[From: *Documentation in Action*, (eds) Jesse H.Shera, Allen Kent,
James W.Perry (New York: Reinhold, 1956)
Chapter XXIII

Invitational Meeting on Machine Translation

JAMES W. PERRY

Introduction

The practical utilization of recorded knowledge often requires its communication between persons who read and write different languages. In dealing with the translation problem which thereby arises, a variety of solutions are possible or conceivable.

Perhaps the simplest solution, in theory at least, would be to adopt some one language as standard throughout the world for the purpose of communicating recorded knowledge. This would mean, of course, that persons who do not know the selected language would have to learn it well enough to read and to write it. This approach would place most of the translation burden on the authors of written documents. At various stages in history one language or another has been widely used for written communication by persons who may not speak it as their mother tongue. Thus, during the Middle Ages, Latin was widely used for learned communication. More recently, particularly in the latter part of the 19th century and the early years of the present century, there was a tendency for German to dominate the scene in the field of chemistry. In more recent years, there has been a trend to ever wider use of English, particularly in the fields of science and technology. Such trends have not, historically speaking, worked out to the logical conclusion of universal acceptance of some one language. The rise of modern science was accompanied by increasing use of the vernacular tongues as media for communicating results of observation and research. The defeat of Germany in World War I ushered in a decline in the relative importance of German in chemistry and other scientific fields. Perhaps English will manage to achieve the universal acceptance which Latin was unable to maintain even within the confines of Europe, and which German only temporarily and to a limited degree attained. Rapidly increasing interest in science, technology, and industry in the heavily populated countries of the orient, may prove in our day to be an insuperable barrier to the universal adoption of English. In those

countries under Soviet domination, there is little evidence of any trend toward favoring English at the expense of Russian, Rather, the contrary is to be observed with Russian finding powerful support both politically and culturally by the Soviet State, particularly in its satellites.

Artificial languages, such as Esperanto and Interlingua, have at no time achieved the measure of international acceptance enjoyed at different times by Latin, German, and English,

Solution of the translation problem has been achieved, at least partially, by requiring candidates for advanced research degrees, particularly the Ph.D., to acquire a reading knowledge of German and French or possibly some other languages of importance to their chosen field of specialization. At least two factors have tended to discourage this approach during recent years. One of these is the ever-widening range of knowledge in every field of science and learning. With so much to be read in English, graduate students are likely to underestimate the advantages and benefits of being able to comprehend publications in foreign languages. Another factor that also tends to discourage graduate students is the widening circle of foreign languages that have become or are becoming important in science, technology and other fields of learning. As a consequence, enthusiasm for language study on the part of candidates for advanced degrees has tended to decline. Furthermore, learning to read German and French no longer provides as thorough a degree of coverage of foreign material in many fields as was the case twenty or thirty years ago. The rise of other languages, e.g. Russian, to positions of importance for learned communication can scarcely be expected to encourage language study on the part of graduate students. As a consequence, the increase in the importance of such languages as Russian during recent years has not been paralleled by more widespread ability to read them on the part of persons active in research.

Another approach to the translation problem is to conduct large scale continuing translation projects. Some of the more important Russian scientific periodicals are, in fact, being translated and republished in English. The costs involved effectively discourage this solution of the translation problem except for material which is likely to find widespread interest and use. Material of more specialized character or of less obvious importance can scarcely be exploited economically by any such translation methods, A dearth of persons having the capabilities required for translating scientific and technical material also imposes limits on what can be accomplished by applying traditional translation methods to make effective use of knowledge recorded in foreign languages.

Translation from one language to another has become, therefor, *an* increasingly difficult problem in utilizing a large fraction of our available recorded knowledge.

In attempting to provide means for meeting this situation, considerable experimental work and thinking has been directed to the possibility of using automatic electronic equipment as aids to translation. The purpose of this invitational meeting was to bring together various persons who have been interested in developing and applying various aids to translation.

This field of investigation—sometimes referred to as machine translation—is still for the most part in the exploratory phase. At the present time, there are no installations in which automatic equipment is being used, on an operational basis, to achieve translation from one language to another. Nevertheless discussions at this invitational meeting revealed that extensive agreement as to eventual benefits and the practicality of attaining them exists among those who have been devoting thought and attention to the development of machine translation. As might be expected, the principle area of disagreement was in the realm of research strategy and tactics. In particular, disagreement was observed as to what can be reasonably expected to be accomplished in the foreseeable future in providing effective machine translating techniques for operational use.

