MECHANICAL TRANSLATION

DEVOTED TO THE TRANSLATION OF LANGUAGES WITH THE AID Of MACHINES

VOLUME TWO, NUMBER TWO NOVEMBER, NINETEEN FIFTY FIVE

News

LONDON CONFERENCE

Three papers on mechanical translation were presented at the 3rd London Symposium on Information Theory, September 12th to 16th, 1955. One, by A. D. Booth and J. *P.* Cleave of Birkbeck College, is scheduled for the next issue of MT; another, by S. Ceccato and E. Maretti of Milan, will be published in an early issue of Methodos. The third, by V. H. Yngve, appears in this issue. Interesting discussions followed all three of the papers and they will be included in the proceedings of the conference which are to be published in 1956 by Butterworth's Scientific Publications.

M.T. GROUP AT CAMBRIDGE

We were happy to learn of activity in the field of mechanical translation at Cambridge, England. The Cambridge Language Research Group (not affiliated with the University) has been formed and had a three-day meeting, August 2nd to 4th. Members of the group are: E. W. Bastin, R. A. Crossland, M. A. K. Halliday, M. T. Hoskyns, M. Masterman, A. F. Parker-Rhodes, R. H. Richens, T. J. Smiley, R. H. Thouless, N. K. Willson and C. M. Winn.

The meeting was also attended by Prof. R. B. Braithwaite, J. Bronowski, A. F. R. Brown, J. Chadwick, Mr. Collinge, Prof. J. R. Firth, D.K. Ghosh, Mr. Hope, C. W. Kilminster, I. R. King, C. D. Parsons, Prof. C. L. Stevenson, and Prof. O. L. Zangwill.

M.I.T. APPOINTMENT

Y. Bar-Hillel, after spending two years at Hebrew University in Jerusalem, has returned to the U.S. for a year's stay at the M.I.T. Research Laboratory of Electronics. He will spend part of his time on the problem of the organization and mechanization of information search systems.

DEMONSTRATION MACHINE

A small German-to-English relay-type mechanical translator of limited capacity (with a stored vocabulary of sixty representative German words and their English equivalents, each containing not more than 7 Letters) is now under construction at the University of Washington, Seattle. Development of the machine has been a joint project of the Electrical Engineering and Far Eastern Departments. It is intended to demonstrate the mechanical identification of the constituents of German compound substantives and to exemplify the direct transfer from the input to the output of words shared by both languages. Both procedures are based on the research of Professor Erwin Reifler of the Far Eastern Department (see MT, Vol. 2, No. 1, July. 1955, and Machine Translation of Languages, pp. 147-148). They permit a substantial reduction of the stored vocabulary and represent *a* mechanical solution to the problem of input words not found in the memory. Associated with the project have been T. M. Stout, W. R. Hill, R. E. Wall, Jr., and R. S. Waggoner.

*Sentence-for-sentence translation**

Victor H. Yngve, Research Laboratory of Electronics and Department of Modern Languages, Massachusetts Institute of Technology

Introduction

Recent advances in linguistics, in information theory, and in digital data-handling techniques promise to make possible the translation of languages by machine. This paper 1 proposes a system for translating languages by machine — with the hope that when such a system is worked out in detail, some of the language barriers can be overcome. It is hoped, too, that the translations will have an accuracy and readability that will make them welcome to readers of scientific and technical literature.

Word-for-word translation could be handled easily by modern data-handling techniques. For this reason, much of the work that has been done up to this time in the field of mechanical translation has been concerned with the possibilities of word-for-word translation[2,3]. A word-for-word translation consists of merely substituting for each word of one language a word or words from the other language. The word order is preserved. Of course, the machine would deal only with the written form of the languages, the input being from a keyboard and the output from a printer. Word-for-word translations have been shown to be surprisingly good and they may be quite worth while. But they are far from perfect.

Some of the most serious difficulties confronting us, if we want to translate, arise from the fact that there is not a one-to-one correspondence between the vocabularies of different languages. In a word-for-word translation it is necessary to list alternative translations for most of the words, arid the choice among them is left up to the ultimate reader, who must make his way through a multiple-choice guessing game. The inclusion of multiple choices confuses the reader or editor to the extent that he is unduly slowed down, even though he can frequently glean the correct meaning after study. Another great problem is that the word order — frequently quite different in the two languages — further obscures the meaning for the reader. Lastly, there are the more subtle difficulties of idioms and the particular quaint and different ways that various languages have of expressing the same simple things. While it has been suggested in the past that rough word-for-word translations could be put into final shape by a human editor, the ideal situation is that the machine should do the whole job. The system proposed here is believed to be capable of producing translations that are considerably better than word-for-word translations .

The solution of the problems of multiple meaning, word order, idiom, and the general obscurity of the meaning when translation is carried out on a word-for-word basis is to be found in translating on a sentence-for-sentence basis. Nearly all of these problems can be solved by a human translator on a sentence-for-sentence basis. By this we mean that each sentence is translated without reference to the other sentences of the article. This procedure can be simulated experimentally by separating a text into sentences and submitting each for translation to a separate person who would not have the benefit of seeing any of the other sentences. In most instances an adequate translation of each sentence would result. Very little would be lost by discarding all of the context outside of one sentence length.

There are striking parallels between language and error-correcting codes. Language is a redundant code, and we are here proposing to deal with code blocks longer than one word, namely, with blocks of a sentence length. Our problem is to specify the constraints that operate in the languages out to a sentence length. This will be difficult because languages are so complex in their structure. However, we shall attempt to specify these constraints, or at least to lay the foundation for such a specification.

The Nature of the Process

A communication system may be looked upon as having a message source, an encoder, a statement of the rules of the code or a codebook for encoding, a decoder, a statement of the rules of the code or a codebook for decoding, and a destination. (See Fig. 1.) The function of the message source is to select the message from among the ensemble of possible messages. The function of the rules of the code or the codebook

---

*This paper was presented at the Third London Symposium on Information Theory, September 12 to 17, 1955. A shortened version with discussion will be published in the proceedings of the conference under the title *Information Theory* by Butterworths Scientific Publications in 1956. An earlier version of some of the ideas contained in this paper can be found in Chapter 14 of reference 2. This work was supported in part by the Signal Corps, the Office of Scientific Research (Air Research and Development Command), and the Office of Naval Research; and in part by the National Science Foundation.

