

The State of the Art of Automatic Language

Translation: An Appraisal

By ANTHONY G. OETTINGER

The Computation Laboratory of Harvard University

Fourteen years have elapsed since the appearance of Warren Weaver's memorandum "Translation" (Weaver, 1949), which is generally regarded as responsible for contemporary interest in automatic language translation. Shortly thereafter, a handful of investigators, including Bar-Hillel and Yngve at M.I.T., Booth at the University of London and the writer at Harvard, had become attracted by the problem. A recent publication (National Science Foundation, 1962) lists nearly 50 research groups currently studying automatic language translation and related areas. The size of recent Soviet bibliographies of Western work (Ravich, 1963) and American bibliographies of Soviet work (Walkowicz, 1962) attests to this field's own vigorous contribution to information retrieval problems.

What has been accomplished in the meanwhile? By sober scientific standards, there has been considerable but unspectacular progress. On the other hand, the false expectations set up by the blatant press agency (cf. Taube, 1961; King and Chang, 1963) which has consistently marred the field since its inception remain largely unfulfilled.

What does the future offer? For the serious investigator willing to assimilate the best theoretical and practical discipline that contemporary linguistics, discrete mathematics, and computer sciences have to offer, there is no lack of challenging and significant problems whose solution is not so far beyond our grasp as to lead to hopelessness. Results of moderate usefulness are at hand for those who wish to apply them with due regard to their limitations; a gradual increase in usefulness may well be prophesied. There should remain little scope for spectacular claims by the crackpot or the charlatan: this aspect of the field has already been so thoroughly exploited that any further effort is bound to be repetitious.

The demand for automatic language translation stems from the belief that nations are unable to cope with an ever rising need for the transfer of technical information generated in one language into some other language. This need is an important aspect of what is frequently but with some exaggeration called the "information explosion". In this fertile soil, the notion of what Bar-Hillel has called "fully automatic high quality mechanical translation (FAHQMT)", planted by over-zealous propagandists for automatic translation on both sides of the Iron Curtain and nurtured by the wishful thinking of potential users, blossomed like a vigorous weed. Bar-Hillel's acid critique of the notion of FAHQMT (Bar-Hillel, 1960, 1963) is too well known to bear repeating here, especially since I agree with its main outlines.

What, then, are realistic goals? On the practical side, one can reasonably ask to what extent machines can assist or replace people in various phases of the translation process or, more broadly, in the whole process leading from the initial drafting of a document to the eventual assimilation of its contents by an interested reader. This is, essentially, a typical engineering problem. There can be no abstract and perfect solution valid for all places and times. Economic, technical, social and military factors control the extent to which the results of scientific investigations can be usefully applied. A natural evolutionary approach has been hampered somewhat by the perennial high hopes of FAHQMT just around the corner, but is now not unreasonable to predict that the scope of useful man/machine collaboration will, on the whole, tend to increase in a gradual and unspectacular way.

It is generally accepted that a command of the subject matter is more important to a technical translator than a command of the languages in which he is working, and technically competent polyglots are naturally scarcer than either monolingual technical specialists or polyglot laymen.

If the objective is to assist the monolingual technical specialist, who presumably is the man who needs to scan and can understand the content of foreign technical literature, then devices far less chimeric than fully automatic high quality translators might be helpful.

One of these is an automatic dictionary proposed by the writer over ten years ago. In such a system, source language words appearing at the input would be matched on a word-by-word basis by a sequence of their most frequent correspondents in the target language. As pointed out earlier (Oettinger, 1955, 1960):

“A reader proficient in the subject matter of a text as translated by an automatic dictionary can readily assign this text to one of three categories according to whether (a) he can extract sufficient information from the text to meet his needs, (b) the text is irrelevant and may be ignored, or (c) the text is not fully understandable, but is of sufficient potential interest and importance to merit translation in the ordinary manner. Hence, the use of an automatic dictionary would ensure that only texts which merit translation, and for which the output of the automatic dictionary is not fully satisfactory, will be brought to the attention of professional translators.”

