[From: *Information Processing 1962*: proceedings of IFIP 62, Munich, 27 August – 1 Sept. 1962 (Amsterdam: North-Holland, 1963)]

# RULES OF INTERPRETATION — AN APPROACH TO THE PROBLEM OF COMPUTATION IN THE SEMANTICS OF NATURAL LANGUAGE

M. KAY
University of California, Berkeley, Cal., USA

## 1. Traditional Grammar And Descriptive Linguistics

It is commonly held that a complete description of any language must have two parts, grammar and vocabulary. The study of vocabulary is concerned with describing the primitive symbols used in the language and their relations to features of the external world. Grammar accounts for the ways in which these symbols are combined to form more complex structures. The aspects of language normally thought of as belonging to grammar, fall under three heads:—

i) Features of language, whether they be primitive symbols like words, or more complex structures like phrases and sentences, whose primary function seems to be that of expressing *relations* among other elements. On this basis, for example, a distinction is frequently made (between *full words,* which represent features of the external world, and *function words,* which serve to indicate relations between them.

ii) Linguistic elements, readily distinguishable by some peculiarity of form or usage, which make up closed and usually fairly small sets. The phonemes of a language form such a set and so do inflectional endings, articles, prepositions, auxiliary verbs and the like.

iii) Features which the usage of a particular language requires in certain situations and which are not, therefore, at the discretion of the speaker to include or leave out as he pleases. We may think of them as a sort of tax which the language levies on its users. It is, for example, difficult for speakers of English to talk about objects without mentioning whether one or more is in question, even though this information may have no direct relevance to the speaker's message. Similarly, an indication must be given, according to a standard formula, of the time of occurrence of an event relative to some frame of reference. The Chinese are exempt from both of these levies though they are subject to others. Often the required feature is of an entirely formal kind as in the case of gender in most of the Indo-European languages.

For any given language it is normally found that these three categories overlap heavily.

It is ascribed to the credit of *structural* or *descriptive* linguistics, that a much simpler scheme has been put in the place of the above. Many of the vague and ill-defined terms have either been defined more sharply or shown to be redundant. In short, a scientific discipline capable of holding up its head with the purest of natural science, now stands in the place of what was previously at best poetic. This paper puts forward a different view, maintaining that the linguists have vigorously thrown away at least one baby with the bath-water. A comparison of the old and new approaches shows that what we used to call *grammar* was, in reality, concerned with two very different types of phenomenon, one of which has become the proper study of descriptive linguistics, while the other came near to being banished into complete obscurity. At least one of the babies which went with the bath-water, though it may not have seemed very robust, deserved better than this Spartan treatment. I shall characterize those features of language which interest the descriptive linguist, as *rules of formation,* and distinguish them sharply from the outcast baby which is the subject of this paper and which I shall call *rules of interpretation.**

## 2. Rules Of Formation

The rules of formation of a language include all that it is necessary to know to distinguish an expression belonging to the language from one belonging to another language or to no language at all. They contain budget of primitive symbols and specify what sequences of these symbols are acceptable as expressions in the language. They may also go further to distinguish certain classes of expressions, such as sentences, but this is not necessary. The classic definition of rules of formation, due to Chomsky, is as follows:

> "The fundamental aim of the linguistic analysis of a language *L* is to separate the *grammatical* sequences which are the sentences of *L,* from the *ungrammatical* sequences which are not sentences of *L,* and to study the structure of the grammatical sequences. The grammar of *L* will thus be a device that generates all the grammatical sequences of *L* and none of the ungrammatical ones."[1])

For the purpose of our discussion, we may leave aside the question of whether a device which will generate the grammatical sequences of *L* is formally equivalent to a device which will recognize the grammatical sequences of *L;* in either case it is true that only

* These terms are not intended to recall the work of Carnap with which they are only remotely and fortuitously connected

rules of formation are involved. Both devices would require an inventory of rules of formation for the language, but that is not to say that they would require the same rules, because the rules of a language will, in general, not be unique. Rules of formation also have the following interesting property. Given a text in an unknown language, it is possible to start compiling a list of rules of formation. As more text is considered, the list will become more like the one required for the total language, and the probability that each rule is applicable to the total language will become steadily greater. Thus, deriving a set of rules of formation for a text is a problem of the classic black-box type. The rules of formation characterize a language as a formal system of a certain kind without regard to its function as a symbolic system. It is true that linguists use the fact that language is also a symbolic system in arriving at some of their conclusions, but the final statement involves only rules of formation.