Limitations on Translation

A consideration of basic importance was referred to several times, particularly in connection with discussion of the question of defining practical goals for the development of automatic aids to translation. This basic consideration which concerns the very nature not only of translation, but also of communication by language, might be stated as follows. Language is a means for communicating with the aid of symbols which represent concepts and ideas. For successful communication by means of language, it is therefore necessary that certain symbols be associated with certain concepts or ideas in the mind of the person receiving the message. Communication by language may fail in either or both of two fundamentally different ways. The person to whom a message is being directed may not *be* acquainted with the concepts or ideas to which the symbols are supposed to direct attention. Such a situation may arise, for example, when the person who is receiving 3 message is of a different cultural background than the person who is sending it. Differences in cultural background may arise as the result of differences in nationality or in professional specialization.

Thus a message from a mathematician may mean relatively little to a chemist because of the latter's unfamiliarity with certain mathematical concepts. (The converse is also true, of course. A paper on organic chemistry may be virtually meaningless to a mathematician.) Such differences in cultural background may also be the result of differences in national customs. For example, a Frenchman who has lived his entire life without seeing or hearing of baseball would scarcely be expected to be acquainted with many of the ideas and concepts peculiar to that game such as "homerun", "double-play," "stolen base," or "shortstop." No conceivable single French word or phrase could evoke in the mind of our Frenchman the same associations that the above mentioned English terms evoke in the minds of most Americans. In the literal sense, at least, such baseball terms are quite untranslatable for a Frenchman who knows nothing of the game. It might, of course, be possible to convey to him an impression closely related to say "homerun" by using an appropriate French word or phrase to indicate a dramatic, successful stroke. But in so doing, the allusion to the American game of baseball with all its associations and connotations would, of course, be lost. At best, only a general translation of one aspect of the concept "homerun" would be achieved. A similar situation exists when the attempt is made to translate poetry that is based on allusions to local customs or traditions or to episodes in national history. Translation of the epics of one national culture into the language of another nation may present very difficult problems. The same is true of many jokes and other forms of humor.

If similarity in cultural background has acquainted both the sender and the received of a message with a common set of basic ideas and concepts than communication by means of symbols to represent such ideas and concepts becomes possible. If, however, the sender and the receiver of a message customarily associate different symbolism with the same concepts and ideas, it may be said that they speak different languages. In such a case successful communication may be achieved by conversion of messages expressed in one set of symbols to another set of symbols, that is to say by translation. In science and technology, such translation is usually successful in achieving communication, as the ideas and concepts with which scientific and technical people work are the result of more or less deliberate cooperation and interchange of ideas between the scientific and the technical people of all nations. Translation is also usually achievable in the field of military science and tactics which are strongly influenced by the capabilities

of various weapons, and by the application of various scientific and technical principles to make their design. In such realms of activity as art, religion, politics, law, and business, differences in customs and underlying philosophy may result in the existence of impediments to understanding of a similar nature to those encountered in the case of our Frenchman who had not become acquainted with baseball terminology.

Realistic Goals for Machine Translation

As in other fields of research, investigations in machine translation may be undertaken either from the point of view of pure research or practical development. With certain languages at least, the situation is such that practical advantages appear immediately achievable without the necessity for extensive preliminary research. With other languages, the situation appears to be less promising, as far as the immediate possibilities are concerned. As discussed subsequently, the nature and structure of various languages present a variety of problems in developing automatic devices to facilitate or to accomplish translation.

As noted above, the translation of scientific and technical material appears to be a particularly promising field in which to apply automatic devices. In considering problems in this field, discussion at the invitational meeting pointed out that providing a complete and detailed translation for a foreign language paper on a scientific and technical subject requires two types of skill. One of these is linguistic in nature—namely the ability to make use of the two languages from which the paper is to be translated, and into which it is to be rendered. The other skill is the ability to understand the subject content of the material, and this may require an expert's knowledge of the field to which the subject content of the paper pertains. Consequently, in working out a complete and detailed translation of a highly technical paper it may be necessary for two persons to work together, one of them having the necessary knowledge of languages, and the other contributing an understanding of subject content, especially a knowledge of the specialized terminology and technical expressions of the output language. From one point of view, the person having linguistic knowledge may be regarded as being assisted by the subject expert in rendering the meaning of the paper into another. From another point of view, the linguist might, perhaps, be regarded as a sort of living dictionary and interpreting device for the subject specialist who proceeds to rewrite the paper in the output language. Translation may, in fact, be accomplished without the linguist at any time achieving full un-

derstanding of the paper because of lack of understanding of the technical concepts involved. Machine translation in such cases might be regarded as a rather special case of such cooperation, with the linguist being replaced, to a certain degree at least, by an automatic device of one type or another.

