APPENDIX II.

LOGICAL PROGRAMMING RESEARCH IN THE UNIVERSITY OF WASHINGTON MACHINE TRANSLATION PROJECT[1]

BY

LEW R. MICKLESEN

AND

ROBERT E. WALL, JR.

This paper considers the problem of the automatic alteration of messages by their transformation from one linguistic system to another. This process is called "machine translation." The linguistic system in which the message was originally expressed is designated as the source language, and the linguistic system into which the message is to be transformed is designated as the target language.

Machine translation, as it is now conceived, is concerned with the translation of written material from one language to another. At the present time it is considered acceptable that the output be an unconventional and prosaic translation of the input material. The problems which must be solved in order to translate from one language into another may be classed under three broad headings: first, the specification of the source language; second, the specification of the target language; and third, the correlation between these two specifications. Contemporary research in machine translation does not necessarily follow consistently the lines implied by this classification; but, ultimately, these three areas constitute the essential elements of information. After the specifications have been made, they are expressed by program algorithms which can then be used with digital data-processing equipment to translate written source-language material into a printed form of the target language.

The ultimate goal in machine translation research is the formulation of an algorithm to effect translation from one linguistic system to another; nevertheless the fundamental problem confronting researchers at present is the discovery procedure for the elaboration of the specifications. The entirely nontrivial nature of this task should be immediately apparent. The multiplicity of referents of a given sentence, and the multiplicity of connections or syntactic relationships between words create a picture of initially bewildering complexity. Actually, within each individual linguistic system ambiguity is not infrequently encountered (witness the constant constructions of puns good and bad), but it is highly improbable. In the sentence: "He took his case to court" there is usually no ambiguity in the word "case" for any native speaker of English even though the form "case" in isolation has a multiplicity of meanings. Even when no ambiguity exists for the native speaker it is often extremely difficult to specify uniquely the patterns by which these ambiguities are resolved; indeed, one of the important tasks of linguistic science is concerned with just this problem. To resolve these ambiguities, it is necessary to describe the abstract system of a given language in such a way that any proper form or sequence of forms, and only the proper ones, can be generated by reference to the description. Looking at the problem from the standpoint of the reader or listener, the linguist wants to know just exactly how the reader or listener understands a message. The understanding of a given oral message has two aspects: sound and meaning. Much linguistic research has already been devoted to the "sound" or acoustic aspect of speech, considerably less attention of linguists has been drawn to the aspect of meaning in a message, and very little is known about the interdependence of sound and meaning.

The word "case" in the message "he took his case to court" is a particularly difficult example of the problem of meaning because there seems to be no general syntactic means by which native speakers of English recognize that "case" acquires its particular meaning in this instance. Rather, the co-occurrence of "case" and "court" seems to be decisive. Extensive syntactic research might reveal that "case" and "court" belong to particular classes of nouns, whose co-occurrence would be governed by a general rule of syntax for speakers of English. For the present, however, the "case" problem may be consigned to the area of non-grammatical meaning. This area of meaning will have to be left alone until its extent has been determined. This problem can be solved only after its present companion area, the area of grammatical meaning, has been defined. Any and all categories or classes of linguistic forms that express some relationship among the forms of a given message are representative of grammatical meaning. Many of these categories are already well known; for example, the past tense, objective case, and continuous aspect in English. Many of their subclasses and other classes remain to be discovered. Consider briefly the following English sentence: "What is meaning in this case?" How may the grammatical status of "meaning" in the above sentence be specified? Is it a verbal noun or a part of the analytical form of the continuous aspect "is meaning"? "Meaning" here is a verbal noun because the verb "mean" may transform into the verbal phrase "finite form of 'to be' + verb stem + ing" only when it precedes the construction "to + verb stem." This grammatical information is the kind of information to be exploited in determining the how's and why's of a given linguistic system.

---

[1]The work described in this paper was supported by the Rome Air Development Center, Contract AF 30(602)-1566, AF 30(602)-1827. Paper prepared for the International Conference for Standards in a Common Language for Machine Searching and Translation, Cleveland, Ohio, September 6-12, 1959.