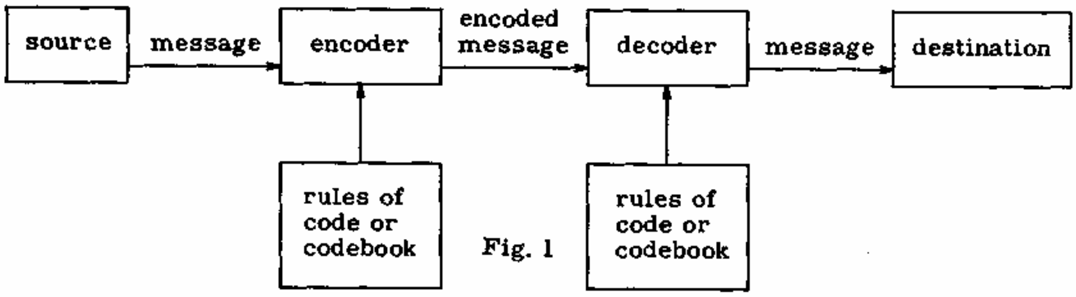

Fig. 1

is to supply the constraints of the code to which the encoded message must conform. In general, the encoded message is in a more redundant form than the original message. The function of the decoder is to recognize the features of the encoded message that represent constraints of the code, remove them, and supply the destination with a message that is a recognizable representation of the original message. This characterization of a communication system can be used with advantage to represent language communication only if great care is used in interpreting the various concepts. To this we shall now turn our attention.

In the case of language communication there is no difficulty in specifying what is meant by the concept of an encoded message if we restrict ourselves to the conventional written representations of the languages. Such written representations can be expressed in binary or other convenient form. What we might mean by "message, " however, is very difficult to specify exactly. Here we encounter some of the many difficulties with "meaning" that have plagued linguists. In the first place, it is very difficult to separate a message source from an encoder when the same individual performs both tasks. The message here would be, approximately, some representation of the "meaning" that the individual could express in the different languages that he might know; it would be something common to all of the different language representations. The message that arrives at the destination would be the receiver's understanding of the meaning, and might not, in fact, be the same as the message that left the source, but usually it is approximately the same if the individuals using the language understand each other. The decoder might not recover the original message, but another, and then there would be a misunderstanding. The decoder might extract a message quite different from the one intended by the message source, as a result of a confusion between message and constraints, and this might happen if the rules used by the decoder are not exactly equivalent to the rules used by the encoder. In this case, some of the constraints supplied by the encoder might not be recognized as constraints by the decoder, but interpreted instead as part of the message. For example, the encoded form of the message might be "Can you tell me where the railroad station is ?" and the decoder might extract such a message as "This person speaks English with an American accent." Or, as another example, the child who receives encoded messages in a language gradually accumulates information about the rules of the language and how to use it.

We now shift our attention from communication systems employing a single code or language, to systems which translate from one code or language into another. A code translation system can be looked upon as being much the same as the above representation of a communication system, but with the operations carried out in a different order; the positions of the encoder and the decoder are reversed. (See Fig. 2.) If the

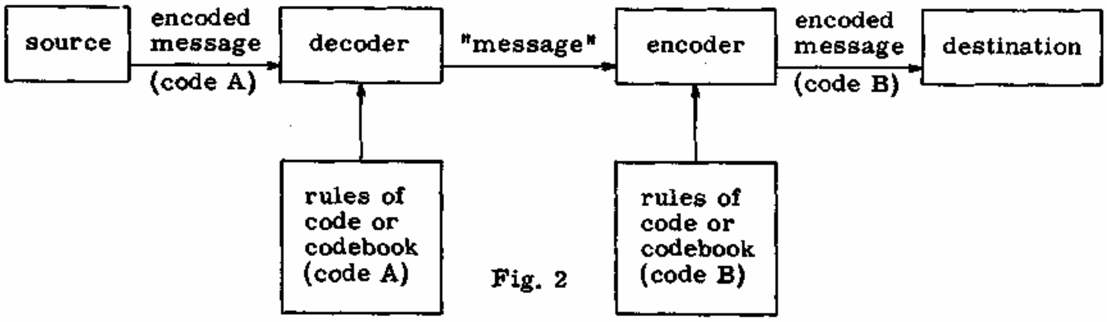

Fig. 2

codes are very similar, or in some sense equivalent, it may not be necessary to first decode and then encode. It may be necessary only to partially decode. If the two codes are very different, it may be simpler to decode to a minimally redundant form of the original message before encoding in the new code. We would like to consider the process of language translation as a two-step process: first, a decoding, or at least a partial decoding; then a recoding into another of the hundreds of known languages. The difficulties associated with word-for-word translations arise from the use of only a partial decoding, that is, a decoding based on the word instead of the sentence or some larger block.

We can assume that most material in science and engineering is translatable, or expressible in all languages of interest. An expression and its translation differ from one another in that they conform to the different constraints imposed by two languages. They are the same in that they have the same meaning. This meaning can be represented by some less redundant expression that is implicit in both language representations and that can be obtained by stripping off from one of them the trappings associated with that particular language. This representation might be called a transition language. Attempts at a specification of the structure of the "message" may get us into some of the difficulties associated with "meaning" but a description of the same thing as a transition language comes naturally from a description of the constraints of the two languages, since the transition language is just a representation of the freedom of choice left after the constraints of the languages have been taken into account.

Many of the constraints of language are quite constant. Grammar and syntax are rather stable. But there are other constraints that are peculiar to each user of the language, each field of discourse, each cultural background. A restriction can perhaps be made in mechanical translation to one field of discourse so that it will be easier to specify the constraints. Since language is a very complicated coding system, and in fact not a closed system, but an open one in that new words, constructions, and innovations are constantly being introduced by various users, the complete determination of the constraints is practically impossible. The best that one can do is to determine an approximate description of the constraints that operate; thus our translations will remain approximate.

What we mean by the concept of transition language in a language translation process can be illustrated by the word-for-word translation case. Booth[4] pointed out that one could not go directly from the words of one language to the words of another language with a digital computer of reasonable size, but that it would be more economical to go through the intermediate step of finding the addresses of the output words. These addresses are in a less redundant form than the original words, and for the purpose of this discussion they will be considered as the transition language. What we mean by transition language in a mechanical translation process is the explicit directions for encoding which are derived by the decoder from the incoming text.

The practical feasibility of mechanical translation hinges upon the memory requirements for specifying the rules of the code, or the structure of the languages. Word-for-word translation is feasible because present-day digital data handling techniques can provide memories large enough to store a dictionary. In other words, we can use a codebook technique for decoding and encoding on a word-for-word basis. If we want to translate on a sentence-for-sentence basis, we must find some method for specifying the structures of the languages which is compact enough to fit into practical memories. Obviously we cannot extend the dictionary concept by listing all of the sentences in the language with their translations. There are certainly in excess of 10^{50} sentences less than 20 words in length in a language like English.

