Artificial Intelligence Project--RLE and MIT Computation Center
Memo 38 -- Machine Understanding of Linguistic Information--
A Survey and Proposal

by Bert Raphael

I. Introduction

For the past few months I have been studying the problem of how to make a computer understand linguistic information (in some generally accepted sense of "understand"). I have listened to courses on Linguistic Structure (Dr. Chomsky) and Mechanical Translation (Dr. Yngve), and read the works of various linguists, logicians, psychologists, and computer programmers on subjects ranging through semantics, information retrieval, and mechanical translation.

The remainder of this paper is divided into two parts: a survey of various ideas and results appearing in the current literature (with some editorial comment); and a proposal for future work to include a computer system for storing and extracting semantic information.

II. Background Literature

A. General Semantics

"Semantics" is generally studied from one of two viewpoints: pure or descriptive. Pure semantics, the kind studied by Carnap,[1]* deals with the properties of artificially-constructed formal systems (which may or may not have analogues in the real world), with respect to rules for sentence formation and designation of formal models and truth values. We shall rather be concerned with descriptive semantics, an empirical search for rules governing truth and meaningfulness of sentences in natural language.

We quickly encounter the paradox of having to use words with which to discuss the meaning of words, in fact even of the word "meaning". Any attempt to distinguish between object language and meta-language seems strained and artificial. A common device is to define "meaning" with a very specialized sense, or refuse that it can be defined altogether. Quine, tongue in cheek, recognizes this difficulty in the following paragraph [2]

"One must remember that an expression's meaning (if we are

* Subscripts refer to Bibliography entries

to admit such things as meanings) is not to be confused with the object, if any, that the expression designates. Sentences do not designate at all..., though words in them may; sentences are simply not singular terms., But sentences still have meanings (if we admit such things as meanings); and the meaning of an eternal sentence is the object designated by the singular term found by bracketing the sentence. That singular term will have a meaning in turn (if we are prodigal enough with meanings), but it will presumably be something further. Under this approach the meaning (if such there be) of the non-eternal sentence "The door is open" is not a proposition..." ---implying that the illusive meaning of "The door is open," is some complete intuitive set of circumstances surrounding a particular occasion on which the statement "The door is open," was uttered. Clearly this kind of concept does not lend itself to computer usage.

Ziff[3] is more precise in making the following distinction: words may have meaning (but not significance); utterances (phrases, sentences) may have significance (but not meaning). However he states that an analysis of the significance of a whole utterance cannot be completed without an analysis of the meanings of the words in the utterance.

Another similar approach is that of Ullmann[4] who considers a word as the smallest significant unit with isolated "content" whereas phrases, sentences, etc., express relations between the things which are symbolized by individual words. Here "meaning" is defined as "a reciprocal relationship between the name and the sense, which enables the one to call up the other." By "sense" is meant the thought or reference to an object or association (referent) which is represented by the word (symbol). Note that meaning here relates word with thought about object, not necessarily with object itself.

In this connection let us mention Walpole's "25 definition routes."[5] He observes that a word <u>may</u> be defined by direct symbolization, i.e. associating it with an observable object but also any other kind of association, connection, or characteristic, such as location or state or legal relation, can be used; and the precise connection is generally described verbally. Walpole also notes (and we shall use this fact later) that some word relationships, such as whole to part, or generalization to special case, determine partial orderings of large classes of words (especially common nouns) into tree structures. However the class of abstract nouns or "fictions," which do not name any object in any specific sense-experience, do not lend themselves to such ordering.

B. Grammar and Meaning

Thus far we have discussed meaning while ignoring the grammar (syntax) of language -- but clearly grammar must be considered since an ungrammatical sentence is not likely to be very meaningful.

A "grammar" is usually defined as a set of rules for generating the grammatical sentences of a language, and none of the ungrammatical ones. Deriving a grammar is an empirical business, since the ultimate test of whether a statement is grammatical or not is to ask a native speaker. Considering only the functional roles of words in sentences (their "parts-of-speech"), but not their meanings in any sense, Chomsky[6] develops various models for English grammar: phrase structure is a simple concept and works for a small part of the language, but is generally inadequate; transformational schemes are probably adequate, but are complicated and difficult to complete or test.

