# USER'S GUIDE TO THE SOLAR SEMANTIC ANALYSIS FILE

SYSTEM DEVELOPMENT CORPORATION

PREPARED FOR
ADVANCED RESEARCH PROJECTS AGENCY

APRIL 1975

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>TM-5292/001/00 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER<br>AD-A009 328 |
| 4. TITLE (and Subtitle)<br>User's Guide to the SOLAR Semantic Analysis File | | 5. TYPE OF REPORT & PERIOD COVERED<br>Special<br>Technical |
| | | 6. PERFORMING ORG. REPORT NUMBER<br>TM-5292/001/00 |
| 7. AUTHOR(s)<br>Tom Bye<br>Timothy Diller<br>John Olney | | 8. CONTRACT OR GRANT NUMBER(s)<br>DAHC15-73-C-0080 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>System Development Corporation<br>2500 Colorado Avenue, Santa Monica, Cal. 90406 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>ARPA Order 2254<br>Program Code 5D30 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS | | 12. REPORT DATE<br>4-31-75 |
| | | 13. NUMBER OF PAGES<br>39 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)<br>ARPA/Information Processing Techniques<br>1400 Wilson Boulevard<br>Arlington, Virginia 22209 | | 15. SECURITY CLASS. (of this report)<br>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 semantic analysis 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.



DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

ABSTRACT

This document contains a general explanation of the semantic analysis 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.

# 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.

(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 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. This file is intended to be used in close conjunction with the next two files described below.

---

(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) 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[2] will provide a series of displays showing most of the other words in the English vocabulary that stand in a morphological, definitional, synonymitive, antonymitive, or thesaural

---

[2]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 [3] 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[4] will, when complete, contain representative 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.

---

[3]Although this file has not yet been produced, the programs needed to build it are close to final checkout.

[4]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/C2

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

7. Diller, T., & T. Bye. (1975) "User's Guide to the SOLAR Theoretical Backgrounds File" SDC Document TM-5292/002/00

8. Diller, T., T. Bye & J. Olney. (1975) "User's Guide to the SOLAR Semantic Component File" SDC Document TM-5292/003/00

9. Diller, T., & P. Heath. (1975) "User's Guide to the SOLAR KWIC File" SDC Document TM-5292/008/00

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

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

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

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

## 2. FILE DESIGN

The file of semantic analyses is intended to make readily available a collection of lexical facts which have heretofore been widely dispersed throughout the linguistic literature. The file obviates to a great extent the following tasks:

- determining where analyses of a given word have been given.(5)
- obtaining copies of the potentially useful articles,
- screening out those articles that do not seem to make a useful contribution, and
- extracting from the remaining articles the information relevant to the given word.

The file also aids the user by presenting the collected facts within a uniform framework.

The file of semantic analyses was designed with the following criteria in mind. First, the file should bring together as much semantic and syntactic data pertinent to a certain subset of English words as is feasible and desirable. The quantity of data per entry is determined in part by our data collection procedure and in part by the richness of the analysis sources.(6) The 'subset' qualification is

---

(5)The difficulty in locating documents treating a given word stems from the nature of the indexing and abstracting services that cover the linguistic literature (e.g., MLA's International Bibliography, Language and Language Behaviour Abstracts, Bibliographie Linguistique, and Language Teaching Abstracts). That is, their subject indices generally do not include the particular words for which semantic analyses are given and they do not include unpublished works.

(6)Our policy has been to cover as wide a range of lexical entries as possible and we have therefore chosen source documents treating several words. As a result, the file includes analyses which can be generously characterized as "partial". Having developed a rather broad base, we are now working with documents that treat fewer words in greater detail.

determined by the lexicons of the users of the archive (primarily the ARPA SUR projects) since the words for which analyses are currently sought are those appearing in the SUR lexicons. The 'feasible' qualification is bounded in part by our level of manning(7) and in part by the accessibility of data sources. The 'desirable' qualification relates to the anticipated utility of the data for speech understanding.

Second, the file should be of practical assistance to researchers engaged in modeling the understanding of English on computers. Since the lexicons currently being analyzed are those of the SUR projects, a vast amount of the English vocabulary was determined to be outside our purview for the present.

Finally, the file should be directly accessible to researchers, and the time required to learn the file structure and data management protocols should be minimal. The file is, accordingly, being placed into a user-oriented SOLAR data management system accessible via the ARPA Network.

An initial concern of ours was whether it would be possible to cull the best analyses from the literature, accepting all theoretical frameworks, and still enter data within a tight set of categories of information without doing violence to these analyses. We felt the necessity of a uniform data format (for data management efficiency) but, at the same time, wished to maintain the distinctiveness of each author's viewpoint. Accordingly, several steps have been taken to retain uniqueness while mitigating conflicts arising from differences

---

(7) Since its inception, our project has averaged one full-time and three part-time researchers.

among the authors' theoretical frameworks and constraints imposed by our own format.