This paragraph is still valid today, except for the regrettable use of the word “translated” in the first sentence. Experiments showed that such material could indeed be used quite effectively by monolingual readers of the target language with adequate technical competence. With some attention to output format to alleviate the visual confusion caused by careless display of many target correspondents for a single source word, and with the introduction of a few helpful morphological and syntactic marks by relatively crude additional devices, including, for example, look-up of each of the more common idioms as one unit, an automatic dictionary can readily be realized.

This is, in fact, the nature of a widely publicized “translation system” developed in the United States primarily to apply a specially designed photographic memory (King and Chang, 1963). The pressures toward immediate high-quality translation, however, lead the designers of this memory device to the self-defeating practice of eliminating the profusion of target correspondents by the simple expedient of allowing in the dictionary only one target correspondent for each source word, except for relatively few instances in which two correspondents were allowed.

The resultant product looks, of course, much neater and cleaner than if all reasonable target correspondents were given. Unfortunately, as tests by one major potential user have recently shown, a reader is quite helpless in the frequent instances where an inappropriate correspondent happened to be in the dictionary: the alternative correspondents that might have enabled him to make an appropriate choice on the basis of his technical knowledge are not available since they were deliberately suppressed. As a result, the original text has to be found, the advice of an ordinary bilingual translator has to be sought and the whole process turns out to be less satisfactory and more wasteful of time, effort, and money than if the machine had been kept out of the picture altogether.

If an automatic dictionary is to be used, therefore, it should be treated plainly as such, and not disguised as a more powerful translator. Similar comments apply to some of the more spectacular Soviet work and to similar practices of ignoring syntactic ambiguities, a point to which we shall return.

There is another current limitation which is more serious, since it affects the economics not only of an automatic dictionary but also of any potentially more powerful device; that is the high cost of transforming an input text into a form suitable for automatic processing. In the very few instances where any kind of record has been kept of such matters, the cost of keypunching or otherwise transcribing a text is extremely high; in fact, considerably higher per word than the cost of machine operations. As a consequence, the cost per word of even the crudest machine output—neglecting all the research and development cost leading to it—is still considerably higher in most cases than that of conventional translation. In the foreseeable future, automatic print-reading devices will handle only materials with great uniformity of layout and of type design, such as, for example, ordinary typewritten material. The value of crude devices like automatic dictionaries in situations where the economic factor is of prime importance is therefore still rather doubtful at this stage, when crude devices are all that we have at hand. It appears, therefore, that for most practical purposes, the *short term* solution to any problems caused by language barriers is to train more bilingual technical specialists.

The extent to which this economic balance can be shifted by careful engineering of the whole translation operation remains an open question, since any prototype is likely to be uneconomical. The fixation on FAHQMT has precluded serious study of what really should be called *machine-aided* translation, a process in which men and machines would cooperate in whatever mixture and manner the unembellished state of computer technology and mathematical linguistics permits.

Most of the research effort of the past decade has been directed to questions of syntactic structure. Answers to these questions were expected to help in removing the all-too-obvious deficiencies of the output of automatic dictionaries. What is not lexical or syntactic is labelled semantic and commonly swept under the rug, where it still remains in spite of the occasional diligent efforts of various scientific housewives.

Although to date the practical impact of this work has been slight, extensive and very significant gains have been made in our understanding

of the structure of language. Some attempts were made to apply newly gained knowledge to the improvement of practical mechanical translation. For example, word order rearrangement, insertion of prepositions, appropriate marks of number or tense, etc. have been introduced on an experimental basis in machine-produced translations from Russian to English, but the major problem of selecting an appropriate target correspondent for a source word on the basis of context remains unsolved, as does the related one of establishing a unique syntactic structure for a sentence that human readers find unambiguous.

In early writings, there was hope that detailed grammars of both source and target languages would be unnecessary for automatic translation since a so-called "transfer grammar" describing the differences between the two languages should be enough. Unfortunately, no one has succeeded in describing the differences without reference to a description of each individual language and, as a consequence, the enormous labor of providing detailed and explicit descriptions of the syntax of several languages has only recently begun in earnest.

A theoretical backdrop to this work has been provided by a very fortunate confluence of results drawn from logic, linguistics, and the theory of automata. The contribution of linguistics has been largely that of Chomsky. Although its abstract and formal character makes much of this work inaccessible to the traditionally trained linguist, an understanding of its essentials, for example through a reading of the relatively non-technical *Syntactic Structures* (Chomsky, 1957) is essential to anyone who wishes to master the literature of automatic language translation, or for that matter, the lively side of the linguistics literature of the past four or five years.