If machine translation be approached from this point of view, the key question then becomes, "How many clues to meaning does a subject expert need in order to be able to understand the contents of a paper in a foreign language in sufficient detail to meet his purposes?' The extent to which a subject expert may need to understand the contents of a paper in the foreign language may vary considerably. It may suffice to know relatively little concerning the paper to arrive at the decision, in the extreme case, that a detailed study of the paper would not be rewarding. In other words, a very general knowledge of subject contents may suffice to establish the possibility of eliminating a paper from further consideration. A somewhat better understanding of the subject content may be required to enable the paper to be indexed, classified, or encoded so that it might be incorporated in a file for subsequent future retrieval in case its subject content may be of pertinent interest to a problem or situation that might arise eventually. Approximately the same degree of understanding of the subject content of a paper can be expected to enable a subject expert to pinpoint those paragraphs or sections of a paper whose complete understanding is of such importance as to warrant determined effort in collaboration with a person having expert linguistic knowledge.

The central problem in developing machine translation in immediately practical form is devising methods and equipment which will provide an optimum margin of benefits over costs in present-day situations. As already noted, one of the principal factors in this connection is the degree of understanding of the original which may be required when dealing with large masses of documents. Important benefits may be achieved if useful purposes can be served by providing a degree of insight that may suffice for screening and processing, e.g. by indexing or coding, so as to enable the document to be identified subsequently as to whether it is of pertinent interest or not. Other factors of importance in determining what form of machine translation may be of optimum practical advantage are of an economic or technical nature.

Economic considerations arise in connection with such questions as the amount of material in a given language which may need to be processed, and the availability of persons who might be able to translate the language in question. In the extreme case of a small

amount of material, for which human translators are readily available, it would scarcely be economically rewarding to provide machine translating facilities. On the other hand, if a relatively large amount of material in a given language must be processed for use, if, for example, large amounts of such material must be indexed and coded for subsequent retrieval, then the ready availability of human translators may or may not be decisive in deciding whether to provide some form or another of mechanical translation. The speed with which automatic electronic equipment can accomplish routine tasks may well make it economically feasible to provide for machine translation even of those languages for which it is not too difficult to find qualified human translators. It is noted in this connection, that the introduction of machine translating facilities, by more effectively directing attention to foreign language material, may well result in an increased interest in foreign language papers, and thus stimulate the more extensive use of large volumes of foreign material that otherwise might escape attention. An important consideration in this connection is the question as yet largely uninvestigated, as to the extent to which automatic translation, even to the limited degree of word by word translation, may be coordinated with indexing and encoding, for subsequent searching, retrieving, and correlating of information. If such coordination of translation and processing for subsequent recall should prove practical, then automation of the processing of foreign language material may be the key to effective utilization of large volumes of information with reduced costs in terms of effort and money.

Varying structural features of different languages also constitute an important factor that can be expected to be of decisive influence in determining the development of optimum forms of machine translation. Many of the more difficult problems, involved in developing machine translation, would not exist if each word or similar readily recognizable set of symbols could be accurately and conveniently represented by some one English word. Various important languages, of which German, Russian, and Chinese may serve as examples, deviate in various ways from such an ideally simple case. It is true that there are a number of important scientific and technical terms whose designation, in these three languages, by words or symbols corresponds to designation by individual English words. This situation is encountered more frequently in Russian than in either German or Chinese. In German, combinations of simple words or roots may form long composite words, whose translation into English may require that the German compound be split into its components, each of which may require

rendition by separate English words. In Chinese, two or more characters, each of which singly may be used to designate an idea, may be used in combination to designate a single concept such as may be represented in English by a single word. Thus, two Chinese symbols are used for each of the English words "oxidize" and "brainwash." Each of the Chinese symbols in such combinations has one or more meanings of its own when taken individually. As already noted, one of the principal problems in German is breaking down compound words into their components. In Chinese a principal problem is discerning what combinations of ideographs must be considered together in order to represent a single idea or concept. It would undoubtedly be excessively pessimistic to conclude that machine translating techniques cannot be worked out for handling such languages as German and Chinese, but it is perhaps evident that breaking down compound words into components or detecting meaningful combinations of individual ideographs does present problems of a somewhat different nature than the mere recognition of individual words whose beginning and end is clearly indicated by blank spaces in a printed line.