First, we treat each author's analysis of a particular word separately, although reference to other analyses may be made. (For each of the words having several analyses in the literature, we are tentatively planning the addition of a "SOLAR analysis" wherein comparison and (partial) integration will be attempted.)

Second, to maintain uniformity, each analysis is entered under identical categories of information. These include an indication of the sense of the word being treated, the semantic components or descriptive constants utilized by the author in defining its core meaning, whatever predicate-argument relationships, selectional features, and/or implications he may specify, and his informal explanations or qualifications, if any. (In cases where the author fails to include an informal commentary, we generally provide one, taking care to label it as our supposition.)

A third step toward reducing the inherent diversity of the analyses is the inclusion of a SOLAR critique. This is put forward with provisional acceptance of the author's framework but with the writer of the critique attempting to put himself into the position of a builder of a speech understanding system.

To ensure that our condensation of each author's analysis does not result in a loss of precision or clarity, we include a separate file in which we attempt to define, as far as possible in the author's own words, the terms he uses as semantic components. We are finding that in many instances, authors have not been concerned with this level of description. Hence, we sometimes have to extrapolate from a variety of

clues and provisionally enter, for one or more semantic components, definitions that appear to match the author's intent.

To ensure that our condensation does not result in incomprehensibility due to lack of context, we have developed a separate file that provides a digest of the author's theoretical orientation together with an inventory of the notational conventions he utilizes. We hope this will enable the reader (if he desires) to reconstruct the particular framework within which a given analysis was presented.

## 3. DEFINITION OF FIELDS

There are currently 23 fields into which data are entered when a semantic analysis is built. The type and format of the data to be entered in each field are described below.

SOLAR Word:

The term being analyzed can be either a single word or a phrase. It is generally the canonical, uninflected form of a word. (Since words appearing in the SUR lexicons are both inflected and uninflected, we have chosen to call the terms in this field SOLAR words rather than SUR words.) In the companion SOLAR word index file, all inflected forms appearing in the SUR lexicons will reference the semantic analyses given for the uninflected form.

SUR Domain:

A connection between the analyses entered for a given word and the SUR lexicons containing that word or one of its inflected forms is made by entering the name of one or more of the lexicons in this field. Each name is entered with a separate field identifier for data retrieval purposes. The list of names currently employed is as follows:(*)

AP (Associated Press Releases--CMU)
APSYS (Associated Press Releases--CMU)
CHESS (Chess Playing Lexicon--CMU)

---

(*)CMU refers to the Carnegie-Mellon University SUR project; BBN refers to the Bolt Beranek and Newman SUR project; SDC refers to the SUR project at System Development Corporation; SRI refers to the Stanford Research Institute project currently providing direct support to the SDC SUR project.

CHT (Second Chess Playing Lexicon--CMU)
DESCAL (Desk Calculator Lexicon--CMU)
DOCTOR (Medical Lexicon--CMU)
PACFLEET (Warships--SDC)
SMALLWORD (Core of Lunar Rocks Lexicon--BBN)
SUPS (Subset of Jane's Fighting Ships--SDC)
TRAVEL (Business Travel Lexicon--BBN)
VOCAB (Pump and Faucet Repair Lexicon--SRI)
WORDS (Expansion of SMALLWORD--BBN)

Additional lexicons will be added as they are produced by the SUR projects.

The domain field can also be used to identify a semantic field to which a SOLAR word belongs. Ikegami's verbs of motion (Ikegami, 1969) and Lehrer's cooking words (Lehrer, in press) illustrate this possibility.

Semantic Analysis Ntr:

Since more than one semantic analysis may be entered for a given word, this field permits the inclusion of a value uniquely identifying each analysis. The value is of the form 'S0009'.(9) This value is typically used during retrieval or updating to uniquely identify the analysis of interest.

Source:

Since each word may be treated by more than one author, provision has been made to keep each such analysis separate. All data within a particular analysis relate to the particular source being analyzed at the time. The source is entered as follows:

---

(9)'S' stands for 'semantic analysis', '0' stands for an optional digit and '9' indicates an obligatory digit.

Surname, Firstname. Year. Title.

A corresponding bibliographic citation containing all particulars is i each case available in the companion "Bibliography" file.

Author's Sense:

Where the author has restricted his analysis to a subset of the senses of the term, that restriction is given in this field. When no explicit sense restriction is given, but sentences illustrating the us of the term are provided, those sentences best illustrating the sense( apparently being treated are entered here. If no sense restriction ca be inferred, we note that fact.