Our problem, then, is to discover the constraints of the language so that we can design practical encoders and decoders. Our problem is that of the linguist who would discover such constraints by careful observation of encoded messages. The following example from coding will illustrate some important aspects of the problem of discovering constraints. We are given the data that the following four binary digit sequences are some of those allowed in the code. We are to determine the constraints of the code.

```
10101010    01001011
11100001    01100110
```

Here, as in the case of studying the structure of language, we do not have an exhaustive list of the allowed sequences. We can only make tentative hypotheses as to the exact form of the constraints and then see if they predict the existence of other observable sequences. Thus we might guess that one of the constraints in the code above is that the number of 0's and 1's is the same. The hypothesis will fall as soon as the sequence 00000000 is observed. Of course the linguist would make short work of the simple coding problem and would soon discover that there are only 16 different allowed sequences. If he were clever, he might deduce the rules of the code (the structure of the language) before he had obtained samples of all of the sequences. He might discover that the second four digits are identical with the first four digits if there is an even number of 1's in the first four; and that if the number of 1's in the first four digits is odd, the second four digits are the complement of the first four, formed by replacing 0's with 1's, and 1's with 0's. Having this specification of the rules of the code, he can say that it takes four digits to specify the message, the other four being completely determined by them. He might then say that we can take the first four digits as the message. He could equally well have chosen any four independent digits, such as the last four, or the middle four. This corresponds merely to assigning to the 16 messages 16 numbers in different order. The code has error-correcting properties, as does language. If one of the eight digits is in error, its location can be deduced by comparing the first four digits with the last four digits, and checking the parity of the first four. If there are two errors, either the first and last four digits differ in two places, or there are no differences, and the parity of the first four digits is odd.

The solution to our little coding problem is satisfactory in that we have a very compact statement of the constraints of the code. However, if we want to utilize the code in an actual communication channel, we have to design an encoder and a decoder. It may be that there are other simple statements of the rules that might be more suitable for the processes of encoding or decoding. In fact, there are other such representations, since the code above is equivalent to the Hamming code[5] of this length, for which the rules for encoding and decoding can be stated entirely in terms of parity checks. The code is also equivalent to the Muller-Reed code[6,7] of this length which uses a majority rule test in decoding. The three statements of the rules of the code are all valid. The choice of the representation of the rules of a language depends partly upon the use for which it is intended, and it is quite possible that one choice would be made for use in encoding and another choice would be made for use in decoding. In other words, the rules of a language may be phrased in a number of equivalent ways. For use in translating machines, they must be operational, that is, they must be appropriate for use in a machine that operates by a predetermined program[8].

The coding example given above illustrates five points about the language problems connected with mechanical translation. First, the rules of the code must be determined from an examination of the received messages. Second, there is no unique specification of the message. Third, there is redundancy which is useful for error correction. Fourth, there may be many equivalent formulations of the rules of the code. Fifth, the choice of a formulation depends partly upon the use for which it is intended.

If our purpose is translation, there is one further consideration. The choice of the form of the rules is also dependent upon which two languages are involved in translation and also in which direction translation is being carried out. It is very likely that the rules of English will have to be restated in various forms, depending on whether one wants to translate into German, out of German, into Russian, out of Russian, and so on. The reason is that certain relations can be found between different languages which can be used to simplify the process of decoding and encoding for the purposes of translation. The form of the transition language that forms the intermediate step in translation will be different with different language pairs.

We have pointed out that we want to translate on a sentence-for-sentence basis; that the feasibility of being able to do this depends upon whether or not we can state the structures of the languages in a form that is sufficiently compact for storing in a machine memory; and that the form of the statements of the structures must conform to certain other requirements, chief among them being that they be appropriate for use in decoders and encoders. We now proceed to discuss the problem of specifying language structure for use in mechanical translation processes.

Structure of Language from the Point of View of the Encoder

We want to consider, first, the form of the rules from the point of view of the encoder because they are simpler to explain and correspond more closely to other points of view commonly encountered. The encoder combines the message with the rules of the language in order to form the encoded message.

We want to limit the encoder to the words of the language. Of the various ways of doing this, perhaps the only one that seems feasible is to list the words of the language in a dictionary and to store this dictionary in the machine. Whether or not an attempt is made to reduce the number of entries in the dictionary by the use of a stem-affix routine — as is proposed by several authors — or by a method of splitting up compound words[9], depends upon whether it will be more economical to supply the required routine or to supply the additional storage space needed to list in full all of the words in their various inflected forms.

We want to encode in blocks of a sentence length. Since the words are to be listed in a dictionary, it seems appropriate to inquire whether a dictionary type of list could be used to assist in the encoding into sentences. It is certainly clear that it would be impossible to list all of the sentences of the language in a dictionary. In fact, an attempt to list all two-word sequences would require a dictionary of impractical size. The length of the list required to accommodate all structures of a code depends upon the redundancy of the structures, but more important, • upon the size of the signaling alphabet and the length of the sequences. The use of words as a signaling alphabet and the use of sequences of sentence length is completely out of question because of the practical impossibility of listing and storing enough sentences.

In order to reduce the signaling alphabet, the concept of part of speech is introduced. Larger structures are stated in terms of sequence of parts of speech instead of sequences of words. By the introduction of the concept of part of speech, we have factored the message into two parts. First of all, there is a sentence composed of a sequence of parts of speech, and the encoder has the opportunity of choice from among the various allowed sequences. Second, there is a further opportunity for choice front among the words that have the privilege of occurrence[10] for each part of speech. In language, these two possibilities for choice correspond to structural meaning and lexical meaning. As an illustration of structural meaning, take the sentence, "The man had painted the house." A German sentence with approximately the same meaning as the one above, translated on a word-for-word basis, would be, "The man had the house painted." Here the words are the same, but the structural meaning is different.

As an example of the economy introduced by the concept of part of speech, consider the Markov source (See Fig. 3.) which will generate over 10^{21} English sentences using a vocabulary of about 35 words. By the use of the concept of part of speech, whole lists of words are considered as equivalent so that with the 10 parts of speech there is only a small number of sentence types. It is estimated that there are millions of possible sentence types of which this diagram represents only a few. The structural meaning is indicated by the sentence type or the choice of path through the diagram, the lexical meanings are indicated by the further choice of the individual words from each list.

The introduction of part of speech and the factoring of the message into a lexical and a structural part has reduced the total number of the possible representations of sentences. The number of different structures, however, is still too large to list in a dictionary. The further step that we propose to take is to take advantage of regularities in the sentence types. For example, the first three states in the diagram (Fig. 3) and their connecting lines may be found included intact in many different sentence types and often more than once in a given sentence type. Just as we have grouped several words together to make a part of speech, we may group several paths together to form a phrase. If this program is carried out in its full elaboration, we are left with a number of intermediate levels of structure between the word and the sentence, such as various types of phrases and clauses. The levels are to be chosen in such a way that the total number of listed structures is reduced to a number that can be handled in a machine memory. Preliminary work seems to show that this can be achieved if the parts of speech number in the hundreds.