The work of Emil Post and later logicians on so-called combinatorial systems (Davis, 1958) has provided the mathematical tools necessary for the understanding of various useful models of language proposed in the last decade. The hierarchy of abstract machines ranging from finite automata through pushdown storage transducers to unrestricted Turing machines, elaborated within the theory of automata, has recently found a close match in the hierarchy of finite-state grammars, context-free phrase structure-grammars, and higher grammars elaborated within theoretical linguistics (Chomsky 1962, Evey, 1963).

As one illustration of the light that theoretical investigations have shed on problems of application, systems such as dependency grammars, projective grammars, pushdown store grammars, categorial grammars

and context-free phrase-structure grammars—all of which have been advocated as vehicles for syntactic description by various research groups—are all now known to be abstractly equivalent to one another in the sense that, roughly speaking, any language that can be described by one of these grammars can equally well be described by any other in the group. This does not mean that the intuitive or pedagogical appeal of all these grammars is necessarily the same, nor does it guarantee that any of them can be fitted to any given natural language to anyone's theoretical or empirical satisfaction. In fact, the latter question is currently still a matter of active controversy, and we shall return to it.

One reason for the popularity of context-free phrase-structure grammars (PSG's) is their direct kinship to the older immediate constituent theories of language structure which are the formal or informal basis for the parsing methods most of us have been taught at school. In such a system a sentence is treated as if composed of immediate constituents such as *subject* and *predicate*, each of which in turn breaks down into further constituents such as *verb* and *object*. Characteristically the diagrams usually drawn from the structure of sentences described in this way are the familiar trees.

It is well known that certain rare types of linked structures (e.g., such strings as ABCD...ABCD...) are beyond the scope of context-free phrase-structure grammars although instances of such constructions may be found in sentences of natural languages. There is no more reason, however, to use this fact for rejecting PSG's as a potentially useful framework for the grammar of languages such as English or Russian, than there is for accepting the other extreme argument that a finite-state grammar should be adequate since the set of all sentences that any machine is ever to be called upon to process will in fact be finite, not only finite-state. The latter argument is clearly absurd since only a little experimentation quickly reveals that a grammar so conceived would be so complex, unintelligible and unstable as to be practically useless.

The case for or against phrase-structure grammars is less clear cut, since success to date with these grammars has been remarkable but not complete, so that it is not altogether clear whether it would be more efficient and elegant to replace them by theoretically more powerful grammars or to complement them by ad hoc techniques that take advantage of natural limits that apparently are placed on the types of sentences that actually occur as opposed to those that could potentially be uttered (Yngve, 1961).

Bar-Hillel's fear that "Such grammars seem definitely to be inadequate in practice, in the sense that the number and complexity of grammatical rules of this type, in order to achieve a tolerable, if not perfect degree of accuracy, will have to be so immense as to defeat the practical purpose of establishing these rules" does not seem warranted in practice. In our own experience at Harvard, a grammar essentially of the phrase-structure type, which started out with 3500 rules, was recently reduced to 2100 rules with some increase in power on the way (Kuno, 1963; Kuno and Oettinger, 1962). Projected refinements should enable us to reduce the size of the grammar to nearer 1000 rules for essentially the same power.

Moreover, the degree of theoretical understanding of phrase-structure grammars that has recently been obtained is still absent for any more powerful grammar. Theorems of Chomsky (1962) and of Evey (1963) have shown, for example, that the set of all languages that can be either accepted or generated by non-deterministic pushdown store transducers is precisely the set of all context-free phrase-structure languages. Earlier conjectures of Oettinger (1961) regarding the role of pushdown stores in syntactic analysis have thus been confirmed and, although other mechanisms have been suggested (Matthews, 1962; Sakai, 1962) or implemented (Robinson, 1962), there is now good reason for regarding the pushdown store transducer as a "natural" device and not merely as a convenient programming trick, although the great simplicity of the pushdown store transducer does lend itself to very attractive and efficient modes of programming.