Although syntactics purports to ignore meanings, the boundary line between grammar and semantics is quite hazy. For example, some linguists classify the so-called "mass nouns"(e.g."water") as a separate grammatical group since they don't take the article; however the distinction between "I want meat" and "I want a steak" is generally considered to be a semantic one.

Ziff defines meaningfullness in terms of rigidity of grammatical structure. Words which are necessary in a particular grammatical configuration, such as frequent occurrences of "to," "do," "the," etc., are said to have no meaning. On the other hand words which could be replaced by a large number of alternatives within a given grammatical context are considered very meaningful. Simmons[7] makes this distinction between *function* words and *content* words even more sharp, as we shall see later. We shall accept these notions only to the extent that they do indicate a close connection between grammatical and semantic classifications.

C. Existing Computer Programs

Several computer programs have been written in the past few years which are relevant to the present discussion. Some of the features of these are summarized below.

1) Feldman's analysis for simple English. (8) This program, designed to translate simple English commands into tooling machine instructions, performs the necessary syntactic analysis on a certain class of simple sentences which are limited to a small, prescribed vocabulary. Vocabulary items are identified with part-of-speech, and certain are marked as "key" words.

2) Phillips' "Question-Answering Routine"[9] Again a dictionary of all needed words and their parts of speech is provided. Sentences are analyzed on the basis of a phrase-structure grammar, and placed in a canonical tabular form. The question is assumed to involve only certain factors, and is transformed to be matched against the sentences in the text, until the appropriate information is found.

3) Simmons' grammatical coding.[10] By making proper note of "function" words (prepositions, auxilliary verbs, conjunctions,etc.) which are keys to grammatical structure, suffixes, and empirical rules of grammatical context (allowable pairs and triplets of syntactic forms), this program can scan arbitrary text at a high rate of speed and tag each word with its appropriate part of speech, with a high degree of accuracy.

4) SYNTHEX.[7] This system, designed to "synthesize human language behavior," starts by scanning arbitrary text, performing the grammatical coding described in 3) above and indexing all occurrences of content (non-function) words. When a question is read, its content words are identified and relevant sections of text extracted by use of the index. An answer is then composed based on a matching process similar to Phillips', although probably somewhat more elaborate.

5) "Baseball"[11] This program, written in the IPL-V programming language, is designed to answer any reasonable verbal English question about the results of a set of baseball games. The data is a certain tree structure containing all the information about all games. Questions are analyzed syntactically with the aid of a dictionary, and the resulting forms completed by references to the data structure. Tabulated answer information is then printed out. The dictionary has a set of values of certain attributes for each word, such as part of speech, whether part of an idiom, and "meaning". "Meaning," which only appears for certain words, is canonical translation within the context of the program; e.g. the meaning of "who" is "Team-? " This procedure is adequate for the problem, but would be

difficult to generalize for application to wider contexts.

6) Sable's semantic structures.[12] Sable constructs a set-inclusion tree for storing items in a limited technical vocabulary. The information retrieval problem is simplified by using a simple scheme for coding the location of items on the tree by level and node count, and information about related items can be obtained from their tree locations. Our proposal below will involve a general application of similar semantic tree-structure principles.

It is interesting to note that all the above programs are based on algorithmic syntactic schemes. While it is presently generally recognized that use of semantic notions will be essential for difficult language-processing problems, useful specific programmable schemes based on semantic ideas have thus far been too elusive. The current state of this phase of the problem is discussed in the following section.

D. Applications of Semantics.

The necessity for having semantic information available has become most apparent in recent work on mechanical translation of natural language. Dictionary look-up and syntactic context schemes have failed miserably, and the only help a bilingual human can give when asked his translation method is to assert that he reads in one language, "understands" what was read, and then re-expresses the same "idea" in the new language. It is apparent that the same processes are involved in human reading comprehension and question-answering procedures and are therefore probably the key to efficient language processing by machine.