As an illustration of the use of an analogous level structure in coding, we can turn to the error-proof codes of Elias[11] . In these codes, "words" are formed according to some error-correcting code, such as one of those already mentioned, in which there are message digits and check digits. After a sequence of words has been sent, a phrase is made by adding a series

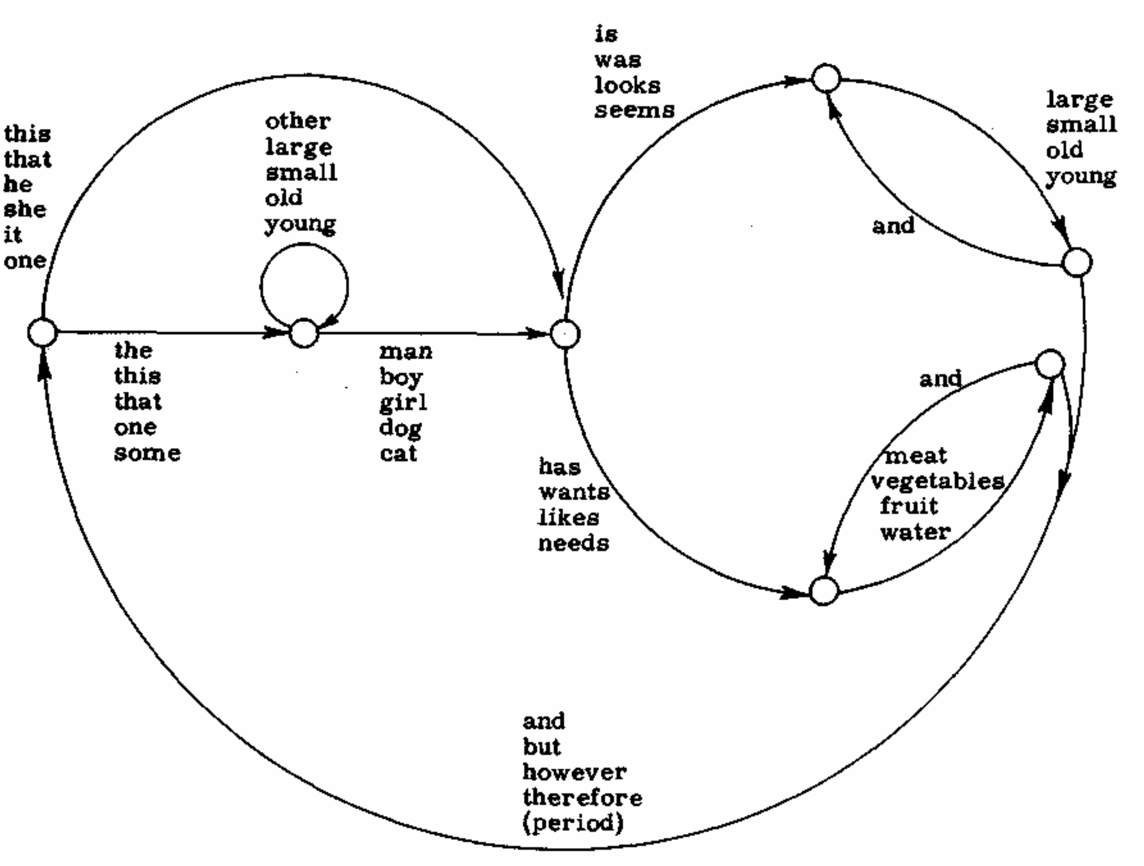

Fig. 3

of check words so that the whole structure has error-correcting properties on the phrase level as well as on the word level. The process is iterated as often as desired.

A somewhat closer analogy to language could be constructed by dividing the words into parts of speech (indicated, for instance, by the first digit so that we would have two parts of speech). A sentence of seven words in this code is represented by the seven rows of the diagram (Fig. 4). The structural meaning is indicated by the binary digits marked A, and these are checked by check digits marked B. The lexical meanings are indicated by the rows of III. In each word, AIII or BIII is checked by the digits C . In this code, the parts of speech are clearly and explicitly marked in the absence of noise by certain features (the first digit) in each word; in language, parts of speech are not always very clearly marked by grammatical affixes or the like. In language, there is no explicit separation into message symbols and symbols furnished by the constraints of the code, but our assumption that each sentence can be translated into another language leads us to look for an implicit separation .

A I I I C C C
A I I I C C C
A I I I C C C
A I I I C C C
B I I I C C C
B I I I C C C
B I I I C C C

Fig. 4

Our rules of language from the point of view of the encoder, then, are somewhat as follows. Select a sentence from among the sequences of clause types. For each clause type, select a clause from among the allowed sequences of phrase types. For each phrase, select a sequence of parts of speech. For each part of speech, select a word. In the translation process, the information required for the selections at each stage must be obtained from the decoder and may be called the "message" represented in the transition language.

Structure of Language from the Point of View of the Decoder

So far, the structure of language has been looked at from the point of view of the encoder which encodes in a given output language the "message" provided for it by the decoder. The rules for decoding language into some representation of the "message" are not just the reverse of the rules for encoding. If they were, mechanical translation would be much easier to accomplish than it appears to be. The difference between the point of view of the decoder and the encoder is just the difference between analysis and synthesis. The difference is illustrated in error-correcting codes that are easy to encode according to rules, but for which no rules are known for decoding in the presence of noise, although the message can be recovered by the use of a code book. In language, the difficulties in decoding are not the result of noise; they are the result of certain characteristics of the encoding scheme.

Decoding would be very simple with the error-correcting code using two parts of speech (Fig. 4). Decoding would be simple and direct because the part of speech of each word is clearly marked by its first digit. This is true to a certain extent in languages that have inflectional endings and grammatical affixes; more so in some languages than in others. Much attention has been paid to these affixes for purposes of mechanical translation. But the fact remains that even in the most highly inflected languages, the parts of speech are imperfectly indicated by affixes on the words. The problem is even worse than that: a given word form may belong to more than one part of speech, and there is no way at all to tell which part of speech it is representing in a certain sentence by looking at the word itself. The context, or the rest of the sentence must be examined. The lists of words that the encoder uses for each part of speech overlap, so that a given word may appear on several lists. In Fig. 3 it can be seen that several of the words appear in more than one list. The proper translation of these words into a language other than English requires a knowledge of the list from which the word was chosen. The decoder has this problem of deducing from which list the word was chosen. The statement that a word may belong to several parts of speech is just another way of saying that it may have several meanings. The concept of part of speech may be extended to include not only the usual grammatical distinctions, but in addition the distinctions that usually would be called multiple meanings.

Probably all languages exhibit the phenomena of multiple meaning, and one word making shift for more than one part of speech. It is interesting to speculate as to whether there is any utility to this phenomena, or whether it is just excess baggage, a human failing, another way in which our language does not come up to ideal. One word — one meaning would presumably make our language more precise and would eliminate the basis for many pointless arguments and much genuine misunderstanding. It has been proposed that language be changed to approach the ideal of one word — one meaning so that mechanical translation would be easier[12]. Some of the advantages accruing from the phenomena of multiple meaning might be as follows: There is an economy of the vocabulary because part of the burden of carrying meaning is transferred to the word sequence. The number of different structures available in a code goes as V^n, where V is the vocabulary size and n is the length of the sequences. In order to take advantage of the larger number of structures available, the words must acquire multiple meanings. There is the introduction of the possibility of the metaphoric extension of the meaning of words so that old words can be used for new concepts. There is the possibility of using a near synonym if a word with the exact meaning is not at hand, and of modifying the meaning of the near synonym to that intended by putting it in an appropriate context.

