the authors' names as they would appear in a bibliography. The second, having a repeating group structure, contains each author's name individually (joint authors being separated) and is used for retrieval in an author index. The title, journal, volume and

issue numbers, and publisher, together with the year, month, and day of publication, are all present as separate fields. A field called 'PAGES' exists to indicate the pages covered by an article in its publication source.

For articles published in a book, an editor can be specified as well as the book title. For reports, a sponsoring agency can be entered, together with a report number. For books, one can enter the total pages, cost, Library of Congress number and/or ISBN number, and where the book has been reviewed.

Each entry can be given as many key words as desired. An abstract can be entered, and one can also enter notes. One or more locations of copies of the document can also be specified. The date the entry is made can be recorded, as well as who created the entry. Reference to individual citations is simplified by the existence of a unique sequence number on each citation.

3. DEFINITION OF FIELDS

There are currently 30 fields into which data are entered when a bibliographic citation is built. Each is discussed below in turn in regard to the type and format of the data to be entered. The number following the heading matches that found on the citation entry worksheets.

AUTHOR: 1)

Each entry is provided automatically with a bibliographic-style author name based on the data entered in the next field. The following formats, together with their parallels containing (Ed.) or (Eds.), are generated:

Surname, Firstname
Surname, Firstname (, Firstname Surname,) and Firstname Surname
Surname, Firstname, et al.

E.g.,

Chomsky, N. and J. Ross
Lakoff, G., et al.

AUTHOR: 102) 2)

Every entry must contain an author name. A single author's name takes the format: Surname, Firstname. If the entire first name is known, it only should be used and a middle name or middle initial should not appear. If only initials are known, as many as are known should be entered with periods after each one. If the author is not known, the value 'Anonymous' should be entered. If there is no author but instead an editor, the names are entered in the same format, but with (Ed.) after at least one of the editors. Each author is a separate 102) 2)

entry.

TITLE: 3)

It a book is being cited, the title appears without underlining or quotes. For all other document types, the title is enclosed in double quotes, i.e., "...".

JOURNAL: 4)

No format is specified for journal names; however, full names are preferred rather than abbreviations. Obscure journals are identified by the place of publication following the journal title.

PUBLISHER: 5)

The publisher's name, followed by a period, precedes the place of publication. The place is optional when the publisher's name contains the location or is sufficiently well-known; e.g.,

Prentice-Hall. Englewood Cliffs, New Jersey
MIT Press.

INSTIT: 5)

For unpublished papers and reports, the institutional affiliation of the author (at the time of writing) is entered, followed by the institution's location (if not obvious). For talks and lectures, the INSTIT field indicates where the lecture was delivered. If INSTIT is a university, abbreviations are acceptable without periods; e.g.,

RAND. Santa Monica, Calif.
UCLA

YEAR (of publication): 6)

A value is obligatory for this field. The value must be either a four-digit number (e.g., 1973) or, if the date is partially unknown, a combination of digits and X's (e.g., 197X or 19XX).

MO (of publication): 7)

The value must be one of the following: JAN, FEB, MAR, APR, MAY, JUN, JUL, AUG, SEP, OCT, NOV, DEC.

DAY (of publication): 8)

The value must be a one- or two-digit number.

VOL.NBR: 9)

There are no format restrictions on how volume and issue number appear; however, the following format is suggested: vol.nbr (e.g., 6.2). Arabic numerals are preferred over Roman for ease in reading. If the issue number is unknown, the volume alone (without a trailing period) is entered. The issue number is not entered if the volume is unknown.

PAGES: 10)

This field indicates the beginning and ending pages of the document. Both numbers must be entered, and their full value is preferred. Thus, if the document begins on page 301 and ends on page 309, the entry should read 'pp. 301-309', not 'pp. 301-9'. If an article is only a single page in length, that page number appears in both blanks.

EDITOR: 11)

A single editor is entered in the following format:

Surname, Firstname (Ed.)