But the enormity of the task carries with it a fundamental problem of approach. The ideal approach requires the exhaustive description of both target and source languages as separate entities through utilization of the most advanced techniques and models of language that modern linguistic science can provide. Implicit in this approach is the incontrovertible fact that all permissible structures in each language must be specified before a proper algorithm for effecting the equivalence of a target-language structure to a structure in the source language can be constructed. Since this approach assumes, at least in its theoretical outlines, the totality of language, certain procedural principles have to be imposed in order to cope with the vast amount of material. The use of texts or corpora, as employed traditionally in descriptive linguistics, must be augmented because huge quantities of text would have to be processed and even then would never assure the isolation of all structural possibilities. In addition, controlled modulation of all the known constructs in both source and target languages must be effected to reveal all permissible permutations. Linguistic science has to impose a hierarchy on the problem to reduce its complexity and lay bare the complex interrelationships characteristic of syntax.

There are also practical limitations which should be placed on linguistic research in machine translation but which may be difficult to apply, at least in the analytical stage of the research. This concerns the at present practical and perhaps obligatory limitation of machine translation to scientific literature. There is considerable doubt that machine translation can ever be effectively utilized for the translation of artistic literature, where there is a considerably greater range in the choice and use not only of words but also of constructs. The artistic and effective usages of a novel, for example, are undoubtedly completely foreign to a scientific treatise. If such constructs could be recognized and if there is absolute certainty that they would not be found in scientific literature, they may be disregarded in the analytical stage. In the ideal approach, the constructs of the source language would be matched with corresponding constructs in the target language. A bilingual dictionary for a given field or fields of science, in which every source-language form and its target-language alternatives are linked by tags with appropriate constructs, would be utilized. This translation system would operate, at least theoretically, as effectively as multiple nongrammatical meaning would allow.

Since machine translation is to be limited to scientific literature, and because of the manifold problems inherent in an exhaustive analysis, various empirical approaches have found wide favor in the machine translation world. The ingredients usually encountered in all of these approaches are the following: the use of a continually expanding corpus of scientific literature, the compilation of a bilingual lexicon on the basis of the corpus, the elaboration of a continually expanding and developing algorithm effecting translation of the corpus and based on the constructs discovered within the corpus.

In the preceding discussion the general problems involved in automatic language translation have been discussed along with certain procedures which may be used for the solution of these problems. In the next, attention will be focused on the research at the University of Washington. The major research effort of the University of Washington Machine Translation Project has been that of compiling a translation lexicon. This lexicon now consists of approximately 170,000 entries. The magnitude of the task involved in compiling this lexicon may be appreciated by observing that considerably over a ton of IBM cards is required to store the lexicon. This lexicon is apparently quite complete as far as general scientific language is concerned but needs considerable augmentation before complete translations can be made in any specific scientific field. The analysis which is required of Russian and English for the construction of program algorithms, on the other hand, has just begun.

The work at the University of Washington has not followed, in order, source-language specification, then target-language specification, and lastly a correlation specification. Rather, the effort has pursued tasks which embrace all three of these areas at once. The University of Washington MT operational lexicon, for instance, contains a minimum specification of both the grammatical and the non-grammatical meanings of words or idioms in the source language in terms of words and idioms of the target language on the basis of a word-for-word translation. The operational lexicon contains, therefore, a partial but adequate specification of the source language in terms of the target language. Emphasis must be placed on the fact that these specifications were made for source-language words in isolation and contain superfluous alternatives to be eliminated in a given context.

The program algorithms written so far have been concerned with resolving multiple meaning; superfluous alternatives are eliminated by consideration of the context. Essentially the algorithms are based on specifications of the source language. Such problems as word-order rearrangement which are completely dependent upon the specification of the target language, have been investigated only cursorily.

The algorithms developed were written for the IBM 650 computer. The 2,000-word storage capacity of the device is inadequate for the translation lexicon and is also too small to allow complete storage of all the processing programs. Instead of storing the translation lexicon in the computer memory and performing the dictionary search automatically, the lexicon is stored on IBM cards; and dictionary search is accomplished by hand. The dictionary search involved in translating a text passage proceeds by extracting from the card file the dictionary entry corresponding to each word in the text passage. It is necessary, of course, that copies of the cards be made for the individual dictionary entries since the same word will often appear several times in a particular text passage. After this manual dictionary search is completed, the entries are stacked in text order and the text passage is ready for processing.