Since the lists of words for the different parts of speech used by the encoder overlap, there is the possibility that the same sequence of words may result from different intended structural meanings. In fact, this sometimes happens when the encoder is not careful, and we have a case of ambiguity. Sometimes the choice of an ambiguous sequence is intentional, and we have a pun. Puns, in general, cannot be translated, and we have to assume that unintentional ambiguity is at a minimum in the carefully written material that we want to translate.

The task of the decoder in a translation process is to furnish the information required by the encoder so that it can make the appropriate selections on each level of structure. This information is implicit in the incoming sequence of words and must be made explicit. The decoder is given only the words of the incoming text and their arrangement into sentences. It must reconstruct the assignment of the words to the parts of speech intended by the encoder, and must make the structural meaning explicit so that it can be translated. The decoder must resolve the problems of multiple meaning of words or structures in case these meanings are expressed in several ways in the other language. The decoder has available two things: the words, and the context surrounding each of the words. The appropriate starting point for describing the structure of language from the point of view of the decoder is to classify the words of the language and the contexts of the language. The classification proceeds on the assumption that there is no ambiguity, that the assignment of words to parts of speech can be done by the decoder either by examining the form of the words themselves or by examining the context.

The classification of the words must be a unique one. Each word must be assigned to one and only one class. These we shall call word classes. In order to set up word classes, we classify together all word forms that are mutually substitutable in all sentences and behave similarly in translation. In practice, one of the difficulties of making such a classification is the problem of how detailed the classification should be. Certain criteria of usage must be ignored or in the end each word class will have only one word in it. As examples of the sort of classification that is intended, "a" and "the" would be assigned to different classes because "a* cannot be used with plural nouns. "To" and "from" would be assigned to different word classes because "to" is a marker of the infinitive. "Man" and "boy" would be assigned to different word classes because you can man a boat. But "exact" and "correct" would not be separated merely because one can exact a promise but correct an impression. Preliminary experimentation has indicated that the number of word classes needed for translating the structural meaning is of the order of many hundreds.

The classification of contexts is very closely connected with the setting up of word classes. A sentence can be considered as a sequence of positions. Each position is filled by a word and surrounded by a context. Since we have classified words into word classes, each position in the sentence has associated with it a word class which can be determined uniquely by looking the word up in a special dictionary. The number of sentence length sequences of word classes is much fewer than the number of sentences. All sentences that have the same sequence of word classes are considered equivalent . The context of a given position in a sentence can be represented by the sequence of word classes preceding the position and the sequence of word classes following the position, but all within one sentence length. It is these contexts that we propose to classify. We classify together all contexts that allow the substitution of words from the same set of word classes. We thus have set up both word classes and context classes.

The relationship between the word classes and the context classes can be illustrated by a very large matrix. The columns of the matrix represent all of the word positions in any finite sample of the language. The rows of the matrix represent different word forms in the vocabulary of the language. Each square in the matrix is marked with an X if the word corresponding to that row will fit into the context surrounding the position corresponding to that column. All words that have identical rows of X's belong to the same word class. All contexts that have identical columns of X's belong to the same context class.

The word classes and the context classes can be set up in such a way that the sentence sequence of context classes contains just the information that we require for specifying the original parts of speech — and thus the structural meanings — as well as the information that we require for resolving many of the multiple meanings of the words and of the larger structures.

The structure of language from the point of view of the decoder is as follows. Words are listed in a dictionary from which we can obtain for each its assignment to a word class. Sequences of word classes are also listed in the dictionary, together with their designations in terms of phrase types. Sequences of these phrase types are also listed in the dictionary, and so on, until we have sentence types. The procedure for the decoder is to look up in the dictionary the longest sequences that it can find listed, proceeding from word class sequences to phrase sequences, to clause sequences and so on. At each look-up step, the dictionary gives explicit expressions that lead in the end to a discovery of the context classes of each position. From this we obtain, for each word, its original assignment to a part of speech, and the structural meaning. Thus we have the "message" or explicit directions for use in the encoder.

Conclusion

The mechanical translation of languages on a sentence-for-sentence basis is conceived of as a two-step process. First, the incoming text is decoded by means of a decoder working with the constraints of the input language expressed in dictionary form and based on word classes and context classes. The result of the decoding operation is a representation of the "message," which is just the directions that the encoder needs to re-encode into the output language by using the constraints of the output language expressed in dictionary form and based on parts of speech. An assessment of the worth or the fidelity of the resulting translations must await completion of the detailed work required to set up the dictionaries and to work out the system in all detail. It is certain that the resulting translations will be better than any word-for-word translations.

Acknowledgment

The author is deeply appreciative of the opportunity that he has had for discussing these matters with his colleagues at the Research Laboratory of Electronics, Massachusetts Institute of Technology. He is particularly indebted to R. F. Fano, P. Elias, F. Lukoff, and N. Chomsky for their valuable suggestions and comments.

References

1 An earlier version of some of the ideas contained in this paper can be found in Chapter 14 of reference 2.

2 Machine Translation of Languages, edited by W. N. Locke and A. D. Booth, The Technology Press of M.I.T. and John Wiley and Sons, Inc., New York; Chapman and Hall, Ltd., London (1955).

3 See various issues of Mechanical Translation, a journal published at Room 14N-307, Massachusetts Institute of Technology, Cambridge 39, Mass., U.S.A.

4 Page 45 of reference 2.

5 R. W. Hamming, "Error detecting and error correcting codes, " Bell System Tech. J. 31, 504-522 (1952).

6 D. E. Muller, "Metric Properties of Boolean Algebra and their Application to Switching Circuits, " Report No. 46, Digital Computer Laboratory, University of Illinois (April 1953).

7 I. S. Reed. "A class of multiple error-correcting codes and the decoding scheme, " Trans. I.R.E. (PGIT) 4. 38-49 (1954).

8 Y. Bar-Hillel, "The present state of research on mechanical translation, " American Documentation. 2, 229-237 (1951).

9 E. Reifler, "Mechanical determination of the constituents of German substantive compounds, " Mechanical Translation. II, No. 1 (July. 1955).

10 L. Bloomfield, Language, Henry Holt and Company, Inc., New York (1933).

11 P. Elias, "Error-free coding," Trans. I.R.E. (PGIT) 4, 30-37 (1954).

12 Chapter 10 of reference 2

An experimental study of ambiguity and context*

Abraham Kaplan, Department of Philosophy, University of California, Los Angeles

Ambiguity is the common cold of the pathology of language. The logician recognizes equivocation as a frequent source of fallacious reasoning. The student of propaganda and public opinion sees in ambiguity an enormous obstacle to successful communication. Even the sciences are not altogether free of verbalistic disputes that turn on confused multiple meanings of key terms.