## 3. Rules Of Interpretation

Consider now the case of an intelligent young Martian making his first trip to the Earth. He has prepared himself by memorizing a grammar of the English language written by the foremost American linguists, and also a compendious vocabulary. He knows what all the words in the vocabulary denote, and he can construct and analyse the syntactic structure of any English sentence. To his dismay, when he eventually arrives on Earth he finds that he is barely able to communicate. He can buy simple things in shops by pronouncing the one-word name of the thing he wants, but he is at a loss when the effect cannot be achieved with a single word. Why is this? Has he not learned the grammar so that he can produce faultless sentences? Unfortunately, all the books he studied were silent on the question of the meanings of phrases and sentences, so that he is quite unable to predict the consequences of uttering any expression, even though he knows it to be well-formed and that the meanings of the words are appropriate to the message he wishes to convey. He knows the words "red" and "brick" and what they mean, and the grammar tells him that the phrases "red brick" and "brick red" are both acceptable in English, but people respond to them in different ways. He may have observed that people occasionally, say "Please pass the salt", but he does not know why they choose that expression instead of "Please salt the pass", nor under what circumstances it would be appropriate to say the latter.

A rule which is used to derive the meaning of a phrase or sentence from the meanings of its constituents is a rule of interpretation. When we say that in English the word "red" may occur before "brick" to form the phrase "red brick", we are making a remark about the rules of formation of English; when we say that the word "red" may be used to qualify the word "brick", we are making a remark about the rules of interpretation of English. The particular rule which is invoked to interpret a phrase like "red brick" seems at first, to be very simple. A red brick is clearly that which is at once red and a brick. The phrase is applicable to those features of the world to which the words "red" and "brick" are individually applicable and the required result is given by a logical product. However, this implies that the relation of qualification is symmetrical and that "red brick" has the same meaning as "brick red".

A great many difficulties attend the study of rules of interpretation in natural language. It might seem that a useful first step would be to compile a list of some commonly used and easily identified rules. However, even so simple and informal an activity as this may be impossible, and would almost certainly produce more difficulties than it resolved. Consider the simple phrase "snow man". Clearly a snow man is a man made of snow. The qualified word names an object made of the substance which the qualifier names. This is a common situation and many examples can be adduced. An ice man, on the other hand, is not a man made of ice, but a man who delivers ice. The difference between "snow man" and "ice man" is greater than the differences between the meanings of the individual words would lead one to expect, and this may be accounted for by invoking two different rules of interpretation. However, there are at least two other equally plausible ways in which the difference could be accounted for. The first involves admitting "snow man" or "ice man" or both, as idioms, that is, as lexical primitives to the internal, constitutions of which rules of interpretation do not apply. Another possibility is to say that one of the meanings of "ice" is "who delivers ice" and one of the meanings of "snow" is "made of snow". In this case, a single rule of interpretation suffices to account for both phrases.

There is another strong objection to attempting to list rules of interpretation intuitively. The result of such an enterprise will be a list of phrases or phrase types against each of which one or more paraphrases will be entered. It will thus resemble, on the one hand, a traveller's phrase book and on the other, an ill-disciplined assortment of grammatical transformations. In any case, it can do little but complicate the early stages of the investigation. The rules applicable to a given phrase are sufficiently opaque in themselves to make it unwise to undertake their simultaneous study in pairs of partially parallel phrases.

## 4. A Model for Qualification

The approach advocated here, involves setting up a minimal calculus using, in the first instance, only one type of relation; that of *qualification.* This calculus is used as the basis for a model of the system rules of interpretation of a language. By causing the model to transform words into phrases and *vice versa,* and by making an intuitive judgement of the similarity in meaning between the words and the corresponding phrases, we hope to arrive at some estimate of the types of relations involved and their roles in the total economy of the system.

We shall use lower-case letters *(a, b, c...)* as variables. The constants of the system are those English words which may function as nouns or adjectives. The expression *ab* is used to stand for "*a* qualifies *b*". We shall write $ab = c$ when *ab* has the same meaning as *c.* We now set the following restrictions on the relation of qualification for the purpose of this model:—

i) The reflexive law: $aa = a$

ii) The anti-symmetric law: $(ab = ba) \rightarrow (a = b)$

iii) The transitive law does *not* apply: $(ab)\cdot(bc) \not\rightarrow ac$

iv) The several modifiers of a single item are unordered: $a(bc) = b(ac)$.