Webster-7 Part of Speech:

When the SCIAR term being analyzed is homographic, the homograph number, term, and part of speech of the relevant W7 definition(s) are entered here. (The actual definitions of this homograph are entered i the following field.)(10)

Webster-7 Sense(s):

To provide the reader with some basis of comparison as he reads t analysis, the W7 sense(s) most closely related to the one(s) the autho is apparently investigating are included in this field. The repeating group structure permits each sense to be entered and retrieved

---

(10)The material in this file taken from Webster's Seventh New Collegiate Dictionary, copyright 1967 by G. & C. Merriam Company, publishers of Merriam-Webster dictionaries, is used by permission. Al rights to such material are reserved by G. & C. Merriam Company.

individually. Thus, sense 1a is entered separately from sense 1b and sense 2.

Preliminary Qualification:

Two basic types of information will be found in this field. First, if an author has placed a circumscription or caveat upon his analysis of a particular word, that limitation is entered here. Some qualifications are syntactic in nature (e.g., limiting an analysis to the intransitive sense of a verb); others are semantic (e.g., specifying the particular sense(s) to be analyzed); while still others are pragmatic (e.g., specifying the non-linguistic context within which a particular word-sense is being considered). Second, if a characterization of the author's theoretical orientation has already been entered in the 'Theoretical Backgrounds' file, the identifying number of the T-entry is given; otherwise we simply note the fact that no such characterization has yet been entered.

Predicate-Argument:

If the SOLAR word can be used as a logical predicate, its underlying predicate-argument structure is found in this field. Because the format is open, we usually adhere to the one used by the author of the analysis. Characteristics of the individual arguments are found in the field titled Selectional Features.

Components:

The legitimacy of a componential semantic analysis has been commented on by Weinreich (1968, p. 28) as follows:

> "That the senses of many morphemes (or lexemes) of a language are indeed analyzable into components can, I think, be supported by such evidence as anthropologists have been developing in their studies of special vocabularies and which linguists have been unearthing in certain areas of "general" vocabulary.... Nevertheless it is sobering to realize that in every language many morphemes or lexemes yield no clear-cut componential analysis, and for many others such analysis is hard to conceive of altogether...."

Stimulated by Weinreich's writings, Hofmann has investigated the relation between components of meaning for given lexical items and presented the following conclusions:

> In review, semantic atoms are motivated by minimal semantic differences which are systematic in the lexicon &/or the morphology of a language. We have dealt only with the concept or descriptive meaning of lexical items, but have shown that (a) it can be analysed into atoms of meaning, (b) it is, in general, not the intersection of the sets designated by the atoms, (c) indeed, it is not even an ordered list of atoms (features, components), & (d) it is not restricted to ordered sequences (or trees) of atoms. Rather, some atoms are ordered relative to others, while other atoms may not be ordered relative to each other. Several types of order can be distinguished, & I conclude that a fully general network of relations is needed. (1974, p. 17)

Weinreich's caveat regarding the difficulty (impossibility?) of finding components for certain lexemes is supported in our collection of analyses. Not all words analyzed have data entered under Components. Hofmann's claim that a fully general network of relations is needed to represent the interconnection of semantic components has also been sustained.

Lacking a graphics display capability and hence being incapable of representing a network as it might be drawn, SOLAR (while holding the theoretical bias maintained by Hofmann) has accommodated itself to data

management by providing both a Component field and a Component Composition field, the latter being a repository for comments on the relations holding between components.

If, in the author's analysis, semantic components (= semantic markers, = semantic primitives) have been utilized, they are entered a separate components here. If some relation of order is indicated or deemed necessary, comments to that effect are entered in the following field. Where the author has integrated these components into a networ structure, we enter the components (nodes in the network) individually and treat the relations (arcs) between them specially in the following field. In any case, the components themselves are treated further in the SOLAR semantic component file (see the companion document entitled "User's Guide to the SOLAR Semantic Component File").

Component Composition:

As noted above, comments on the relationships existing among the individual components are entered here. In many cases relationships only implicit in the author's analysis are made explicit here. In thos cases where the author explicitly indicates interconnections between components, these interconnections are entered as quoted material. If some components function as terms relating other components, that is pointed out. Analyses written from a generative semantics framework receive special treatment within this field since the representation of meaning for a lexical entry is in terms of primitive concepts embedded in tree structures. Analyses arranging components in even more complex network structures likewise receive expanded prose treatments.

Subcategorial Restrictions:

This field incorporates restrictions on the syntactic category of an argument or the structure of arguments accompanying the predicate. For example, if a particular predicate accepts only a sentential object that fact will be entered here.

Rule Features:

For analyses written within a generative framework, provision has been made for indicating the lexical features proposed as triggering or permitting the operation of particular syntactic rules. Thus, within the area of nominalization and complementation, one can indicate, for example, features governing extraposition, equi-NP deletion, and complementizer insertion. Within the area of case placement, features governing passivization and subject or object raising can be included. Similarly, the many determiner idiosyncrasies can be specified here.