The processing programs had to be divided into several parts in order to effect their application. To process text material, the first part of the processing program and then the text card decks are loaded into the computer. The computer executes the programs and punches out another card deck which is just like the input text deck except for the processing accomplished by the program. The changes appear in modifications of the tags, i.e., the coded grammatical and nongrammatical information which is stored with each individual entry. The text deck from this round of processing is then placed immediately behind the program deck for the next round of processing, and both decks are fed into the machine again. This is repeated until all processing has been completed. The output deck may then be introduced into the accounting machine to print out the translation.

These programs examine the context, and, on the basis of syntactic patterns which they are designed to detect, make modifications of the tags associated with the individual entries.

To accomplish this processing, a limited context of semantic units[2] is stored in the machine at one time: three semantic units before the semantic unit being processed; two semantic units after the semantic unit being processed; and the last substantive and the last verb. The total context which may be examined at any one time by the program is, therefore, eight semantic units. This figure of eight semantic units was arrived at as a compromise between computer storage capabilities and processing effectiveness. It must be emphasized that this limitation to an 8-word context was an enforced one: the minimum context storage for satisfactory translation is at least one complete sentence. Storage of eight semantic units allowed over half of the 2000-word capacity of the machine to be available for the processing programs, and at the same time the 8-word context was estimated to be sufficient to solve more than 85% of the occurrences of the syntactic patterns which were programmed.

Since it was not possible to store all programs on the memory drum at one time, the programs were divided into four parts. The programs of the first three parts are concerned with the actual syntactic processing while the fourth part performs special functions. The first part, called the First Round of Processing, is concerned with the solution of multiple meaning associated with the elements of a noun phrase linked by agreement and with the elements of a prepositional phrase linked by government. The second part, called the Second Round of Processing, is concerned with the solution of multiple meaning associated with substantives by establishing some substantive-verb and substantive-substantive patterns linked by agreement and some substantive-substantive, substantive-numeral and substantive-verb patterns linked by government. The third part, called the Third Round of Processing, is concerned with the solution of multiple meaning associated with verbs by establishing some substantive-verb patterns linked by agreement and some verb-infinitive and verb-adjective patterns linked by government and with the solution of a few multiple-form-class problems.

The actual processing performed by these three rounds will be amply exemplified below, but first the fourth part of the programs, called the Interpret Routine, must be discussed because it affects the results of the first three rounds. The Interpret Routine performs two functions. First of all, after an examination of the tags, it inserts in the translation English prepositions whose function is expressed in Russian by inflections. The reason for including the insertion of prepositions in the Interpret Routine is very simple. All rounds of processing narrow the number of case possibilities for some entries. The prepositions inserted depend on the remaining case possibilities; hence the insertion of the prepositions must be postponed until final processing.

The second function of the Interpret Routine is execution of the individual-entry subroutines. Individual-entry subroutines are processing programs which apply to one entry alone and are consequently stored as an integral part of the individual entries. For example, the Russian preposition "в" may govern either the locative or the accusative cases. The English equivalents of "в" are "in/to/at/on/of/like." If, in a particular instance, "в" governs the locative, the equivalents "to/of/like" may be deleted. Since this deletion applies only to the entry for "в," it would be wasteful of general programming storage to include the deletion routine for "в" in the general program. In the procedure described, the deletion program is stored with the entry for "в" in the large lexicon and is executed during the Interpret Routine. If, during the three rounds of processing, the case governed by "в" has been narrowed down to locative, then in the Interpret Routine the deletion program will eliminate "to/of/like" from the translation.[3]

Individual-entry subroutines are executed in the Interpret Routine because of the address system used in the IBM 650. This address system includes with each program step the address (the instruction address) of the next program step to be executed. If a routine is stored randomly, then each instruction address must be modified every time the program is stored since, in general, the routine will be stored in a different place each time. The individual-entry subroutines are stored as integral parts of the entries; and, since the entries are stored randomly in the Second and Third Rounds of Processing, the subroutines are also stored randomly. In the Interpret Routine only one entry is stored at a time, and it has a fixed location. As a consequence, the individual-entry subroutines can be executed conveniently in the Interpret Routine but could only be effected by complex initializing in either the Second or Third Rounds of Processing.