Given the two formulae

$$ab = c \quad (1)$$

$$\text{and} \quad dc = e \quad (2)$$

we may substitute $ab$ for $c$ in (2) and write

$$d(ab) = e. \quad (3)$$

Now consider a simple example. The following formulae are given initially:

$$\text{young human} = \text{child} \quad (4)$$

$$\text{male child} = \text{boy} \quad (5)$$

$$\text{male human} = \text{man}. \quad (6)$$

Substituting for "child" in (5), we obtain

$$\text{male (young human)} = \text{boy} \quad (7)$$

which, by (iv), we may rewrite as

$$\text{young (male human)} = \text{boy}. \quad (8)$$

Lastly, from (6) we may substitute "man" for "male human" in (8) and obtain

$$\text{young man} = \text{boy}. \quad (9)$$

A great many examples can be constructed in this way, some producing results that are intuitively acceptable and others producing nonsense. At this stage, our only object is to separate those cases where a particular relation holds, from those where it does not, in the hope that formal criteria will emerge for distinguishing them. If there are sufficient cases where the rule does apply, we may then go on to construct new rules to account for the residue and so on.

The main difficulty here, as in all language study, is that the volume of data to be examined is exceedingly large, and unless a systematic discovery procedure can be devised (using mechanical aids wherever possible) the study of rules of interpretation is likely to remain at the trivial level of our example. In this case, we must not be surprised if the notion that descriptive linguistics provides the only really serious approach to the study of language, continues to gain ground. Let us therefore, turn our attention to the possibility of constructing a mechanical model which will enable us to examine the consequences of a rule like the one we have suggested, when applied to large bodies of data. The model will be a very simple one.

## 5. Computation Methods

The relation of qualification as we have defined it, together with a great many others of the same general kind, can conveniently be embodied in a computer program similar to that developed by Feigenbaum for simulating verbal-learning behaviour[2]). Data are presented to the program in the form of equations or definitions, e.g. "child = young human" or "A child is a young human". Each word or phrase is associated with a node in a *net* which is *grown* in a piecemeal fashion to accommodate incoming data, fig. 1. The program is able to complete parts of the net which are not supplied by the data. The method is entirely straightforward. In fig. 1, the downward path from the node corresponding to "human being" to "boy" is by way of the lines marked "young" and "male", in that order. According to our hypothesis, qualifiers are unordered and it should therefore be possible to take the "male" path first. This would take us as far as "man". We may now construct a path from here to "boy" and mark it

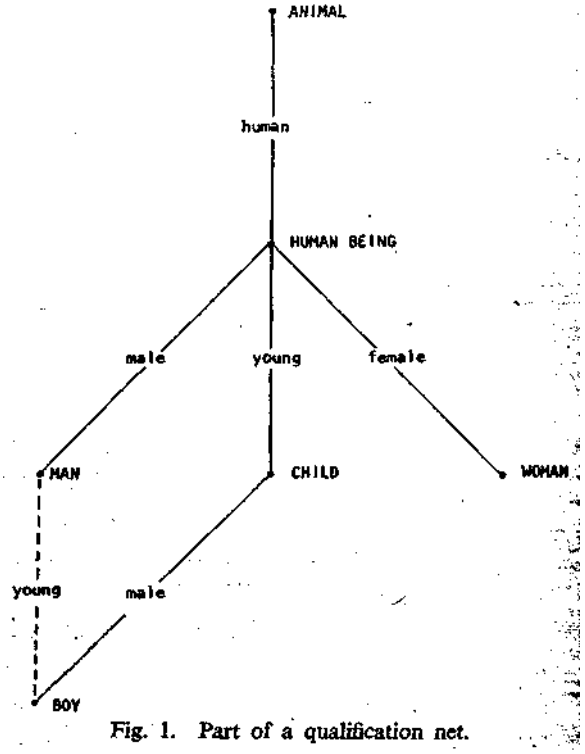


**Fig. 1. Part of a qualification net.**

"young". As, the net grows larger, the proportion of paths constructed by the program may be expected to increase, and it is on the basis of these paths that the efficacy of the rules will be judged.