Although short-hand feature notation is sometimes employed, we generally enter data here in prose and attempt to state the utility of the feature in terms of its effects on surface structure. To enhance the readability of this field, we discuss in the 'Theoretical Backgrounds' file, features that appear to be idiosyncratic to the metatheory being used. Rule features that have some direct correlation with surface structure realizations are discussed from that perspective.

Selectional Restrictions:

Limitations on the co-occurrence of word sets with the word being analyzed may be specified either in terms of the particular words involved or in terms of (complexes of) semantic features delimiting the

corresponding word sets. Depending on the word being analyzed and the metatheory within which the analysis is made, either or both approaches may be followed. The former (collocational) approach is typically used to indicate those prepositions, if any, that signal particular cases when they occur in construction with the analyzed term. The latter (feature) approach is typically used to specify semantic restrictions holding between the predicate and the nominal arguments accompanying it.

When an author has followed the feature approach, we have included in this field both features that are intended to be utilized at the point of lexical insertion (generally considered true selectional features) and features that are intended to be used at the point of interpretation (spoken of by Weinreich (1966) as 'transfer' features and by Leech (1970) as 'ascription' features).

Within this field one might also find an indication of whether a particular case role is obligatory or optional and, if optional, whether the case can be inferred from context.

We enter selectional features separately from presuppositions for two reasons, one pragmatic and the other theoretical. First, most analyses written by linguists have used the term 'selectional restriction'. Only recently has the term 'presupposition' come into common usage. Since confusion is avoided by preserving an author's terminology wherever possible, we place all comments originally considered 'selectional restrictions' under this identifier. Second, it is not at all clear that selectional restrictions are equivalent to presuppositions. As Keenan (1969, p. 145) points out, selectional restrictions, unlike presuppositions, are stated in terms of predicates that are not present in the sentence. Further, the failure of a

selectional restriction to hold must be interpreted differently from the failure of a presupposition. The former means either that an incorrect appraisal of the properties of what a nominal argument refers to has been made or that a trope is being used which permits such deviance. The latter has often meant that no referent can be found.

Presuppositions:

The data in this and the three fields following are entered according to the labels chosen by the author. That is, if the author chooses to designate some proposition as a presupposition rather than as an implication, entailment, or assertion, we adhere to the designaton he has selected. We have on occasion, however, commented on the appropriateness of the label chosen if the author appears to depart from "accepted" usage of the term. Since the definition of the field is thus left in the main to the author, we will simply try to point out in our description of these four fields what has frequently been placed under the labels given.

"Presupposition" has been used both by linguists and philosophers to refer to a wide variety of phenomena. The divergence in opinion as to what the term means centers on the issue of what the constituents of a presupposition relation should be taken to be and extends to the consequences of a presupposition failure.

Among the candidates which have been proposed as that which presupposes are the speaker, his utterance, underlying propositions, and individual predicates. Likewise, that which is presupposed has been variously identified with the situation or state of affairs holding at the time of the utterance, with a proposition, and with the truth value

of a proposition.

With regard to the consequences of a presupposition failure, different authors have suggested that such a failure results in ungrammaticality, meaninglessness, inappropriateness, falsity, failure to perform a speech-act, and lack of a truth value.

The reader wishing a further discussion of these various views is referred to Appendix III: Notes on Presupposition.

Entailment:

Virtually all linguists, and most philosophers who have been concerned with natural language, use the term 'entailment' to designate those inferences that can be deduced from a sentence by virtue of (a) the cognitive meaning of its constituent words and phrases and/or (b) its logical form, i.e., those inferences that can be made with the aid of various logical inference rules. Thus, "John is a bachelor" is considered to entail "John is an adult male", "John is unmarried", etc.

'Entail' is often used in statements about the truth value of propositions expressed by sentential complements of particular verbs. Thus, one finds in the literature statements like "'discover' entails that its complement is true".

Assertions:

Hooper and Thompson (1973, p. 473) in their treatment of sentential complements, make the following remarks about assertions:

> The assertion of a sentence is its core meaning or main proposition. In most cases the assertion of a declarative sentence is found in the main clause. The assertion of a sentence may be identified as that part which can be negated or questioned by the usual application of the processes of

> negation and interrogation. It is usually assumed that all assertions are speaker assertions. We will claim here, however, that some embedded statements have the characteristics of assertions, as can be seen when the tests of negation and questioning are strictly applied.
>
> A single sentence--for example, one consisting of two coordinately conjoined Ss--may contain more than one assertion. Less obvious, however, is the fact that there are also some subordinate clauses that are asserted, even though they are slightly subordinate to the main assertion of the sentence.