If there are co-editors, the names are as follows:

Surname, Firstname, and Firstname Surname (Eds.)

Translators, transcribers, compilers, and annotators also appear here with appropriate indication in parentheses; e.g.,

Kastovsky, Dieter (Ed.)
Lehmann, W., and R. Stachowitz (Eds.)

SPO. OR: 12)

The name of the agency sponsoring the production of the document is entered here. No particular format is required.

ABSTRACT: 113) 13)

Concise annotations of documents are entered in normal textual format. The maximum allowable length of text between the ABSTRACT field numbers is 256 characters, and the lines provided on the data collection sheets are intended to keep the values below this threshold.

NOTES: 114) 14)

Noteworthy information includes such things as earlier publication, reprinting, place and date of pr ntation, source of material, translation facts, revision details, conference or symposium that the document records, availability and source (if unpublished or a report), and plans for publication. The same restriction on length (256 characters) exists here as for abstracts.

KEYWORDS: 115) 15)

Terms characterizing the document that will be useful in citation retrieval may be either single or multiple word terms. Such search terms are entered one per field number. Determination of the search terms to be entered for a particular citation is accomplished in two ways. For documents obtained from machine-readable collections, the keywords accompanying the citation (if any) are accepted as is and entered in this field. For these and the hand-collected citations, SOLAR staff can add additional keywords as deemed appropriate. The SOLAR staff periodically monitors the keywords and standardizes them to ensure an acceptable degree of consistency.

When a document analyzes a particular lexical item, that word is enclosed in single quotes, e.g., 'happy'. Semantic primitives and phones are enclosed in square brackets upon printout but, because of keypunching restrictions, they must be entered between @@ and %%, e.g., @@HUMAN%% and @@t%%. Phonemic symbols appear between slants, e.g., /t/.

TOTAL PAGES: 16)

The total number of pages records only the number of Arabic-numbered pages. Prefatory material having Roman numeration is ignored. Any number up to four digits in length may be entered.

BELL SOURCE: 17)

The citations received from Bell Laboratories (see fn. 7, p. 24) contained, within a single field, information about the journal, publisher, volume and number, pages, editor, and sponsor. Since the parsing of that information was not a straightforward task, we have

incorporated the information into this special field as it was received. All relevant data will thus be retrievable.

REPORT NBR: 18)

This field allows inclusion of the report number for papers produced by contract research. The format used is that given by the originating institution. SDC documents have report numbers like TM-1234/001/00.

DOC TYPE: 19)

Only the values A, B, C, J, M, P, R, and U are legitimate. These values are already specified on the citation entry worksheets.

COST: 20)

The price of the document is entered according to the format: dollars.cents, e.g., $9.50.

LOCATION: 121) 21)

Although no particular format is required, tradition favors the use of the owner's surname plus any other distinguishing indicators as to where a copy of the document is kept.

LIB. CONG. NBR: 22)

The Library of Congress number found in most books can be entered here.

SUBJECT ID: 23)

This field exists to permit the addition of unique identifiers for large clusters of entries dealing with a particular subject. This identifier is maintained separately from the keywords for two reasons. First, it ensures the integrity of its uniqueness. That is, since the entry of keywords is unrestricted, several persons entering documents may use the same keyword. The value chosen for this field must be checked with the SOLAR staff before it is used.

Second, it eliminates keywords having so many entries as to be unwieldy and practically useless in a keyword index. This field permits the collection of citations for disparate domains into a single data base while still allowing separation when desired.

SEQ NBR: 24)

The sequence or document ID number is unique for each citation. It has two major functions: it allows cross-referencing between printed indices to the file and it allows efficient updating and file maintenance. The SOLAR staff is responsible for entering this value.

REVIEWED BY: 125) 25)

Although no particular format is required, tradition favors indicating the reviewer's name, the location of his review, and the date; e.g.,

Chapin, Paul. Foundations of Language, Vol. 8. 1972, pp. 298-303.

