# MACHINE LANGUAGE TRANSLATION STUDY

First quarterly progress report

## 1 May 1959 - 31 July 1959

Research on the machine translation of German

Contract No. DA 36-039 SC 78911
File No. 18678-PM-59-91-91 (6909)

W. P. LEHMANN, Chief Investigator

## TABLE OF CONTENTS

Purpose:

The aim of the contract is to conduct a study and analysis of the problems of translating foreign languages into English by automatic means.

The work falls into two types of study and analysis: 1) linguistic; 2) programming. Since the foreign language under study is German, our linguistic analysis is directed at it. As analysis of German is more complete, programming techniques for its translation to English will be worked out.

Three teams of linguists have been established: one consisting of Dr. Winter working with four assistants; another consisting of Dr. Werbow working with two assistants; the third consisting of Dr. Parker with two assistants. Each of these teams is dealing with specific problems in the analysis of German.

The fourth team consists of two programmers working with Mr. Pendergraft.

Related projects, with bibliography giving reference to their work, are discussed in section 4 of the body of the report.

## Abstract:

Our technical report No. 1 summarizes previous work at the University of Texas in the context of work in machine translation generally. Progress has been made in the analysis of German, particularly at the morphological and syntactic levels.

The lengthy study by Dr. Winter's group on identification of subjects, summarized in the report, indicates that most ambiguities can be eliminated. We plan to eliminate them completely, but instead may have to give alternate renditions of some sentences which are ambiguous in German or we may signal in some fashion that an ambiguity is present and in this way suggest to the reader that he make the proper interpretation of the two possible.

The study on the classification of verbs by Dr. Werbow and Miss Sträussnigg, presented in the report, indicates the linguistic analysis necessary for programming. The classification will also serve to show the type of analysis being carried out on other elements of the grammar.

Work on programming has been largely exploratory, with surveys of work elsewhere, as indicated in the report, and study of the possibilities of analysis by machine.

## Publications, lectures, reports, conferences:

The linguists in our group took part in the conference on English syntax, held at the University of Texas June 16-19 under the direction of Professor A. A. Hill. The aim of the conference was to attempt to bring closer together two approaches to syntax, one based on signals grammar, the other on transformational grammar. Professor Martin Joos of the University of Wisconsin discussed the first; Dr. Noam Chomsky of the Massachusetts Institute of Technology the second. Invited participants from various institutions represented both approaches. The most important result of the conference was greater understanding of both positions and their relationships.

## 1. Background

### A. Early work in machine translation.

Work in machine translation, including ours at The University of Texas, has been based on the assumption that thorough linguistic descriptions of source and target languages are essential before programming can be undertaken. All machine translation projects have accordingly set out to secure such descriptions from previous linguistic analyses or to obtain them through their own analyses. Even when previous analyses of languages have been extensive, scholars in machine translation have found them inadequate and incomplete. The work which has been expended on machine translation has therefore been largely directed at providing complete descriptions of selected languages. An insight into the reasons for this work may be furnished by noting the make-up of previous linguistic descriptions, and their deficiencies for work in machine translation.

Until recently linguistic analysis was wholly traditional, with bases in meaning, logic and the grammar of Latin and Greek; grammatical categories were set up on these bases rather than on the basis of formal markers in the language to be analyzed. English, for example, was equipped with an objective case in noun inflection, a future tense and a passive voice in verb inflection; grammars of English prescribed patterns that seemed to conform to logic and rules of Latin, such as 'it is I' rather than the general 'it is me'. Since categories based on meaning could not be utilized for machine coding, it was fortunate for machine translation that linguists began to abandon this type of description from approximately 1925, even though it still is prominent in our text-books. The type of descriptions now produced by linguists establishes categories and classes in any given language solely on the basis of the formal differences which are evident in that language. Since linguists attempt to determine the relationship and the structure of categories and classes, the present type of linguistic analysis is often referred to as structural linguistics; another less widely used name is 'signals grammar', based on the establishment of grammatical classes and categories only when texts contain a formal signal for them.