They then go on to discuss five classes of verbs taking sentential complements, only some of whose complements are asserted. SOLAR has included this field for the entry of such idiosyncratic data.

Other writers have connected assertions to the complements of verbs such as 'say'. Thus, in the sentence "John says Bill has already left", the speaker is reporting an assertion made by John. Recognition of this property of certain verbs can also be made here.

Implications:

Linguists and philosophers alike have viewed implication as a group of relations, one of which is entailment.(11) Linguists working with contrary-to-fact conditionals have noted the relationship of their semantic interpretation to that of the "if-then" connective (and hence to material implication -cf. footnote 11). For example, given the sentence "If Harry would have come, we could all be gone by now", the

---

(11)Philosophers have recognized two principle types of implications: strict (or logical) implication, which is indistinguishable from entailment (the latter is defined as "the relation that exists between two propositions one of which is deducible from the other" (The Encyclopedia of Philosophy, Vol 5, p. 63)); and material implication (which is defined as the interpretation of the "if-then" connective according to which "'if A then B' is true in all cases except when A is true and B false" (The Encyclopedia of Philosophy, Vol. 5, p. 66)).

proposition "Harry didn't come" may be said to be implied (by modus tollens). Such implications will generally not be treated in SCLAR since the implications can not be tied to a single lexical entry.

Recently, some writers (e.g. Grice (1968)) have distinguished unde the labels "conversational implicatures" and "invited inferences" certain implications that do not possess the logical properties of either material implication or entailment. Conversational implicatures include those inferences deducible by virtue of the context of the utterance, common presumptions about the nature of the world, conversational conventions, etc. A proposition conversationally implie may correspond directly to a surface structure construct. Thus, the complements of certain verbs form the nuclei of conversational implicatures. For example, "John opened the door" is an implicature of "John remembered to open the door".(12) In other cases, the implied proposition does not correspond directly to a surface structure construct but must be inferred from the semantic interpretation of the sentence as a whole. For example, "John is not here" is a conversational implicature of "John left an hour ago". Note that in both of these examples the former proposition is not simply entailed by the latter, since, in the first example, something may have prevented John from opening the door and, in the second, John may have returned.(13) But if nothing more is said, the hearer is entitled to assume that the implied proposition is true on the grounds that the

---

(12)Note that negation of the embedding verb also negates the implied proposition. "John didn't remember to open the door" implies "John didn't open the door".

(13)In contrast, "the door was open" is entailed by "John remembered that the door was open".

speaker will otherwise have violated a conversational convention and said something misleading.

Some authors (e.g., Lehrer (1973, p. 175)) treat implication as an evaluation of the truth or falsity of a proposition expressed by a complement. A particular sentence may imply that the speaker believes that the proposition expressed by a complement is true or false, as in: "Bill knows that John is here".(14)

Connotations:

As Hall (1972) has pointed out, connotations can be divided (in an idealized world) into two types: those held in common by members of a speech community and those unique to an individual. The latter are clearly inappropriate in the context of our archiving objectives. The former encompass a rather broad range of phenomena.

The label 'connotation' is sometimes applied to the association of a word or phrase with a particular dialect. Thus, to use Hall's example, 'to muscle in on' is associated with the speech of gangsters.

'Connotation' has also been used to label the emotive values signalled by the use of a word. Thus, in referring to a newspaper as a 'fishwrapper' or to a public servant or politician as a 'bureaucrat', the speaker is indicating an emotional attitude.

No formal structure for the collection of such data is proposed and the author's own formalization is generally used.

---

(14)Lehrer is obviously not rigorously distinguishing "presupposition", "implication", and "entailment" from each other, however, since all three terms are applied (in the same paragraph) to the relationship holding between "know" and its complement.

Informal Explanation:

Although some authors provide a primarily formal description of a term, while others provide only a prose treatment, the great majority provide a brief formalization and follow that up with commentary expanding upon the formalization. Such informal comments are entered here together with any interpolations or clarifications the SOLAR analyst deems useful.

Final Qualification:

Where the author has added a footnote to his analysis (either literally or figuratively), it is entered as a final qualification.

SOLAR Comments:

Comments by the SOLAR analyst are made with provisional acceptance of the author's framework. That is, criticism of the author's theoretical orientation is minimized and attention is focused on any lacunae, vagueness, ambiguities, or contradictions that may be found in his analysis. Comparisons with other analyses of the same word sense are also included where instructive.

Users are invited to comment on analyses (cf. section titled 'DATA COLLECTION' for details) and, when the inclusion of such comments is acceptable to both the user and the SOLAR staff, those comments are entered here with appropriate attribution.

Analyst:

This final field simply indicates who put together the SOLAR entry.