BOOK: 26)

The title of a book containing a collection of articles by different authors is entered here without underlining or quotation marks.

ENTERED BY: 27)

The last name of the one entering the citation appears here.

DATE ENTERED: 28)

The date on which the citation was entered on the worksheet appears as three sets of numbers, e.g., 11-21-73, where the order is month-day-year.

BELL ID NBR: 29)

Each citation received from Bell Laboratories (see fn. 7, p. 24) contained a uniquely identifying number. We have preserved that number here for anyone who may want to access a citation and already knows the Bell ID number.

ISBN NBR: 30)

The International Standard Bibliographic Number, if one exists, is entered here, with the label 'ISBN' omitted. This number is found in the prefatory pages of a book and identifies the country of publication, the publisher, the title and edition.

4. CITATION RETRIEVAL

The information in the bibliographic citation file is available in two modes. First, one may access the data base by on-line queries to the SOLAR data management system. Second, one can peruse keyword and author indices covering the entire file. These are produced and printed following intervals of document additions.

4.1 ON-LINE ACCESS

All SOLAR files reside in the SDC SOLAR data management system.(*) Since the system is self-documenting and exceptionally user-oriented, our guidance here in the use of the system is quite general.

The SOLAR data management system resides within the CMS time-sharing system running on an IBM 370/145 at SDC. CMS is accessible through the ARPA Network via either TELNET or TIP connections.

(1) To connect to SDC CMS via a TIP, make sure your terminal is set to full duplex and type:

@T <SP> C <SP> L <CR> 'transmit on linefeed'

@L <SP> 8 <CR> 'log to host #8 (SDC)'

The response to you should be:

OPEN 'TIP says you are now connected'

SDC 370/145 TELNET 'SDC net msg'

VM-370 ONLINE 'SDC time-sharing msg'

(*)The SOLAR data management system has come into existence largely because of the selfless, diligent, and competent work of Roy Gates. Through his efforts the system was made compatible with the CMS time-sharing system and the initial compilations were accomplished. Dwight Hare also gave generously of his time and expertise.

'period is the login prompt'

At this point CMS is expecting you to login.

(2) To login, type: LOGIN SOLAR <CR>. SOLAR will then print some sign-on messages and take care of mounting disk packs (if necessary). You will then be asked to sign our visitors log. The signal for your response throughout your interaction with SOLAR will be a hyphen (-) in column 1. Please wait for that prompt before typing. Finish each input by striking the carriage return <CR> key. Terminal input may be either upper case, lower case, or a mixture.

(3) To obtain an introduction to the SOLAR CMS, ask for the new-user format when given that option. Or, type: "EXPLAIN SUMMARY" <CR> (with quotes). SOLAR will then give you a briefing on searching and printing procedures, command names, and program messages.

(4) To obtain an introduction to the bibliographic citation file, type: "EXPLAIN DATABASE" <CR> (with quotes). This will elicit the following table together with an explanation of the various categories of information.

ABBREV	CATEGORY	SEARCHABLE
AU	AUTHOR	
AI	INDIVIDUAL AUTHOR	X
TI	TITLE	
JO	JOURNAL	
PU	PUBLISHER	
PY	PUBLICATION YEAR	X
MO	PUBLICATION MONTH	
LA	PUBLICATION DAY	
VO	VOLUME	
PA	PAGES COVERED	
ED	EDITOR	
SP	SPONSOR	
AB	ABSTRACT	
NO	NOTES	
IT	INDEX TERM	X
TP	TOTAL PAGES	
BS	BFIL SOURCE	