Special importance attaches to ambiguity as a result of the growing interest in the possibilities of mass translation: rapid and routine translation of large bodies of material. The simplest expedient, as a first approximation, is word by word translation — a word for word substitution carried out by essentially clerical methods, very possibly by machine. But word for word substitution is hardly usable when the words of both languages are even moderately ambiguous.

It is a familiar fact that ambiguity of isolated words is reduced by the contexts of their occurrence. The total behavioral situation in which language functions is decisive in determining what will be communicated. For many problems, however, (and in particular, that of mass translation), the behavioral situation is not accessible. The 'context' (itself an ambiguous word) must here be taken to consist of the verbal setting in which the word to be interpreted occurs, i.e., the other words with which it is being used.

The problem of this study is to determine to what extent and in what ways verbal setting reduces ambiguity. Is ambiguity primarily a feature of words in isolation, or does it persist to some extent even in context? What part of the context is most effective in reducing ambiguity — for instance, how is the ambiguity of a selected word affected by the words immediately preceding and following it, as compared with the effect of the entire sentence in which it occurs? Does it matter whether the immediate context consists solely of particles ? How is the reduction in ambiguity affected by the linguistic sensitivity of the translator? By the multiplicity of senses of the isolated word? By the clarity of the word; that is, the ease with which its multiple senses are identified? These are the questions to which this study is addressed.

Two important restrictions on this study are to be noted.

In the first place, it deals with ambiguity of single words, not homonyms (word types, not word tokens[1]): the four letters "blow" actually may constitute a single word, semantically and grammatically speaking, or may be one of several homonyms — a) to send forth a current of air, b) a wind or gale, c) a blossoming or blooming, or d) a forcible act or effort. There is no doubt that the setting usually allows us to distinguish nouns from verbs, for example, hence among homonyms which are different parts of speech. The problem here will be to distinguish the multiple senses of a single word. For instance, the verb "blow" has several senses: a) producing a noise by blowing, b) panting or puffing, c) talking loudly or boastfully, and so on. These are related senses, and as a group quite distinct from the senses of the homonym "blow" which means "to blossom." The ambiguity with which this study is concerned is thus more subtle than homonymy. Whatever analysis is to be given of the distinction between homonyms and single words, it is reasonable to suppose that the effect of context on homonym-ambiguity is more marked than that of the single-word-ambiguity here dealt with.

A second restriction on the study is this. It is not concerned with what ambiguity actually occurs in written material. The attempt is to determine the reduction of ambiguity by context, and not the actual frequencies with which ambiguities and their reductions occur. To be sure, the material selected is presumed to be sufficiently representative of actual discourse to make the results of practical relevance. But this presumption is not itself being tested here. All the cases studied are actual cases; the contexts were selected from published texts and were not constructed for the study. Nor were words selected on the basis of the kinds of contexts in which they occurred, except for certain formal requirements described below.

Procedure

A group of "translators" was presented with a set of words, each with a number of possible meanings to be judged applicable or not. The words were first presented in isolation, then in certain standard contexts.

---

*Reprinted with permission of the Rand Corporation from their report P18, dated November 30, 1950, which has been out of print for several years.

[1] For a discussion of this distinction, and a comprehensive survey of contemporary semantics, see C. W. Morris, Signs, Language, and Behavior, 1946.

The sample was derived entirely from the literature of pure and applied mathematics. This selection was made partly because of the background of the translators used in the experiment, partly because it is commonly supposed that such material involves less ambiguity than non-scientific writing, or even that of some other scientific disciplines. The specific books used are as follows:

	No. of Samples
Alexander, J., *Colloid Chemistry.* Vol. III, Chemical Catalog Co., 1931	15
Holmboe J. et al., *Dynamic Meteorology*, Wiley, 1945	15
Lefschetz S., *Introduction to Topology* Princeton, 1949	9
Moulton, F. R., *Introduction to Celestial Mechanics*, Macmillan, 1914	15
v. Neumann J and Morgenstern,O., *Theory of Games and Economic Behavior,* Princeton, 1947	15
Richter W., *Fundamentals of Industrial Electronic Circuits,* McGraw Hill, 1947	15
Stuhlman O., *Introduction to Biophysics,* Wiley, 1948	14
Weyl H., *Philosophy of Mathematics* and Natural Science, Princeton,1949	12
Williams C.D. and Harris E. C., *Structural Design in Metals,* Ronald Press, 1949	15
Zemansky, M. W., *Heat and Thermodynamics,* McGraw Hill, 1943	15
Total	140

The contexts were provided by sentences selected at random from these books, not drawn, for example, solely from prosy introductory chapters. On the other hand, "symbol-heavy" sentences which would require either specialized knowledge or considerable portions of text for their interpretation were omitted. Sentences were selected to vary in length from 15 to 40 words; occasionally, dependent clauses irrelevant to the clause in which the key word occurred were omitted. The distribution of sentence lengths was:

Number of Words	Number of Sentences
15 - 19	33
20 - 24	56
25 - 29	39
30 - 34	8
35 - 39	4
Total	140

The key words selected were limited to nouns, verbs, and adjectives; these are the major carriers of the content of any discourse, and probably more markedly exhibit ambiguities. The position of the word in the sentence was varied at random, to avoid overemphasis on the special contexts constituted by opening and closing phrases. The first and last two words of the sentence were never selected, so that contexts could be restricted to a single sentence. No mark of punctuation was allowed to occur within two words on each side of the key word, so as to simplify the appraisal of the effect of verbal setting. Only words of sufficiently general use to be included in the Fifth Edition of *Webster's Collegiate Dictionary* were chosen; and it was required that the dictionary distinguish at least three senses of the word.

Although frequency of use was not a criterion of selection, it was afterwards found that all of the 140 words selected appear in *The Teacher's Word-Book of 30,000 Words*.[2] Seventy-four of the words are among the thousand most frequent words in the English language; of these, forty-four are among the first 500. The following is the frequency of occurrence per million words in the Thorndike-Lorge count:

Frequency	Number of cases
Over 100	76
50 - 99	31
25 - 49	18
2 - 24	15
Total	140

The actual key words used in the sample are listed in Table I.

For each word, a number of possible senses was listed, obtained from the dictionary entry for that word. The fully inflected form of the word was used — e.g., the plural or past tense if this was the form of its occurrence. It was required that the senses listed be clearly distinguishable (in the judgment of the experimenter) from one another; this did not by any means coincide with the numbered senses in the dictionary entry. Obsolete, archaic, colloquial, and highly technical senses were omitted. A maximum of ten senses was selected. Wherever necessary, the total number of senses was made up to ten by adding an appropriate num-

[2] By E. L. Thorndike and I. Lorge, Columbia University Press, 1944.