# 4. DATA RETRIEVAL

The information in the semantic analysis file is available in two modes: via on-line queries to the SOLAR data management system (DMS) over the ARPA Network and by listings distributed by the SOLAR staff.

## 4.1 ON-LINE ACCESS

All SOLAR files reside in the SDC SOLAR data management system.(15) 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> 0 <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' |
| . | 'period is the login prompt' |

At this point CMS is expecting you to login.

---

(15)The SOLAR data management system has come into existence largely because of the tireless, 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 Harm also gave generously of his time and expertise.

(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 inpu 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 DMS, 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 access the semantic analysis file,[16] type: "FILE SEMANAL <CR>.

(5) To obtain an introduction to the semantic analysis file, type "EXPLAIN DATABASE" <CR>. This will elicit the following table together with an explanation of the various categories of information.

| ABBREV | CATEGORY | SEARCHABLE |
|---|---|---|
| WD | SOLAR WORD | X |
| DO | SUB DOMAIN | X |
| SN | ANALYSIS NBR | X |
| SO | SOURCE | X |
| AU | AUTHOR'S SENSE | |
| WP | W7 PART OF SPEECH | X |
| WS | W7 SENSES | |
| PQ | PRELIMINARY QUALIFICATION | |
| PA | PREDICATE-ARGUMENT | |
| CO | COMPONENTS | X |
| CC | COMPONENT COMPOSITION | |
| SR | SUBCATEGORIAL FEATURES | |
| RU | RULE FEATURES | |
| SE | SELECTIONAL FEATURES | |

---

[16]The SOLAR DMS initially accesses the bibliographic citation file.

| | |
|---|---|
| PR | PRESUPPOSITIONS |
| EN | ENTAILMENTS |
| AS | ASSERTIONS |
| IM | IMPLICATIONS |
| CN | CONNOTATIONS |
| EX | INFORMAL EXPLANATIONS |
| QU | FINAL QUALIFICATIONS |
| CR | SOLAR CRITIQUE |
| AN | ANALYST |

(6) To search for semantic analyses of interest to you, type in a expected value for one of the searchable categories. For example, type with (WD) or CAUS# (CO) <CR>. The search terms must be entered unpunctuated. The # sign stands for an indeterminate string of characters. The category in parentheses limits the search to a single field.

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

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

| COMMAND | FIELDS RETURNED |
|---|---|
| "PRINT" | WD, SN, and SO |
| "PRINT SENSE" | AU, WP, AND WS |
| "PRINT COMPONENTS" | PA, CO, CC, and EX |
| "PRINT CONTEXTS | SB, RU, SE, PR, EN, AS, IM and CN |
| "PRINT QUALIFICATIONS" | PQ, QU, and CR |
| "PRINT FULL" | All Fields |

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

(8) 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.

(9) To switch to another data file, type: "FILE <FNAME>" <CR>. E.g., "FILE COMPON" <CR>. To ascertain the files available, type "FILE ?" <CR>.

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

## 4.2 COMPOSED LISTINGS

The semantic analyses are being made available in printed form as well as on-line.(17) Users wishing to receive these listings should request them from Tim Diller. The user is advised, however, that the on-line version is likely to be more current than the printouts, which will be produced only at intervals of significant accretion.

---

(17) Not all users are expected to have access to the ARPA Network and some analyses may be considered unsuitable for terminal printout because of their length.

## 5. DATA COLLECTION

The file of semantic analyses is being built manually by locating and reading documents relevant to the 3,000 words in the speech understanding research lexicons, extracting the essence of the authors' analyses, writing critiques of them, and entering these data on sheets for keypunching.(18)

Documents are located by scanning periodicals, the output of bibliographic services, proceedings of conferences, and references cited in pertinent articles. (Quasi-)formal semantic analyses have appeared in the linguistics literature only in recent years but already cover a wide range of words. For words that have received multiple semantic analyses, one of two situations generally holds. Either the analyses were done independently, with no intention of criticizing or bolstering a previous analysis, or -- less often -- the analyses were put forward as competitors, sometimes primarily with a view to defending a particular theoretical framework rather than improving the understanding of the word's meaning or use.

The process of extracting analyses from the literature is influenced to some extent by the nature of the analyses found. For the independently constructed analyses, SOLAR is providing an opportunity for comparison. For the groups of competing analyses, SOLAR is suppressing polemical remarks and focusing on what is said about the word's meaning.

---

(18)The data entered in this file have been collected for the most part by Tom Bye, with assistance from Martin Mould and Tim Diller.