An example of a sentence illustrating the effect of each round of processing follows. This sentence has been taken from the original corpus of the University of Washington Machine Translation Project; specifically this is sentence 2 of Text Passage No. 1. To facilitate comparison by the reader, all forms of the sentence are printed together; a detailed discussion of each form in turn follows the presentation of the last form. The sentence appears in five forms: the original Russian, the word-for-word translation on the basis of the University of Washington MT Operational Lexicon,[4] the results of the First and Second Rounds of Processing plus the Interpret Routine, and the results of the First, Second and Third Rounds of Processing plus the Interpret

---

[2] A single free or bound meaningful symbol or symbol sequence, and any group of free symbol sequences which is idiomatic in terms of source-target semantics. See Reifler, Erwin, "Some New MT Terms," in Linguistic and Engineering Studies in the Automatic Translation of Scientific Russian into English, Technical Report prepared for Intelligence Laboratory, Rome Air Development Center, Griffiss Air Force Base, New York, 1958.

[3] Note that the processing of "в" is not an example of the insertion of English prepositions corresponding to Russian inflections.

[4] The only exception is the phrase физические свойства which has been treated as a "pseudo-idiom," that is, it will be coded in toto into the memory device. This will enable the automatic translation system to supply the idiomatic translation "physical-properties" (for reasons of consistency in the use of editorial output symbols the present translation system demands a hyphen linking the constituents of such idioms). For the concept of "pseudo-idiom" and its importance for MT lexicography and the improvement of the MT product, see Erwin Reifler's paper, "MT Linguistics and MT Lexicography" in these Proceedings.

Routine.

1. The Original Russian.

   Конструкция эталона, его физические свойства и способ осуществления определяются природой величины, единица которой воспроизводится, и состоянием измерительной техники в данной области измерений.

2. The Word-for-Word Translation.[5]

   construction/design (of)standard, (of)(to/for)(by/with/as)his/its//him/it physical-properties and/even/too method (of)realization(s) are defined/determined/assigned (by/with/as)nature (of)magnitude/quantity(s), unit/one (of)(to/for)(by/with/as)which is-reproduced, and/even/too (by/with/as)state/fortune (of)(to/for)(by/with/as)measuring/-dimensional (of)technics/practice//-technologists in/to/at/on/of/like (of)(to/for)(by/with/as)given (of)(to/for)area/oblast(s)[6] (of)measurements.

3. Translation after the First Round of Processing plus the Interpret Routine.

   construction/design (of)standard, (of)(to/for)(by/with/as)his/its//him/it physical-properties and/even/too method (of)realization(s) are defined/determined/assigned (by/with/as)nature (of)magnitude/quantity(s), unit/one (of)(to/for)(by/with/as)which is-reproduced, and/even/too (by/with/as)state/fortune (of)measuring/-dimensional technics/practice in/on/of given area/oblast (of)measurements.

4. Translation after the First and Second Rounds of Processing plus the Interpret Routine.

   construction/design of standard, (of)(to/for)(by/with/as)his/its//him/it physical-properties and/even/too method of realization are defined/determined/assigned by nature of magnitude/-quantity, unit/one (of)(to/for)(by/with/as)which is-reproduced, and/even/too by state/fortune of measuring/-dimensional technics/practice in/on/of given area/oblast of measurements.

5. Translation after the First, Second, and Third Rounds of Processing plus the Interpret Routine.

   construction/design of standard, his/its physical-properties and method of realization are defined/determined/assigned by nature of magnitude/quantity, unit/one (of)(to/for)(by/-with/as)which is-reproduced, and/even/too by state/fortune of measuring/-dimensional tech-nics/practice in/on/of given area/oblast of measurements.