In addition to such problems in associating individual words or symbols with ideas and concepts, the various languages differ widely with regard to the syntactical problems that they present to persons interested in developing machine translation. In German, for example, the use of separable prefixes with verbs and the inverted word order in dependent clauses present problems of considerable difficulty. In developing machine translation from Russian into English even more difficult problems are encountered when the attempt is made to work out programs for automatic equipment so that the English definite and indefinite articles are supplied in the machine output. The English words "the," "a," and "an" can be inserted, for demonstration purpose, in the output of a machine by special programming designed for selected Russian sentences. This is, of course an entirely different matter than programming the electronic equipment so that it would be able to supply the English definite and indefinite articles correctly for any input Russian sentence. The problems encountered in this connection are of such difficulty that despair was expressed as to the possibility of their solution in developing machine translation in the foreseeable future. The problem of supplying the definite and indefinite article when translating by machine from Russian to English is a particularly difficult one because there are no symbols or word order conventions in Russian that correspond directly or indirectly to the English words "the," "a,"or "an." Even in languages, such as German, in which words corresponding

to "the," "a," and "an" occur, other difficulties may arise because of ambiguities in declension. Thus the German "der" may be used as the definite article with the nominative singular of a masculine noun, or with the genitive or dative of a singular feminine noun or with the genitive plural of a masculine, feminine, or neuter noun. Similar ambiguities as to significance of inflexional endings are encountered in Russian nouns, and noun modifiers, namely adjectives and participles. The situation is made even more difficult by the fact that the significance to be attached to grammatical clues, as indicated by inflexional endings or by word order conventions, is not independent of the words to which such endings are applied or the phrasing in which such words are involved. For example, the translation of the reflexive form of a Russian verb into English occasionally requires the reflexive forms of English verbs to be used. Much more frequently, however, the passive voice of the English verb will be required, or, not infrequently, the Russian reflexive must be translated by the active form of the corresponding English verb. The choice of the proper English verb form is determined therefore not merely by the form of the Russian verb, but by its meaning. Difficulties such as these led to a sharp difference of opinion as to practical goals for machine translating. One opinion was that, during the foreseeable future and certainly not within the next twenty years, it would not be possible to provide by means of automatic equipment complete and full translation of one language into another. A contrary opinion was that an intensive program of fundamental linguistic research would enable the goal of a full and flawless translation from one language to another to be achieved within the foreseeable future.

This sharp difference of opinion served to emphasize the point that it is necessary to distinguish between research in linguistic theory which, as is the case with pure research in any field, may or may not provide practical results within any specified length of time and research programs that are directed to meeting the requirements of a given situation and to provide an optimum margin of benefit over cost. If this latter practical point of view as to the development of machine translation be accepted, then it appears evident that a number of important considerations should be taken into account when planning and conducting this development. One of these is determining what the translation requirements may actually be in a given situation. As noted above, a rather crude word-by-word interpretation may suffice for the purposes of screening documents and also for indexing, coding, and searching documents collected in files. In defining goals, it appears particularly dangerous to make the tacit assumption that only so-called

"complete" translation can serve the requirements of a given situation. A second important practical point in developing operational machine translation procedures is to gear them with indexing and coding methods in order that searching and correlating of material being processed may be facilitated. A third important consideration is the development of more effective automatic equipment than is now available. General purpose computers are far from being optimum in design for use as translating machines.

The meeting closed with a discussion of the difficulty in finding financial support for the development of machine translation. It was felt that the novelty of the idea of translating with the aid of automatic equipment is an adverse factor in this connection. Furthermore, there seems to have been a failure to understand the advantages of what can be provided at the present time, for certain languages at least, by applying present day electronic techniques, without embarking on elaborate programs of linguistic research.

It was decided, at the conclusion of the meeting, to arrange similar meetings of the group in the future at various centers interested in developing machine translation.