The extracted data are written on data collection sheets that have a format similar to that shown in the sample entry of Section 7. 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 contiguous hyphens. Left square brackets are keypunched as double AT signs (e.g., '[' --> '@@') and right square brackets are keypunched as double percent signs (e.g., ']' --> '%%'). During on-line editing, the AT and percent signs are converted back to their original representation.(19)

Once an analysis has been entered in the data management system, a printout is sent to the author of the document from which we extracted the analysis, with a letter soliciting his comments and criticisms. Revisions are then made where appropriate. This serves both as a check on the accuracy of our characterizations and as an opportunity for the inclusion of afterthoughts by the author.

We also solicit commentary on analyses from all archive users. If the contributor of such commentary agrees, such comments can be entered (with appropriate attribution) into the field titled 'SOLAR Comments'. We hope in this way to stimulate consideration of the claims being made

---

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

about the meanings of particular word-senses and to air comments on particular words that may not have been deemed appropriate for inclusion in published articles.

## 6. REFERENCES

Austin, J. (1962) How To Do Things With Words. Harvard University Press.

Bellert, Irena. (1973) "On Inferences and Interpretation of Natural Language Sentences"

Fillmore, Charles. (1968) "Types of Lexical Information" in Working Papers in Linguistics (OSU) Vol. 2, pp. 65-103.

Fillmore, Charles. (1971) "Verbs of Judging: An Exercise in Semantic Description" in Studies in Linguistic Semantics (ed. by Fillmore and Langendoen) Holt, Rinehart, and Winston, pp. 273-289.

Grice, H.P. (1968) "Logic and Conversation" unpublished version of The William James Lectures delivered at Harvard University.

Hall, Robert A., Jr. (1972) "Why a Structural Semantics Is Impossible" Language Sciences (I.U.: Research Center for the Language Sciences) 21:1-6.

Hofmann, T. R. (1974) "Semantic Studies IV: In Support of Semantic Atoms" Language Sciences 32:12-18.

Hooper, J. & S. Thompson (1973) "On the Applicability of Root Transformations" Linguistic Inquiry 4.4:465-497.

Horn, Larry. (1969) "A Presuppositional Analysis of only and even" in Papers from the Fifth Regional Meeting of the Chicago Linguistic Society (ed. by R. Binnick, et al.) University of Chicago, pp. 98-107.

Ikegami, Yoshihiko. (1969) The Semological Structure of English Verbs of Motion.

Kartunnen, L. (1970) "On the Semantics of Complement Sentences" in Papers from the 6th Regional Meeting of the Chicago Linguistic Society pp. 328-339.

Karttunen, L. (1971) "Implicative Verbs" Language 47.340-358.

Katz, J. (1973) "On Defining 'Presupposition'" Linguistic Inquiry 4.2:256-260.

Keenan, E. (1971) "Two Kinds of Presupposition in Natural Language" in Studies in Linguistic Semantics (ed. by Fillmore and Langendoen) Holt, Rinehart, and Winston, pp. 45-52.

Kiparsky, P. and C. Kiparsky (1970) "Fact" in Progress in Linguistics (edited by Bierwisch & Heidolph) Mouton & Co., The Hague.

Kuroda, S-Y. (1969) "Remarks on Selectional Restrictions and Presuppositions" in Studies in Syntax and Semantics (ed. by Kiefer)

pp. 138-167.

Lakoff, George. (1972) "Linguistics and Natural Logic" in Semantics of Natural Language (ed. by Harman and Davidson) Dordrecht: D. Reidel, pp. 545-665.

Lakoff, George. (1971) "The Role of Deduction in Grammar" in Studies in Linguistic Semantics (ed. by Fillmore and Langendoen) Holt, Rinehart, and Winston, pp. 63-72.

Langendoen, D. T. and T. Savin. (1971) "The Projection Problem for Presuppositions" in Studies in Linguistic Semantics (ed. by Fillmore and Langendoen) Holt, Rinehart, and Winston, pp. 55-62.

Leech, G. (1969) Towards a Semantic Description of English. Indiana University Press, Bloomington.

Lehrer, Adrienne. (1973) "Verb and Adverb Interaction" Univ. of Rochester, unpublished MS.

Lehrer, Adrienne. (in press) Semantic Fields and Lexical Structure.

McCawley, James. (1972) "A Program for Logic" in Semantics of Natural Language (ed. by Harman and Davidson) Dordrecht: D. Reidel, pp. 498-544.

Muraki, M. (1972) "Discourse in Presupposition" Papers in Linguistics 5.2.

Schachter, J. (1971) "Presuppostion and Counterfactual Conditional Sentences." UCLA PhD Thesis.

Strawson, P. (1952) Introduction to Logical Theory. London: Methuen.

Weinreich, Uriel. (1966) "Explorations in Semantic Theory" in Current Trends in Linguistics. Vol. III. (ed. by T. Sebeok) The Hague: Mouton, pp. 395-477.