For example, the Harvard multiple-path syntactic analyzer (Kuno and Oettinger, 1962; Kuno 1963), based on a predictive technique originally proposed by Rhodes (1959) has recently been characterized abstractly (Greibach, 1963) as a directed production analyzer (DPA) realized by a nondeterministic pushdown store transducer. Every DPA is the inverse of a context-free phrase-structure generator (PSG) in a standard form with productions $P \rightarrow c P_1 \dots P_k$. It is an inverse in the sense that the DPA will accept as well-formed precisely those strings generated by the PSG. More significantly, Greibach has shown that, for every PSG in the sense of Chomsky, there is a PSG in her standard form which generates precisely the same set of strings; hence every PSG has a DPA as an inverse, and the intuitively evolved multiple-path predictive analyzer therefore turns out to have even greater generality and esthetic appeal than was originally hoped for.

It has been argued by Bar-Hillel, for example, that “transformational grammars have a much better chance of being both adequate and practical (than phrase-structure grammars)”. The use of transformational grammars does have attractive properties. First, transformational grammars can be inherently more powerful than phrase-structure grammars although their precise position between phrase-structure grammars and unrestricted production systems or Turing machines has still not been precisely determined. Second, a phrase-structure grammar that adequately describes a language in the formal sense may not yield intuitively attractive structures. In such cases, the application of appropriate transformations may add perspicuity without altering the abstract power of the grammar. Finally, the transformation approach is essentially the adaptation to linguistics of a well-known mathematical technique highly successful in other realms, namely, that of describing some large set of ultimate elements in terms of a smaller set of canonical elements plus transformations that will, in one direction, combine and modify the canonical elements into the ultimate elements and, in the other direction, decompose ultimate elements into their canonical ancestors.

It is not improbable that the description of any sentence in terms of its canonical base and a set of transformations may be more economical, elegant and perspicuous than an alternative description, where feasible, in pure phrase-structure terms. Nevertheless, some investigators, notably Yngve, have challenged this hypothesis on the basis of experimental data. Most important, however, as far as the comparability of phrase-structure and transformational grammars is concerned, is the fact, made abundantly clear by Chomsky but generally missed or obscured by his followers, that a transformational grammar in the sense of Chomsky must be based on a phrase-structure grammar.

Transformations are matched to the sentences to which they may be applied by means of a phrase marker defined in phrase-structure form. This causes no difficulty in most of the investigations of transformational grammars that have been made to date. The point of departure for these investigations is the view that language should be described by providing a grammar that will generate all and only the sentences of the language in an effective way that also yields useful insights and has adequate explanatory power. Given a kernel or canonical sentence and its phrase marker, together with a set of transformations, it is then easy to determine which among the transformations are applicable to the kernel with the given phrase marker and hence to apply these to yield new sentences with new

transformed phrase markers, that in turn define the further transformations that may be applied.

The situation is quite different when one is concerned with automatic language translation or, more generally, with automatic language data processing. The central problem in this connection is to determine for any given sentence how it might have been generated by whatever grammar is postulated for the language. Of course, no phrase marker is given with a sentence occurring in a natural text; hence there is no immediate clue as to what inverse transformations might be applied for determining the structure of such a sentence by reducing it to canonical form. Proposals that have been made (Matthews, 1962) for analyzing sentences by essentially generating them are, by the admission of their proponents, thoroughly impractical. Other means have had to be found.

Another major problem, which has been largely ignored since it does not present itself in pernicious form when one is generating sentences, is that of ambiguity. Any given grammar may generate a particular sentence in more than one way. If a generator is turned on, and allowed to go forever, sentences will indeed be generated in all the ways consistent with the structure of the given grammar, at least in terms of a gedanken experiment. To guarantee that all acceptable structures are concretely displayed at reasonable cost for any specific given sentence presented for analysis is a much more difficult matter.

This question has, in fact, not even been considered seriously, except within the last two or three years. As a consequence, automatic translation systems, even when they have attempted to make use of syntactic analysis to improve the output of an automatic dictionary, have given at best a single structure to each sentence without any guarantee whatsoever that this single structure is correct and, if indeed correct, the most likely one in a given context. The choice of the most likely interpretation in any given context is one of the major remaining open questions along with that of choosing among multiple target correspondents of course words.