RN	REPORT NUMBER	
DT	DOCUMENT TYPE	X
CO	COST	
LO	LOCATION	
LC	LIBRARY OF CONGRESS NBR	
ST	SUBJECT IDENTIFICATION	X
ID	DOCUMENT ID NUMBER	X
RE	REVIEWED BY	
BO	BOOK ARTICLE IS IN	
EB	ENTERED BY	
DE	DATE ENTERED	
BI	BEIL ID NUMBER	X
IN	ISBN NUMBER	

(5) To search for citations of interest to you, type in an expected value for one of the searchable categories. For example, if searching for documents treating movement transformations, one could type: extrapos# (IT) or cleft# <CR>. The search terms must be entered unpunctuated. The # sign stands for an indeterminate string of characters. The parenthesized field identifier limits the search for the preceding term to that category.

A search can also be made of the non-indexed fields using the STRINGSEARCH facility. Type "EXPLAIN STRINGSEARCH" <CR> (with quotes) for details.

(6) To print data once a citation has been selected, you can use one of the following special print formats:

Command	Fields Returned
"PRINT"	All data in a normal citation
"PRINT SHORT"	Author, title, & publication year
"PRINT FULL"	All fields

It is also possible to tailor your print commands. Type "EXPLAIN PRINT" <CR> (with quotes) for details.

(7) To halt printout of data on your terminal, hit the break key once and wait for the SOLAR prompt (-). Then type: HT <CR> (halt typing). When prompted again, hit <CR> and SOLAR will ask for your next

search statement.

(8) To switch to another data file, type: "FILE <FNAME>" <CR>. For example, "FILE COMPON" <CR>. To ascertain the files available, type "FILES ?" <CR>.

(9) To quit your interaction with SOLAR, type: QUITIT <CR>. SOLAR will then automatically log you out.

4.2 COMPOSED LISTINGS

We have found it useful to provide three types of bibliographic listings at intervals determined by the document entry rate. Other listings can be created upon request.

The first listing is an author index. It is sorted on two fields: the author and the year of publication. In addition to these data, the index indicates, for each entry, its title and sequence number.

The second listing is a keyword index. It also is sorted on two fields: the keyword and the year of publication. In addition, the author, title, and sequence number are given. Since the listing is sorted on high-year, the latest documents on a subject appear first.

The third listing produced is the general bibliography sorted by sequence number. A user can thus determine most of the data available for a particular entry once he has determined the citation he is interested in via the previous two indices.

Readers who desire composed listings are invited to contact the author so that suitable arrangements can be made for the dissemination of data. Because we are particularly interested in enlarging the bibliography, we welcome citation entry in exchange for composed listings. The following section indicates how citations can be added.

THIS
PAGE
IS
MISSING
IN
ORIGINAL
DOCUMENT

5. CITATION ENTRY

The file of bibliographic citations is being built in three ways: by incorporating machine-readable transcripts of bibliographic references, by keypunching citations collected by researchers outside of SDC who wish to add to and utilize the SOLAR citations, and by keypunching citations found by the SOLAR staff as the semantic and conceptual analyses are being done.

Approximately 1,600 entries in the areas of syntax, speech and language disorders, language learning, psycholinguistics, and sociolinguistics have been contributed by the Linguistics and Speech Analysis Dept. of Bell Laboratories, Murray Hill, N.J.(7)

The phonetics department of the University of Michigan contributed about 650 citations in the area of instrumental phonetics.(8)

Through collaboration with Diana Van Lancker (UCLA Linguistics Department) and Marian Macchi (Bell Laboratories) about 625 entries have been added focussing mainly on neurolinguistics, psycholinguistics, speech perception and production, and speech pathologies.(9)

---

(7)Marian Macchi generously provided a transcript of the bibliographic citations available in the BABEL information retrieval system as of September, 1974, in exchange for a similar copy of the SOLAR citations as of approximately the same date.

(8)This collection represents the current status of an ongoing phonetics project under the supervision of Prof. Ian Catford. Nina Macdonald kindly provided us with a copy of the data base.