Between 1925 and today structural linguistics has been occupied with the development of methodology and the description by that methodology of a small group of languages. Much work was done on English; less on German, Russian and other languages. But even the languages subjected to considerable analysis, like English, were described only partially: determination of the sound system, or phonology, and of the form system, or morphology, occupied virtually all the efforts of structural linguists dealing with English. The syntactic system was little studied; the meaning system was scarcely touched. The first scholars who dealt with machine translation realized the importance of syntactic analysis and devoted much of their attention to it. The Georgetown - IBM experiment of 1954, for example, concerned itself largely with syntactic rules. Earlier exploratory work on German by Oswald and Fletcher aimed for the 'mechanical resolution of German syntax patterns' (Modern Language Forum 36. (1951) 1-24.) In this way the requirements of machine translation directed the attention of scholars at all levels of linguistic analysis.

The goals of machine translation differ, however, in several respects from those of structural linguistics. One of these is the concern of machine translation for the comparative structure of languages. Structural linguistics analyzes every language in terms of its own structure; machine translation analyzes further the structural similarities and differences between two or more selected languages. Moreover, structural linguistics deals with the spoken language, while with the present limitations on equipment machine translation must restrict itself to the written language.

Early scholars dealing with the machine translation of German recognized that a pedagogical system had been developed for the translation of German to English, and they began their work with a consideration of this system. It was devised by Professor

C. V. Pollard of The University of Texas and is the essential part of his text-book: The Key to German Translation. Professor Pollard compared the structure of German and English sentences, and with the aid of noun capitalization in German taught students eleven rules for the translation of German into English. Though his approach provided a good jumping-off place, scholars soon found that their rules for machine translation needed to be much more precise and that they could not rely on the capitalized noun as the essential clue to an analysis of the German sentence. For initially in the German sentence, capitalization is essential for all parts of speech, and other elements within the sentence may be capitalized; nouns then are not unambiguously marked. Oswald and Fletcher set up more complex rules for German to English translation; their rules are in turn amplified in Mechanical Resolution of Linguistic Problems by A. D. Booth, L. Brandwood and J. P. Cleave (London, 1958) 125-286. Other scholars have contributed further detailed rules and descriptions. Before evaluating these, with a brief sketch of the contributions made by various groups working on machine translation, we may find it profitable to compare the current status of structural linguistics with the requirements of machine translation.

## B. Structural linguistics and the problems of machine translation.

One of the most important findings of structural linguistics is that a language is composed of a series of levels or hierarchies, and that it is not one simple system. The grammatical hierarchy is completely separate from the phonological; semology comprises a further hierarchy. Further linguistic analysis may set up within these hierarchies further hierarchies, such as a syntactic as opposed to a morphological hierarchy. Of importance now is the finding that language cannot be analyzed into one simple set of units.- Like other non-specialists, some workers in machine translation seem unaware of the hierarchial make-up of language. When proposing linguistic analysis for the programming of German Booth- Brandwood-Cleave, 47, suggest that "The dictionary entries [in any sentence] bear alongside them an indication of the part of speech each usually represents." Such a procedure would be inadequate, as we may illustrate with even the exeedingly simple sentence: "all acids contain hydrogen." Labeling the last three words with the normal part of speech tag, we would call the second and fourth nouns, the third a verb. This analysis would be accurate at the morphological level. If maintained at further levels, it would conceal the similarity between 'acids' and 'they' in this sentence and in the sentence "they contain hydrogen." To indicate the similarity we must set up a further level of analysis. While at the morphological level "acids" is a noun and "they" a pronoun, at the further level, which we may call phrasal, both are nominals. But even this level does not reveal the entire structure of the two sentences, and we must set up a further level, which we may call clausal.

At the clausal level 'all acids' and 'they' in the above pair of sentences are similar sentence components and we may label them 'subjects'.

If we labeled elements of these sentences with the customary part of speech designations, the variety of possible sentences would be unmanageable. With the hierarchial analysis, our two sentences are identical with each other at the clausal level and with further complex sentences like: "All the substances which are in the containers in that cabinet contain hydrogen." For a useful linguistic analysis in machine translation, the hierarchical structure of language will have to be noted.

It has been suggested above that structural linguistics has gradually dealt with increasingly complex elements of language, starting with the phonological hierarchy. The latest hierarchy to be dealt with structurally involves meaning. If we consider phonology the simplest hierarchy and grammar the next in complecity, analysis of meaning follows grammar. The linguistic approach to meaning has been described most thoroughly by Martin Joos in "Semology:

A Linguistic Theory of Meaning," *Studies in Linguistics* 13.53-70 (1958).