The First Round of Processing establishes the agreement characteristic of substantives and their modifying adjectives and the government of the components of a noun phrase by a preposition. In the example cited, there is one instance each where agreement of a substantive with a modifying adjective and where government of a substantive by a preposition can be exploited to resolve multiple-meaning problems. The instance of agreement of a substantive with a modifying adjective is exemplified by the sequence измерительной техники. измерительной can be only feminine in gender and singular in number, but it may be either genitive,

[5] A few basic instructions are due the interested reader of this unprocessed output. Three spaces separate target equivalents for source-language semantic units. A slash separates alternatives, one of which must be chosen. One space separates parts of equivalents to be read together after a proper choice has been made from among alternatives separated by slashes. Parentheses surround alternatives which may or may not be chosen.

[6] The technical term "oblast," an administrative unit, has become a loan word.

dative, instrumental or locative in case. The substantive техники has the possibility of nominative plural and genitive singular feminine.

The ordinary syntactic binary combination of singular adjective plus singular substantive involving agreement demands that the two components share the grammatical categories case, number and gender. In the example under discussion, the adjective измерительной and the substantive техники share the genitive case, the singular number and the feminine gender.

Establishment of agreement between the adjective and substantive in this case prescribes the following deletions. The alternatives "(to/for)(by/with/as)," associated with cases other than the genitive, are removed from the equivalent for измерительной. The alternatives "(of)" and "//technologists" are removed from the equivalent for техники; the first because of the presence of a preceding adjective, the second because of its nongenitive grammatical information.

Government of the components of a noun phrase by a preposition is exemplified above by the prepositional phrase в данной области. Actually, the particular capability of the processing program for the solution of multiple meaning connected with the problem of government is limited to the preposition and the immediately following adjective or substantive. In this case, the preposition "в" may govern either the accusative or the locative case. The form данной, adjective and participle, is only feminine singular but may be either genitive, dative, instrumental or locative case. The syntactic binary combination of preposition plus adjective or substantive demands that one case required by the preposition coincide with one case inherent in the adjective or substantive. In the cited example, the case shared is locative. Coincidence of the case required by the preposition and the case exhibited by the adjective-participle allows the following deletions: the alternatives "to/of/like" associated with the accusative case are removed from the equivalent for the preposition "в." Bear in mind that the deletion of these alternatives entails an individual-entry subroutine; the processing connected with the following adjective, however, is a genuine part of the First Round of Processing. The alternatives "(of)(to/for)(by/with/as)," all grammatical information, are removed from the equivalent for данной because of the immediately preceding governing preposition.

Complete processing of the prepositional phrase в данной области in the First Round must proceed in two steps. The second step is the second instance of agreement between a substantive and its modifying adjective. The adjective данной has already been pinpointed as locative singular feminine. These same three grammatical categories can be matched in the tag for the substantive области and its alternatives "(of)(to/for)" and "(s)" are eliminated on the strength of this grammatical information.

The Second Round of Processing establishes some syntactic constructs involving government of substantives or adjectives by substantives and verbs. In the sentence at hand there are five examples of government of substantives by other substantives and two examples of the government of substantives by verbs. In all cases multiple meaning can be reduced. The condition of government by other substantives in the sentence being processed can be conveniently classified into two groups on the basis of complexity. The simpler condition is represented by the Russian forms, эталона, измерительной, and измерений. These three linguistic forms are all associated only with the grammatical information "(of)" signifying that they are in the genitive case and that this bit of grammatical information may or may not be retained in a given syntactic situation. All three forms are immediately preceded in the sentence by substantives which have the potentiality of governing a directly following substantive or adjective in the genitive case. Since the case governed by these preceding substantives and the case of the immediately following forms coincide, a syntactic linkage can be established; and retention of the grammatical information is prescribed. The parenthesis marks, consequently, are removed from the English preposition "of."

The more complex condition of government by other substantives is represented by the Russian substantives осуществления and величины. They both exhibit the elements of grammatical information "(of)" and "(s)" signifying in this instance that they are genitive singular and nominative and accusative plural. Again the case demanded by the preceding substantives is genitive, prescribing deletion of "(s)" and removal of parentheses from "(of)."