(9)Diana Van Lancker consented to make the bibliography of her PhD thesis (titled "Heterogeneity in Language and Speech: Neurolinguistic Studies") available for computerization, and Bell Labs sponsored its transcription onto SOLAR entry worksheets. Diana then added keywords, proof-read, and edited the worksheets. After the sheets were keypunched at SDC, Diana once again edited the citations, and they were added to SOLAR.

The UCLA Tone Project, under the direction of Prof. Peter Ladefoged, has contributed about 550 citations on the phonetics and phonology of tone.(10) These include studies on the physiology of phonation and pitch control, pitch perception, inherent pitch of vowels, and the interaction of tone with musical melody in tone languages.

The hand collection of citations by the SOLAR staff is done as documents relevant to lexical semantics are located. The SOLAR staff is regularly scanning periodicals, the output of bibliographic services, proceedings of conferences, and references cited in pertinent articles.

All hand-collected citations are written on data collection sheets that have a format like that shown in the following section. The data on these sheets are then keypunched, converted to upper and lower case, and run into the SOLAR data management system. Because all data are keypunched, we have limited the permissible symbolization to the characters available on the IBM 129 keypunch machines, with three exceptions. Dashes are represented by two continguous hyphens. Left square brackets are keypunched as double AT signs (i.e., '[' --> '@@') and right square brackets are keypunched as double percent signs (i.e., ']' --> '%%'). During on-line editing the AT and percent signs are converted back to left and right square brackets.

Two other symbol restrictions are necessary because of data management conventions. First, the symbol '# is reserved to signal the end of a citation. Second, a string consisting of 'space, digits, right

---

(10)The citations were compiled and written up on SOLAR entry sheets by Ian Maddieson and Jack Gandour and keypunched at SDC. They were then entered into SOLAR and printed out in a format suitable for publication in UCLA Working Papers in Phonetics 28 (11-74).

parenthesis, space' (e.g., ' 1967) ') must not be used, since that string is seen as a field identifier by the DMS. Such a string can be avoided either by placing a character (such as a period) immediately after the parenthesis or by putting a period or comma before the parenthesis.

Additions to this file are continually being made, and entries are solicited from all readers. For those readers wishing to enter citations, the following steps should be taken.

1. Obtain document entry worksheets from the SOLAR project.
2. Choose the appropriate document-type entry blank, i.e., for books, articles, reviews, etc.
3. Print or type all known values in their appropriate blanks. Please do not use longhand, so that we can decrease keypunching trauma and needless editing.
4. The distinction between upper and lower case will be lost in keypunching but can be used to make the entry more readable for the keypuncher, if desired.
5. If a paper has been published in more than one place, either create separate entries or make note of its re-publication in the field labeled NOTES.
6. Return the filled-out forms to Mr. Tim Diller, SDC, 2500 Colorado Ave., Santa Monica, California 90406.

6. SAMPLE CITATION ENTRY WORKSHEET

To conserve space, a single citation entry worksheet follows. The other six document type entry worksheets are similar in form, differing only in the fields appropriate to the document type.

'ARTICLE IN A BOOK' Entry Form

AUTHOR: 102)2) PARTEE , BARBARA
102)2) _____ , _____
102)2) _____ , _____
TITLE: 3) "Linguistic Metatheory"

BOOK: 26) A Survey of Linguistic Science

EDITOR: 11) Dingwall, William (Ed.)
PUBLISH: 5) University of Maryland
YEAR[9999]: 6) 1971 PAGES: 10) pp. 650 - 680
LOCATION: 121)2?) Olney SEQ#: 24) 8511
DOCTYPE: 19) C
ENTERED BY: 27) Diller DATE ENTERED: 28) 3 - 14 - 74
NOTES: 114) 14)
114) 14)
ABSTRACT: 113) 13)
113) 13)
KEYWORDS: 115) 15) Theory, Linguistic 115) 15) Interpretive Semantics
115) 15) 115) 15)