In order to understand "understanding," we should look at what is involved in learning new words. Quine[2] describes three basic approaches: 1)pointing in isolation ("stimulus meaning"); 2) from context (performing the necessary inductive inference); 3) by description (definition -- in terms of other words). I feel that while the first is the most fundamental, the third approach is by far the most important in building a vocabulary and recalling "meanings." Ziff[3] defines semantic relations as "...correlations between types of events, or a type of event and a state of affairs, or a type of word and a type of thing, or a word and a thing, and so forth." However since types and states are generally associated with specific words, all of the above except the last may be considered as in a language system. They enter into all kinds of groupings held together by a complex, unstable and highly subjective network of associations: associations between the names and the senses, associations based on similarity or some other relation. It is by their effects that these associative connections make themselves felt;...The sum total of these associative networks is the vocabulary."[4]

One way to deal with the problem of semantics might be to avoid it by translating ordinary language into a formal system which could be handled syntactically.[13] Unfortunately this procedure, if possible at all, would obscure the real problems in a mass of detailed documentation and notation, and be of little general value. At first view Freudenthal's LINCOS[14] may seem like a formal system for describing human behavior; but it is actually quite far from such, and assumes far greater abilities for inductive inference of rules and situations on the part of the receiver than is expected of the usual language student.

Quillian[15] is attempting to represent the semantic content of words as sets of "concepts", which could be combined to represent the meanings of phrases and sentences. With the basic premise that learning a new word involves measuring its values on a set of basic scales, he is trying to build up a repertoire of suitable coordinate scales (which are generally intuitive, unidimensional coordinates: e.g. length, time, hue, etc.), and code the corresponding representations of English words. He also permits defining words in terms of predefined words as coordinates. Some of this work is being programmed for the computer in the COMIT system. My feeling is that the relations between words is more important than the conceptual meaning of individual words in terms of something more basic (assuming each can be suitably found); and therefore a simpler approach which ignores base meanings might be more immediately fruitful.

Sommers[16] is more specifically concerned with permissable word combinations. He first describes a hierarchy of sentence types: 1) Ungrammatical; 2) Grammatical but nonsense; 3) Sensible but false; 4) True. He then argues that the crucial semantic distinction lies between the grammatical declarative sentences which are nonsense, and those which are significant (but may be true or false). Any pair of monadic predicates P_1, P_2 are said to have a sense value $\mathcal{U}(P_1,P_2)$ if there exists any significant sentence conjoining them. Otherwise they have value, $\cup$= $N(P_1, P_2)$. The relation is symmetric and is preserved under the usual logical operations on its arguments, but is not transitive. A stronger sense relation $Q \leftarrow P$ is true if "of (what is)P, it can be significantly said that it is Q...e.g. P = prime minister, Q = quick". This permits the arrangement of these "monadic predicates" into a simple tree, where all words in the same meaning class (e.g. all colors, or all words describing weight) occupy the same node. My main objection to this work is in where the important distinctions lie. Sommers would argue that "the idea is always green" is nonsense, but " the sky is always green" is sensible (since sky may have color, for "the sky is blue" and therefore "the sky is not blue" are significant), although false. Note that "ideas cannot be green" would be considered nonsense, rather than true, by Sommers. I feel the distinction between "nonsense" and "sensible but not true of the real world" to be too hazy to be the basis for a semantic system (such as that based on the $\cup$ and ← relations).

A new book by Nida[17] discusses several types of possible word associations. Chain associations, i.e. linear ordering of terms such as numerals or colors of the spectrum, have very limited applicability. Hierarchical ordering according to generality (class inclusion) is essentially one scheme which we shall adopt, with modifications. I don't believe as Nida does that pronouns have a place in the hierarchy as very general terms. Rather, they should be replaced by their antecedants. Constituent analysis is a scheme similar to Quillians, and again requires the difficult choice of coordinates and assignment of values. It is difficult to distinguish between Nida's "four basic types of semes as fundamental components in semantic structure" and certain well-known grammatical parts of speech-except that the rules for semantic parsing of sentences are much more obscure.

XII. Proposal

Let us now review the major points established in the preceding paragraphs:

1) Schemes based on the syntactic analysis of English language have been successfully programmed to solve certain limited linguistic problems.

2) Such schemes are not adequate, in themselves, for any larger, more general problems.

3) However, syntactic analysis (grammar) must be an important part of any more general method.

4) Understanding seems to be based largely on various associations between words.

5) None of the semantic analysis schemes proposed thus far can be realizable on a computer in the near future, largely because they involve procedures which are too vague and general, in connection with problems which are too large.