TABLE I

Key Words Used

appear
approaches
assume
attached
balance
bears
broad
care
case
cells
change
character
class
classical
clear
close
come
compose
conceived
conditions
connections
consideration
contain
contracts
converted
course
current
cycle
deductions
degree
depending
determined
developed
device
diaphram
direct
dropped
due
elements
established
eye
field
flow
force
formal
found
free
function
general
generation
given
goes
good
ground
heads
heat
induced
introduced
leading
levels
lies
little
load
lower
maintained
make
mass
material
model
motion
narrow
nature
new
normal
note
numbers
observed
origin
part
particle
passes
people
period
phase
place
point
position
possesses
power
produce
product
projection
properties
protection
provides
put
raised
reached
reaction
reference
relations
requires
rest
rise
runs
scale
screen
separated
serve
set
shank
shape
show
skin
slight
solution
spirit
spread
state
strong
study
subject
substance
survey
system
tension
terms
tests
time
tool
transmitting
treated
tubes
types
used
value
view
words
work
world

ber of "false" senses, obtained from dictionary entries for words of the same part of speech. The average number of "correct" senses of the words in the sample was 5.6, approximately the degree of ambiguity in actual discourse.[3] The distribution of words in the sample with various numbers of senses was:

Number of Senses	Number of Words
3	16
4	33
5	30
6	25
7	7
8	14
9	5
10	10
Total	140

---

[3] See G. K. Zipf, Human Behavior and the Principle of Least Effort, Addison-Wesley Press, 1949, p. 30.

Examples of words with the senses listed (including the "false" ones) are given in Table II, below.

The study was carried out with the help of seven "translators", four of whom had considerable training in the mathematical sciences, the other three having only a high school education.

Words were first presented in isolation — the so-called <u>null context.</u> Each translator indicated which of the ten senses for each word appeared to him to be senses in which the word might sometimes be used. In the second phase, seven contexts were employed, derived from the sentence of the actual occurrence of the word. These contexts were:

the word preceding (P1)
the word following (F1)
both of these (B1)
the two words preceding (P2)
the two words following (F2)
both of these (B2)
the entire sentence (S)

TABLE II

<u>Examples of Words and Senses</u>

Starred senses are actual ones. (Of course, no stars were printed in the sheets from which the translators worked.)

<u>appear</u>

1) shine faintly
*2) be obvious or manifest
*3) come before the public
4) come or go near
5) be in great plenty
*6) attend before a tribunal
*7) seem, look
8) pass or move suddenly or quickly
*9) become visible
10) look steadfastly; meditate

<u>approaches</u>

*1) approximations
*2) preliminary steps
3) summaries, epitomes
4) suppressions, suspensions
5) wants, lacks
*6) ways, passages
7) posterior sections
8) dwellings, sojourns
9) skills
*10) advances

<u>assume</u>

1) snatch, seize
2) derived by reasoning or implication
*3) suppose
4) come into possession of
*5) undertake
*6) appropriate, usurp
*7) feign, sham
8) swallow eagerly
9) hold in possession or control
*10) receive, adopt

Words were presented to the translators in one or another of these contexts, and acceptable senses were again indicated by them. The design used had the properties that each translator was presented with all the words in some context or other; each word appeared in all the contexts; each context had all the words in it; and no person faced the same word in more than one context. Thus each subject made two interpretations of each word: once in the null context, and once in some verbal setting.

<u>Results</u>

The <u>accuracy</u> of a translator was measured by the number of his correct characterizations of a listed sense as actually belonging to the word or not: ascriptions of true senses plus denials of false senses. (This measure could be used only for the null context, where the true senses are specified by the dictionary; no such standard is available for occurrences in context.) The seven translators ranged in mean accuracy for all the words from 62% to 84%, around a mean of 75%. The four trained in mathematics averaged 80% accuracy, the other three 70%. Since the isolated words are not distinctively mathematical, the difference is presumably due to general linguistic facility.

The <u>clarity</u> of a word is defined as the mean accuracy attained on it by the seven translators. (Like accuracy, therefore, it applies only to the null context.) The mean clarity for all the words words was 75% (being linked to the mean accuracy). The distribution was:

Clarity (%)	No. of cases
40-49	1
50-59	4
60 - 69	29
70 - 79	57
80 - 89	41
90 - 99	8
Total	140

Unclarity was not due markedly either to a failure to recognize true senses or to a tendency to ascribe false ones. The mean number of true senses was 5.6; of assigned senses, whether true or false, 5.5. Clarity did not show any significant correlation with ambiguity: words with a large number of true senses were, on the whole, neither more nor less clear than those with a small number. Neither was clarity correlated with familiarity, as measured by frequency in the Thorndike-Lorge count. In both cases the correlation was + .1 and not significant.

By the reduction of a context will be meant the ratio of the number of senses assigned to a word occurring in that context to the number assigned to it in the null context by the same translator. The lower this ratio, the more effective is the context in reducing ambiguity. The reduction of the contexts tested was found to be:

Context	Reduction (%)
P1	75
F1	57
B1	47
P2	50
F2	56
B2	44
S	47

The context consisting of one preceding word appears to be least effective in reducing ambiguity, being significantly worse than one word following. One word on each side of the word to be translated is more effective than two preceding or two following. It is noteworthy that two words on each side of the key word are comparable in effect to the entire sentence. The distribution of the various degrees of reduction for each of the contexts is given in the following table.

Reduction (%)	Percent in Context						
	P1	F1	B1	P2	F2	B2	S
0 - 29	37	41	41	38	36	51	60
30 - 59	19	25	28	28	27	27	24
60 - 89	18	14	17	18	22	6	4
99 - 100	11	9	9	10	4	6	4
over 100	15	11	5	6	11	10	8
Total	100	100	100	100	100	100	100

What is the effect of initial ambiguity on its reduction? Do more ambiguous words profit more from context than less ambiguous ones? To answer this question, words of from three to five true senses were separated from those of six to ten: there were 79 cases in the former group, 61 in the latter. The reduction effected by each context for these two groups of words was found to be:

Context	Reduction (%) for less ambiguous words	Reduction (%) for more ambiguous words
P1	65	88
F1	62	51
B1	48	45
P2	56	43
F2	52	61
B2	44	44
S	47	47

As can be seen, there was no consistent direction of difference: the mean reduction was 53.4% for the less ambiguous words, 54.1% for the more ambiguous. It is to be noted that P1 again appears as the worst context; B1 as quite good, and B2 comparable in effect to that of the entire sentence.

The same procedure was used to appraise the effect of clarity on reduction of ambiguity. The sample was evenly divided into words of relatively high and low clarity, as defined above, and reduction separately computed:

Context	Reduction (%) for clear words	Reduction (%) for unclear words
P1	88	62
F1	53	62
B1	47	47
P2	49	52
F2	5?	59
B2	48	41
S	58	36

The effect is again not a consistent one, though it suggests some slight advantage to the initially unclear words, as profiting more from context. The mean reduction was 56.6% for the clear words, and 51.3% for the unclear.