The Second Round of Processing solved two examples of government of a substantive by a verb. One was solved genuinely, the other by chance. The genuine solution was applied to the Russian substantive природой governed by the immediately preceding verb определяются. The substantive is instrumental singular feminine; the verb is a third person plural reflexive form with a passive meaning and therefore has the potentiality of governing a substantive in the instrumental case. The coincidence of the case governed by the verb and the case exhibited by the substantive establishes the binary construction in this sentence and indicates deletion of the alternatives "with/as" and the parentheses. Accidental solution of a verb-plus-substantive-in-the-instrumental construct was applied to the Russian substantive состоянием and the preceding verb воспроизводится . Here too the substantive is instrumental, and the verb is a reflexive form with only a passive meaning; but the substantive here is not governed by the verb. The substantive состоянием is actually governed by the verb определяются, but this verb played no role in the correct processing of состоянием because it is beyond the range of the program. The level of the present processing program does not allow it to consider the implications of the comma after воспроизводится. It is obvious that this kind of processing has serious limitations.

The Third Round of Processing solves multiple meaning associated with some verb forms and with a few multiple-form class words by establishing the different syntactic constructs characteristic of such words. The sentence under discussion contains only examples of multiple-form-class words. A word in the multiple-form class exhibits the syntactic behavior of two or more form classes. There are two such words in the sentence under analysis: его which is at once a prosubstantive and a proadjective, and и which is both a coordinating conjunction and a particle. The program developed for words like его searches the immediate context of его and establishes essentially the presence of a directly following substantive, the idiom физические свойства, and the absence of any preceding verb governing two objects. This contextual information is sufficient in the given case to pinpoint the form class of его as proadjective. Once this decision has been reached, the same program

will perform an operation similar to one of those in the First Round of Processing, i.e., the gender, number, and case information for его and the following substantive will be matched and appropriate grammatical alternatives will be deleted. In this case the substantive may be nominative or accusative plural; so all the grammatical information in the equivalent for его is removed.

The program developed for words like и reveals in the sequence физические свойства и способ an immediately preceding substantive in the nominative or accusative case and an immediately following substantive also in the nominative or accusative case. The presence of a preceding substantive and a following substantive in identical cases is sufficient evidence for the present program to pinpoint и as a coordinating conjunction and to delete the alternatives "even/too." There is a second occurrence of и linking the substantive природой and состоянием but the form природой is located beyond the contextual range of the processing program and no solution was attained.

The effectiveness of this type of processing can be easily and simply determined by reference to the University of Washington dictionary equivalents as a standard. The following formula[7] borrowed from the concepts of statistical communication theory permits calculation of the effectiveness (E) of the three rounds of processing in terms of the solution of multiple meaning:

$$E = \frac{\log_2 (p_1/p') \log_2 (p_2/p'_2) \ldots \log_2(p_n/p'_n)}{\log (p_1) \log (p_2) \ldots \log_2(p_n)} = \frac{\log_2 (p_i)/(p'_i)}{\log_2 (p_i)}$$

where: $p_i = 1/s_i, s_i$ number of i^{th} possible English alternatives in the word-for-word translation for the i^{th} semantic unit of the source language.

$p'_i = 1/s'_i, s'_i$ number of i^{th} possible English alternatives in the processed translation for the i^{th} semantic unit of the source language.

The application of this formula to the output of the University of Washington MT operational lexicon may be illustrated in the Russian phrase:

О лечении нервной импотенции новокаином.

The word-for-word translation of this phrase is:

about/against/with treatment (of)(to/for)(by/with/as)nerve/nervous (of)(to/for)impotence

(by/with/as)novocain

In order to apply the above formula, the number of possible translations must be determined. Equivalent number one has obviously three alternatives and, therefore, three possibilities. In the case of equivalent number three, the reader may choose any one of the six English prepositions or none of them--a total of seven choices--and must choose either "nerve" or "nervous;" consequently, there are 7 x 2, or a total of fourteen possible combinations. In like manner, equivalent number four has four possibilities, and equivalent number five also has four.