Recently, however, two operating systems producing all parsings of a sentence acceptable to a given grammar have been produced for English by Robinson (1963) at the Rand Corporation and by Kuno and Oettinger (1962) at Harvard; a similar system for Russian is being produced by Plath at Harvard.

The operation of such analyzers to date has revealed a far higher degree of legitimate *syntactic* ambiguity in English and in Russian than has been anticipated. This, and a related fuzziness of the boundary between the grammatical and the non-grammatical, raises serious questions about the possibility of effective fully automatic manipulation of English or Russian for any purposes of translation or information retrieval. However, by consistently and exhaustively revealing the extent and the nature of the ambiguity implicit in a grammar, these analyzers open the door to systematic investigations of the problems of reducing syntactic ambiguities, in the long run by better fitting grammars and, in the short run, by providing for appropriately interspersed human intervention. Eventually perhaps even a better understanding of the very mysterious "semantic" processes that enable most reasonable people to interpret most reasonable sentences uniquely most of the time may thereby be achieved.

Dealing with ambiguity is extremely hard for both formal and psychological reasons. Formally, there is a class of unpleasant theoretical results that tell us that the ambiguity problem is recursively unsolvable even for context-free languages of greatly restricted generality (Chomsky and Schützenberger, 1963; Greibach, 1963), i.e., no general algorithm can be found for determining whether or not a given phrase-structure generator will generate some sentence in more than one way. Corresponding results can be obtained for analyzers. The outlook for practically interesting decidable subsets is dim, and so experimental search for special solutions in special cases is our only recourse.

In a grammar that purports to describe a natural language, the question is not so much the existence of ambiguity but, worse yet, matching the ambiguity of the grammar to that observed in the language. Some success has been achieved in incorporating ambiguities in a grammar that actually are recognized in the language. For instance, for sentences such as "They are flying planes.", the Harvard analyzer produces several structures each reflecting one of the distinct interpretations.

Unfortunately, current grammars also have ambiguities which are not recognized in the language. Some of these can be readily eliminated and cause no trouble. Their elimination usually corresponds to an enlargement of the precincts of syntax at the expense of what would otherwise be regarded as semantics.

However, much of the problem of appropriately fitting ambiguity of grammar to that of the language remains today. It is indeed a major

problem even to classify an ambiguity. Is it there because the grammar is at fault? or are we unhappy with it merely because our mind is fixed on one plausible interpretation to the exclusion of others? One's natural tendency would be to answer yes to the first question.

Consider, however, the following sentence: "People who apply for marriage licenses wearing shorts or pedal pushers will be denied licenses." The example may appear to be frivolous, but similar ones may be found throughout the technical literature. Have you thought, for example, that "*People who apply. . . or pedal pushers . . .*" would be denied licenses? Dope pushers would be! Or perhaps it is "*People who apply for... or (who) pedal pushers...*"? People do pedal bicycles. Are they wearing shorts, or are they applying for shorts that happen to be wearing marriage licenses? Will they *be denied* licenses? or will they be *denied licenses*? The present Harvard syntactic analyzer relentlessly exhibits several more (Kuno, 1963).

In summary the outlook is grim for those who still cherish hopes for fully automatic high-quality mechanical translation. For those concerned with a better understanding of the structure of human language and, incidentally, the structure of artificial languages such as those used in computer programming, there can be considerable satisfaction in the extensive progress made in the past decade and a good deal of optimism for the future. Finally, for those concerned with practical problems and willing to consider careful engineering designs based on well planned collaboration between man and machine, much is now known that could be quite usefully applied to machine-aided language data processing and particularly to machine-aided language translation. Emphasis on the training of bilingual technical specialists should, however, be increased in any kind of realistic planning for the short range.

References

- BAR-HILLEL, Y. (1963) *Four Lectures on Algebraic Linguistics and Machine Translation*, Hebrew University, Jerusalem.
- BAR-HILLEL, Y. (1960) "The Present Status of Automatic Translation of Languages," in F. Alt (Ed.) *Advances in Computers*, Academic Press, New York, pp. 91—163.
- CHOMSKY, N. and M. P. SCHÜTZENBERGER (1963) "The Algebraic Theory of Context-Free Languages," in P. Braffort and D. Hirschberg (Eds.) *Computer Programming and Formal Systems*, North Holland, Amsterdam.