The effect of familiarity was appraised in the same way. The seventy-four words which, according to the Thorndike-Lorge count, are among the thousand most frequent in the English language were separated from the remaining sixty-six words in the sample, and reduction again separately computed:

Context	Reduction (%) for frequent words	Reduction (%) for infrequent words
P1	89	59
F1	56	59
B1	49	44
P2	40	62
F2	59	52
B2	44	45
S	51	43

Again there is no consistent effect, though again there is some slight advantage for the less frequently appearing words, their mean reduction being 52.0% as compared with 55.4% for the more frequent ones. It is quite in accord with expectation, of course, that the less clear, less familiar words should profit more by being put in context than those that are clear and familiar to start with. But the results can only be said to be compatible with this expectation, and scarcely to confirm it.

By contrast with these slight effects of doubtful significance are two other factors which appear to be quite important in reducing ambiguity. The first is the semantic content of the context. A context might consist entirely of articles, prepositions, conjunctions, etc., and could be expected to contribute less to a translation than one which also contained words not so poor in semantic content. We may call the first particle contexts, the second substantive contexts. A context was classified as "substantive" if at least one word in it was not a "particle" word. The full list of words in the sample regarded as "particles" (not grammatically, but from the viewpoint of semantic content) is given in Table III, below. The results were the following:

Type of Context	Particle Contexts		Substantive Contexts	
	No. Cases	Reduction (%)	No. Cases	Reduction (%)
P1	89	80	51	66
F1	107	66	33	28
B1	67	54	73	40
P2	56	61	84	43
F2	62	62	78	51
B2	25	45	115	44
S	0	—	140	47

The effect is consistent and unmistakable. The mean reduction for the particle contexts was 61.3%, for the substantive contexts, 45.6%. How effective a context is in reducing ambiguity is a function, therefore, of whether it itself has a semantic content or is functioning primarily syntactically. It is noteworthy that for the B2 context there was no significant difference in reduction; but the small number of cases of B2 particle contexts (25) makes this result suspect.

A second markedly significant factor in reduction of ambiguity by context is the accuracy of the translators. The samples translated by the three most accurate and those by the three least accurate (for the words which they were each interpreting in the context in question) were grouped separately, there being sixty cases for each group. The results were:

Context	Reduction (%) for inaccurate translators	Reduction (%) for accurate translators
P1	109	59
F1	67	51
B1	58	46
P2	57	48
F2	63	52
B2	60	36
S	76	26

TABLE III

List of "Particles"

a
above
against
all
an
and
are
as
at
be
behind
between
by
can
certain
does
done
during
for
from
has
if
in
into
is
it
its
just
let
many
may
must
near
no
not
of
on
one
only
or
other
our
out
over
quite
same
several
shall
since
so
some
than
that
the
their
there
these
they
this
through
thus
to
under
until
us
very
we
when
which
whose
will
with
within
would

The effect is again unmistakable. The inaccurate translators showed a mean reduction, for the various contexts, of 70.0%, while the accurate translators attained a reduction of 45.5%. In the sentential context, the reduction of the accurate group was about three times as great as that of the inaccurate group.

In terms of these two important factors, an appraisal can be made of the optimal reduction of ambiguity by context, considering only the accurate translators, working with substantive contexts. The results are:

Context	No. Cases	Reduction (%)
P1	24	40
F1	13	35
B1	35	33
P2	38	39
F2	29	42
B2	53	36
S	60	26

Conclusions

1. Even for familiar words, no more than about 3/4 of the possible meanings presented are correctly translated as senses in which the words might sometimes be used.

2. The accuracy of such translation varies significantly from person to person, and shows some relation to educational level. Whether this is due to language ability, intelligence, or some other factor was not investigated.

3. There is no consistent direction of error in translation: false senses are as likely to be ascribed to words as are true senses to be unrecognized,

4. How accurately, on the whole, a word is translated bears no marked relation to the number of its actual senses nor to the frequency (within a fairly wide range) of its occurrence in actual discourse.

5. The verbal setting with least effect on reduction of ambiguity is the one word preceding the word to be translated. The greatest effect is that of the entire sentence in which the word occurs.

6. A context consisting of one or two words on each side of the key word has an effectiveness not markedly different from that of the whole sentence.

7. The most important factors affecting contextual reduction of ambiguity are the accuracy

of the translators and whether the verbal setting includes words other than particles. The most practical context is therefore one word on each side, increased to two if one of the context words is a particle.

8. Under optimal conditions (most accurate translators, non-particle contexts, at least one word on each side of the key word) ambiguity is reduced to from 1/4 to 1/3 of the number of senses assigned to the word in isolation. A short verbal setting therefore reduces average ambiguity from about 5 1/2 senses to about 1 1/2 or 2.

ERRATA

from "Mechanical Determination of the Constituents of German Substantive Compounds" by Erwin Reifler, MT 2, 1.

Professor Reifler has called to the attention of the Editors certain errors which slipped into his article. Corrections are listed below, with our apologies:

p. 5, col. 2, 13 lines from bot., read: "The ambiguity is here due to anteposition which,..."
p. 6, col. 1, 1. 21 & 22, instead of "Most capitalized forms. . .as follows," read "The capital memory will therefore contain only:"
p. 6, col. 1, 12 lines from bot., instead of "grass," read "gross"
p. 9, col. 1, 1. 9 read (cf. 3,5)
p. 10, col. 1, 1. 11 for "Vorhange," read "Vorhänge
1. 14 for "Rot," read "Rot-"
1. 15 for "grün," read "-grün"
1. 18 for "Mitburger," read "Mitbürger"
p. 11, col. 1, first line, over "(Senn/", read PLT
1. 19 for "cf. 7/10," read "7/8 & 10"
1. 31 for "P(PLTH)," read P(PLTX)
p. 12, col. 1, 1. 16 for "Arbeiterin-vasion," read "Arbeiterin-vasion"
col. 2, 20 1. from bot. for "cf. 5/Db)", read "(cf. 5/d2)"
p. 13, col. 1, 5 1. from bot. for "largest," read "longest"

bibliography

Erwin Reifler 56
Mechanical Determination of the Constituents of German Substantive Compounds
Mechanical Translation, Vol. 2, No. 1, pp. 3-14 (July 1955)

The paper demonstrates with examples from the German language how compound words can be mechanically "dissected" into their immediate *linguistic* constituents even though no graphic indication of their inner boundaries is present. This is particularly important in the case of extemporized compounds which, being unpredictable, can not be coded into the mechanical memory. The identification of compounds via the mechanical identification of their constituents, all of which are entered in the mechanical memory, is applicable to all compounds of all languages. The number of entries in the mechanical memory can be substantially reduced which, in turn, makes possible a decrease in access time and engineering costs. Considerations of source-target semantic intelligibility and accuracy are given precedence in this solution.

Reifler — Mechanical Determination (cont.)

Author

James W. Perry 57
Translation of Russian Technical Literature by Machine - Notes on Preliminary Experiments
Mechanical Translation, Vol. 2, No. 1, pp. 15-24 (July 1955)

See Abstract number 25, Vol. 1, No. 1.