In the word-for-word translation of the whole phrase there are (3)(14)(4)(4) = 672 possible sequences. If the particular processed translation were

about/against/with treatment of nerve/nervous impotence (by/with/as)novocain

the number of possible sequences would be (3)(2)(4) = 24, and the effectiveness of the translation would be:

$$E = \frac{\log_2 \frac{672}{24}}{\log_2 672} = \frac{\log_2 28}{\log_2 672} = .51$$

This equation may now be used to calculate the effectiveness of the processing routines in the complete Russian sentence utilized above. The numbers of possible sequences for each translation read as follows:

Word-for-Word

(2)(2)(16)(1)(3)(1)(4)(3)(4)(8)(2)(7)(1)(3)(8)(14)(5)(6)(7)(16)(2) = 2.33×10^{12}

First Round of Processing

(2)(2)(16)(1)(3)(1)(4)(3)(4)(8)(2)(7)(1)(3)(8)(4)(2)(3)(1)(2)(2) = 2.38×10^{9}

[7]The reasoning behind the advisability of using a logarithmic function may be found in many good books on information theory. For a good discussion see C. Cherry: On Human Communications, John Wiley & Sons, 1957, p.178.

Second Round of Processing

$(2)(1)(16)(1)(3)(1)(2)(3)(1)(2)(2)(7)(1)(3)(2)(2)(2)(3)(1)(2)(1) = 2.32 \times 10^6$

Third Round of Processing

$(2)(1)(2)(1)(1)(1)(1)(3)(1)(2)(2)(7)(1)(3)(2)(2)(2)(3)(1)(2) = 4.84 \times 10^4$

After inserting the above values in the equation and performing the requisite calculations, the effectiveness of the First Round of Processing (E_1) is:

$$E_1 = \frac{\log_2 \frac{2.33 \times 10^{12}}{2.38 \times 10^9}}{\log_2 2.33 \times 10^{12}} = .242$$

The effectiveness of the First and Second Rounds of Processing (E_2) is:

$$E_2 = \frac{\log_2 \frac{2.33 \times 10^{12}}{2.32 \times 10^6}}{\log_2 2.33 \times 10^{12}} = .483$$

The effectiveness of the First, Second, and Third Rounds of Processing (E_3) is:

$$E_3 = \frac{\log_2 \frac{2.33 \times 10^{12}}{4.84 \times 10^4}}{\log_2 2.33 \times 10^{12}} = .616$$

All the elements of information utilized by the processing routines are traditionally grammatical in nature, and all the problems of multiple meaning solved by the processing routines are based on traditional grammar. There is one apparent exception in the equivalent "technics/practice//technologists" where the alternative "technologists" belongs ostensibly to the area of non-grammatical meaning but was eliminated on the basis of grammatical information.

The only remaining "grammatical" problems in the example involve "his/its," "(of)(to/for)(by/with/as)-which," and "and/even/too." If these problems were all solved, the processing effectiveness (E_4) would be:

$$E_4 = \frac{\log_2 \frac{2.33 \times 10^{12}}{1.15 \times 10^3}}{\log_2 2.33 \times 10^{12}} = .75$$

The ratio of effectiveness of the three rounds of processing to the effectiveness of the instance where all "grammatical" problems are solved is:

$$\frac{E_3}{E_4} = \frac{.616}{.754} = .818$$

In other words, 81.8% of the traditional grammatical problems was solved by the three rounds of processing. An examination of a considerable number of examples has indicated that the three rounds of processing consistently solve about 80-90% of such multiple-meaning problems.

From the preceding discussion it is obvious that the logical processing programs written at the University of Washington need considerable augmentation before satisfactory translations can be realized. The progress has been encouraging, however. In summary, the following points should be stressed:

(1) Three rounds of logical processing were developed to test the effectiveness of such operations on the University of Washington lexicographical work, which, in turn, was done in order to optimize a word-for-word translation. The machine translations obtained are unconventional, but accurate and intelligible.

(2) These rounds of logical processing eliminate superfluous grammatical information in the form of target-language alternatives by reference to individual semantic units co-occurring in a limited context. They make no pretense of lacing together whole constructs of the source language.

(3) The process is limited in any one operation to a context of 8 semantic units. This figure was chosen as a compromise between computer storage capabilities and processing effectiveness with the realization that the minimum context storage for satisfactory translation is one complete sentence.

(4) The three rounds of processing, to the extent that they were applied, solved approximately 80-90% of the multiple-meaning problems of a purely grammatical nature and about 50% of the totality of multiple-meaning problems.

Present research at the University of Washington in the area of logical processing continues to be based on the original lexicographical work but utilizes a considerably more sophisticated body of grammatical information than that described in this paper.