My feelings with regard to point 5) above are based on the following heuristic: "Large problems are easier to solve if the solutions to smaller, similar ones are available." Here "similar" is used in its broadest most ambiguous sense. If we are lucky the small problem will turn out to be a subproblem of the large one, or a special case with obvious routes to generalization (although this might not be apparent at the outset); however the solution will almost always suggest methods which should be tried on the large problem (or, just as important, methods which should not be tried).

I propose to develop a computer program which will have the ability to utilize semantic "background" information while answering questions based on available text material. Texts will be limited to a certain small vocabulary and simple sentence structures, probably selected from children's literature. Grammatical information such as phrase structure rules and the part-of-speech labeling available from Simmon's program[10] will be available to the system. The semantic information will be implicit in the organization of the "dictionary" which will contain divisions and associations of both syntactic and semantic natures, as described below.

Words will first be classified by part of speech. Nouns will then be arranged in a hierarchical structure based on generality class inclusions (this will probably involve re-entrant tree structure). Other associations will also be indicated between words in different parts of the tree, such as part-whole relationships, or just abstractly "related" (i.e. likely to appear in the same sentence, e.g. "dog" and "Bow-wow"). To some extent a verb tree can also be set up, with "move" branching into "walk," "hop," and "spin," for example.

Connections can then be made between nodes in the noun tree and nodes in the verb tree, indicating, say, "it makes sense to use any noun below this point as the subject (or object) of any verb below that corresponding point." Similar structures could be set up within the classes of adjectives and adverbs, and between them and the nouns and verbs which they may modify.

This semantic dictionary will be prepared in advance and available to the initial state of the program. It could conceivably be used then to generate random English sentences, all of which may be true of the real world. However, the question-answering procedure would be as follows:

1) As the text is read, a "thread" is inserted into the dictionary which follows the actual events of the text.

2) If "impossible" word relations or new vocabulary appear, the program will complain.

3) When a question is asked, the words in the question, the text thread, and the structure of the dictionary are all available to be used while composing an answer.

Ambiguous words, oblique meanings, and abstract concepts will be avoided or ignored. The purposes of this study will be

1) to determine whether it is possible to store a significant amount of semantic information, by means of the type of structure described, in a reasonable amount of space.

2) To determine the nature of the search and analysis procedures which are needed to utilize most efficiently this available semantic information.

I shall soon start preparing detailed flow-charts for this program. I would appreciate hearing whatever ideas or suggestions anyone might have.

Bert Raphael

Bibliography

1) Carnap, R. *Meaning and Necessity*, U. of Chicago Press, 1947

2) Quine, W. *Word and Object*, MIT Technology Press, 1960

3) Ziff, P. *Semantic Analysis*, Cornell U. Press, 1960.

4) Ullmann, S. *Words and Their Use*

5) Walpole, H.R. *Semantics: The Nature of Words and their Meanings*.

6) Chomsky, N. *Syntactic Structures*, Mouton and Co., 1957.

7) Simmons, R.F., et al. "Toward the Synthesis of Human Language Behavior," SP-466, Systems Development Corp., Santa Monica, Cal.

8) Feldman, C.G., "A Digital Computer Analysis Method for Simple English Sentences," MIT Electronics Systems Lab. Tech. Memo 6873-TM-9, 1959.

9) Phillips, A.V., "A question Answering Routine," AI Memo 16, 1960

10) Simmons, R.F. "A Computational Approach to Grammatical Coding of English Words," SP-701, SDC, Santa Monica, Cal.

11) Lincoln Laboratories, "Baseball: An Automatic Question-answerer." Proc. WJCC, May 1961.

12) Sable, J.D., "Use of Semantic Structures in Information Systems," Comm. ACM, Jan. 1962.

13) ACF Industries, Avion Div. "Translating from ordinary discourse into formal logic-- a preliminary study," Scientific Report AF CRC-TN-56-770.

14) Freudenthal, H., "LINCOS: Design of a Language for Cosmic Intercourse, North Holland Press, 1960.

15) Quillian, R., "A Revised Design for an Understanding Machine," RLE, MIT, 1962.

16) Sommers, F.T., "Semantic Structures and Automatic Clarification of Linguistic Ambiguity," International Electric Corp, Paramus, N.J., 1961.

17) Nida, "*Toward a Science of Translation* (in preparation).