- CHOMSKY, N. (1962) , “Formal Properties of Grammars,” in R. R. Bush, E. H. Galanter and R. D. Luce (Eds.), *Handbook of Mathematical Psychology*, Vol. 2, Wiley, New York (Forthcoming).
- CHOMSKY, N. (1957), *Syntactic Structures*, Mouton, 's-Gravenhage.
- DAVIS, M. (1958), *Computability and Unsolvability*, McGraw-Hill, New York.
- EVEY, J. (1963) “The Theory and Applications of Pushdown Store Machines”, *Mathematical Linguistics and Automatic Translation*, Report No. NSF-10, The Computation Laboratory of Harvard University.
- GREIBACH, S. (1963a) “Inverses of Phrase Structure Generators,” *Mathematical Linguistics and Automatic Translation*, Report No. NSF-11, The Computation Laboratory of Harvard University.
- GREIBACH, S. (1963b) “Formal Parsing Systems,” (submitted for publication).
- KING, G. W. and H. W. CHANG (1963) “Machine Translation of Chinese,” *Scientific American*, June 1963, pp. 124—135.
- KUNO S. and A. G. OETTINGER (1962), “Multiple-Path Syntactic Analyzer,” *Preprints of the Proc. of the International Federation of Information Processing Congress-62*, Munich, Germany (Proceedings to appear).
- KUNO, S. (1963) *Mathematical Linguistics and Automatic Translation*, Reports No. NSF-8 and NSF-9, The Computation Laboratory of Harvard University.
- LOCKE, W. N. and A. D. BOOTH (Eds.) (1955) *Machine Translation of Languages*, Wiley, New York.
- MATTHEWS, G. H. (1962) “Analysis by Synthesis of Natural Languages,” *Proc. 1961 International Conference on Machine Translation of Languages and Applied Language Analysis*, Her Majesty's Stationery Office, London.
- NATIONAL SCIENCE FOUNDATION (1962), *Current Research and Development in Scientific Documentation*, No. 11.
- OETTINGER, A. G. (1961) “Automatic Syntactic Analysis and the Push-down Store,” *Proceedings of Symposia in Applied Mathematics Vol. XII*, American Mathematical Society, Providence, Rhode Island.
- OETTINGER, A. G. (1960) *Automatic Language Translation: Lexical and Technical Aspect*, Harvard University Press, Cambridge, Mass.
- OETTINGER, A. G. (1955) “The Design of an Automatic Russian-English Technical Dictionary,” in Locke & Booth (1955) pp. 47—66.
- RAVICH, R. D. (Mel'chuk and Tsvejj, Eds.) (1962) *Bibliografija Zarubezhnyx Rabot Po Mashinnomu Perevodu*, Institut Jazykoznanija Akad. Nauk

- SSSR, Moscow. (Also in English translation series, *Foreign Developments in Machine Translation* No. 122, JPRS: 19, 293; OTS: 63—21870).
- RHODES, I. (1959) "A New Approach to the Mechanical Translation of Russian," *National Bureau of Standards* No. 6295, also in *Mechanical Translation*, Vol. 6, November 1961.
- ROBINSON, J. (1962) *Preliminary Codes and Rules for the Automatic Parsing of English*, Memorandum RM-3339-PR, The RAND Corporation, Santa Monica, California.
- SAKAI, I. (1962), "Syntax in Universal Translation," *Proc. 1961 International Conference on Machine Translation of Languages and Applied Language Analysis*, Her Majesty's Stationery Office, London.
- TAUBE, M. (1961) *Computers and Common Sense*, Columbia University Press, New York.
- WALKOWICZ, J. L., (1962) *A Bibliography of Foreign Developments in Machine Translation and Information Processing*, National Bureau of Standards Report No. 7721.
- WEAVER, W. (1949) "Translation" in Locke & Booth (1955) pp. 15—25.
- YNGVE, V. (1961) "A Model and an Hypothesis for Language Structure" *Proc. American Philosophical Society* No. 104, pp. 444—466.