Weinreich, Uriel. (1968) "Problems in the Analysis of Idioms" in Substance and Structure of Language (ed. by J. Puhvel) University of California Press, pp. 23-82.

Wilson, Deidre. (forthcoming) "Presuppositions and Non-Truth-Conditional Semantics" MIT PhD Thesis. draft in hand dated 1973.

## 7. APPENDIX I: SAMPLE ENTRY

Word: GIVE
Domain: AP, PACFLRET, SMALLWORD, TRAVEL, VOCAB
Semanal #: S167
Source: Katz, J. 1972. Semantic Theory.
Author's sense: "...a process in which the possession of something is transferred from one person to another without anything being given in return. ...[7.176] [a] John gave the book to Mary" [p. 348]
W7 sense: vt 1: to make a present of
2a: to grant or bestow by formal action
3a: to put into the possession of another for his use
3c: to commit to the trust or keeping of another
Prelim. qual.: By removing from the definition of "sell" the process semantic marker which indicates the sum of money for which an item is sold, the lexical reading for "give" can be obtained. [paraphrase of Katz on p. 348]
Predicate-args: 'give': (subject-of) (object-of) (indirect-object-of) (inflexional)
Components: ((Condition) (Possesses Y) of X at T-i)
((Condition) (Possesses Y) of Z at T-j)
Component composition: The process of giving is represented by a transition from the state represented by the first component to the state represented by the second component. 'X' represents the subject, 'Y' the object, and 'Z' the indirect object. 'T-i' indicates the initial state and 'T-j' the terminal state.
Selectional features: 'Y' = <(Physical object)>
'X' = <(Human) & (Not-infant)>
'Z' = <(Human) & (Not-infant)>
Solar comments: Katz' definition has the outstanding quality of representing the process involved in giving. It is deficient in other respects, however. First, it fails to indicate that the subject ('X') initiates/causes the process. Second, it does not indicate that in the terminal state (T-j) the subject (X) no longer possesses the object (Y). Third, the selectional restrictions placed on the indirect object (Z) hold for the other verbs of transfer of possession which Katz treats but not for 'give'. Cf. "John gave the calf/the baby some milk." Fourth, there is a minor discrepancy between the "author's sense" and the "selectional features" in the statement of properties that the possessed object (Y) can have. We suggest a broadening of the selectional features of (Y) to include <(legal entity)>, since the possession of legal entities which are not physical objects (e.g., a copyright, power of attorney, etc.) may be transfered without anything being given in return.

# 8. APPENDIX II: SAMPLE ON-LINE INTERACTION

The following interaction is typical of what may be expected in on-line access to this file. Terminal inputs are the lines having a hyphen in column 1.

```
-"file semanal"          <user chooses file to be accessed>

YOU ARE NOW CONNECTED TO THE SEMANAL DATABASE.

SS  1 /C:        <SOLAR asks for first search statement or command>
-find            <user asks for entries having 'find' as searchable term>

PSTG (4)         <SOLAR indicates it found four such entries>

SS  2 /C:        <SOLAR asks for second search statement or command>
-"print"         <user commands printing of preliminary data>

WD- FIND
SN- S233
SO- Fillmore, C. 1971. "Types of Lexical Information"

WD- FIND
SN- S269
SO- Hooper, J. 1974. "On Assertive Predicates"

WD- FIND
SN- S216
SO- Nash-Webber, B. 1971. "Verbs of Composition"

WD- FIND
SN- S545
SO- Cook, W. 1973. "Covert Case Roles"

SS 2 / C:        <SOLAR asks for search statement or command>
-s233            <user asks for entry having 's233' as searchable term>

PSTG (1)         <SOLAR indicates it found one such entry>

SS 3 /C:         <SOLAR asks for search statement or command>
-"ptint"$        <user makes spelling error and deletes line with $>
"print sense"    <user commands printing of sense>

SN- S233
AU- The sense of 'find' being described is illustrated in the
    sentence: 'He may find the eggs'.
WP- 1 find vt
WS- vt 1a: to come upon often accidentally: encounter
```

```
SS  3 / C:              <SOLAR asks for search statement or command>
-"print components"     <user commands printing of components>

SN- S239
CO- ACTOR
CO- BY-CHANCE
CO- ACTIVITY
CO- OUTCOME
CC- The 'activity' of the 'actor' leads 'by chance' to the 'outcome'.
EX- "Sometimes a verb has a built-in reference to the outcome of an
    activity. Conceptually it appears that the actor engages in some
    activity and through the activity may be directed toward some
    specific outcome.
EX- It is the activity itself which (by chance) leads to that
    .
    .
    .

SS  3 /C:               <SOLAR asks for search statement or command>
-
```

## 9. APPENDIX III: NOTES ON PRESUPPOSITION

This section is currently in preparation and will be included in the next edition of this user's guide.