There is not space here to discuss numbers of different rules of interpretation which might be found to be applicable in, say, English. At this stage it is more important to establish that they are as crucial a part of language as are rules of formation. The descriptive linguist, in so far as his interest in the language remains purely academic, is free to define the object of his study in such a way that rules of interpretation are excluded, but the development of machine translation is entirely dependent on their being well understood. A net of the kind described here, but also incorporating other rules, is in fact, directly usable as a dictionary for a machine translation program.

A net generated from a simple rule of interpretation such as the one we described, or from a number of such rules, turns out to have a number of features in common with the lattices of the Cambridge Language Research Unit. The lines in the diagram can, for example, be interpreted as inclusion relations in the straightforward sense that more specific terms fall under more general terms. If a word in a text is replaced by one occurring above it in the net, the sentence in which it occurs will differ in meaning from the original sentence only in being less specific. However, the lattice model was rejected because of the very small number of relations which could be defined in it, and in particular because of the impossibility of defining any but commutative relations. One of the great advantages of the lattice model is that it provides a convenient measure of

semantic similarity, namely lattice distance. This feature is preserved in the net.

## 6. Machine Translation

The theoretical implications of this kind of work will take a long time to emerge, and, as we have remarked. there are a great many questions which cannot be approached at this stage. However, the implications of a device like the semantic net may be much more immediate in machine translation. Every translator knows that the situation in which a word is best translated by a phrase or a phrase by a word, or a group of words is the rule rather than the exception. The most sophisticated approach to the problem of meaning which has so far gained any currency in machine translation is embodied in the list; a one-to-one mapping of the words of one language onto the words of another. If a phrase is to be reproduced by a word, or *vice versa,* the dictionary maker must have predicted it; he must have foreseen that the particular group of words might occur together and that they could best be rendered by a single word. The net provides a way in which a machine program could make substitutions of this kind without the dictionary-maker having foreseen the particular usage. Words and phrases of two or more languages can be stored at the nodes of the net. A node is created only when one of the languages has a single word to fill that position, but there is no requirement that each language should be represented at each node. The strategy of the translation process might then be somewhat as follows. The phrase structure of each input sentence is determined. The grammar which is used for this analysis also furnishes for each phrase an indication of the rule or rules of interpretation appropriate to it. The rules are then applied to the phrases and nodes in the net located for as many of them as possible. Here the strategy of the translation process was described in the last section. If, as is unlikely, a node is found to correspond to a complete sentence, and if the target language is represented at this node, then the sentence can be rendered by the single word. Otherwise, the immediate constituents of the sentence are examined until a word or phrase is found to represent every constituent. It may be that the level of the word is reached in some places without an equivalent having been found in the target language. In this case it is necessary to work up the semantic net still further to construct a phrase which can be used to represent the word; that is to say, a path leading downwards from a node where the target language is represented to the source language word. Machine translation research, however far it may still be from its final objective, has done a great service in revealing an important gap in language study. It is generally agreed that a sophisticated technique for syntactic analysis is one of the first requirements and it is admitted, however reluctantly, that the problem of meaning will have to be faced one day. What is missing is the bridge which links these two. I am claiming that the investigation of rules of interpretation is a matter of the greatest urgency, for providing as they do this link between syntax and semantics, grammar and vocabulary, they may be expected to have consequences for the study of rules of formation. In general, the phrase-structure grammar of a language is not uniquely defined. The decision as to which is the best grammar must be made, at least in part, on the basis of the rules which are required to interpret the phrases which it identifies. Therefore, the study of the rules of interpretation of a language cannot be made to wait until the syntax is complete, for this may make the formulation of the rules impossibly difficult. It is rather the case that the study of rules of formation and rules of interpretation should be conducted in parallel at every step.

## 7. Acknowledgements

I am deeply indebted to D. G. Hays of the RAND Corporation and M. Hollis of the University of California for their invaluable help in developing the ideas in this paper.

## 8. References

1) Chomsky, N.: “Syntactic Structures”, (Mouton, The Hague. 1957).

2) Feigenbaum, E. A.: *The Simulation of Verbal Learning Behaviour,* Proc. of the Western Joint Computer Conference, Los Angeles (1961).