Semology does not deal with the relations between linguistic units and the outside world; it does not, for example, investigate the relationship between the word 'cat' and living beings such as the *felis libyca domestica* or a type of woman or a kind of fish or a powerful tractor. Rather, "semology undertakes to explain the semantic functioning of 'content' morphemes from the interrelations of abstract semological units and forms..." [Joos, 53]. It is important to note that like other current linguistic study semology is formal; further, that it deals with the relationship of entities in a text. Again structural linguistics is analyzing language in a manner which is highly profitable for machine translation. We may note the relevance of this further hierarchy in linguistic analysis by citing the six "contextual clues" which Professor Yngve proposes to use in producing an elegant translation; see "A Framework for Syntactic Translation," _Mechanical Translation_ 4.59-60 (1957).

If we rearrange the sequence of Professor Yngve's clues we find that his second, third and fourth correspond to the structural linguists' grammatical hierarchy, his first, fifth and sixth to the semological hierarchy. Yngve's second clue deals with idioms and compound nouns; his third with syntactic classes based on order; his fourth with selectional relations. His first deals with the field of discourse (if it were ichthyology, the use of 'cat' might already be circumscribed); his fifth with antecedents and his last with contextual clues generally. Since we may in this way observe a close correlation between the procedures of structural linguistics and the requirements of machine translation, the pertinence of structural linguistic methodology for our work may be clear. The correlation may also point up the course of the research in machine translation performed by the Texas group.

## 2. Course of work of the Texas group, in relation to that of other groups.

The first aim of Texas group in machine translation was to provide a complete description of German from a formal point of view, particularly of those segments of the language which were inadequately described. These are primarily syntactic; the morphological classes have been thoroughly described, but need rearranging in accordance with the requirements of machine translation. Similarly, the spelling devices of German, its graphemic hierarchy, is well described but needs the precise formulation required for machine translating. The tasks called for may then be arranged by various hierarchies and levels.

1. The graphemic hierarchy
2. The grammatical hierarchy
   - 2.A. The morphological level
     - a. Inflectional
     - b. Derivational

2. B. The syntactic level

a. Phrasal

b. Clausal

3. The semological hierarchy.

Members of the group have undertaken projects which will fill in gaps in our description of German. Some of the resulting reports will be reproduced below as an indication of our procedures and for their interest to other workers in machine translation.

The graphemic hierarchy.

The graphemic hierarchy is important in determining the limits of the sentence and other entities, such as clauses. One of our early studies analyzed the markers by which sentences were delimited, essentially one of the four marks of punctuation. ? ; : followed by a blank space. The study also corroborated the inadvisability of relying on capitalization as a marker for nouns, which was noted above.

The grammatical hierarchy.

The morphological level: inflection.

In determining inflectional classes, standard grammars of German, such as that by George O. Curme (New York, 1922), can be heavily relied on for the essential facts. Their methods of arrangement and presentation however are not designed for work in machine translation. Accordingly restatements must be made.

We have studied various such re - statements. The one we choose ultimately will depend on our programming procedures.

The fullest published morphological analysis of German for machine translation purposes is contained in Booth-Brandwood-Cleave. Nouns are classified by them into 8 main classes, pp.155-6, noun modifiers, pp.158-60; verbs are classified into various groups, depending on their stems, pp.189-96. One of our appended papers presents a verb classification. It was prepared by R. Sträussnigg in collaboration with Dr. Werbow.

The morphological level: derivation.

Our investigations in derivation have demonstrated to us the need of dealing with bases and derivational elements rather than with complete compounds. Listing of compounds would greatly increase the size of the glossary. But even if there were no objections to a large glossary, there would be a fundamental difficulty in that compounds may be made at will in German. Accordingly one would never have an exhaustive list of them.

In one of the most useful papers dealing with the machine translation of German Professor E. Reifler has proposed a system of analyzing compound nouns into their components, *Mechanical Translation* 2 (1955).

Brandwood-Booth-Cleave, pp. 233-240, also discuss types of compounds and their analysis.

As with the treatment of inflection, the essential steps for dealing with morphological elements have been suggested and may be followed; details will be determined by programming methods.

The syntactic level: phrasal.

Many of the serious difficulties of machine translation are at the syntactic level and much of the previous work has been done here. Among the greatest problems in translating from German to English is the adjectival modifier construction. One of our early studies was devoted to it.

Other phrasal difficulties are caused by adverbial constructions and their order. Adverbial expressions of time, for example, precede those of place in German, while the reverse order holds for English. One of our current studies deals with the various types of adverbial constructions in German and their correspondences in English.

The syntactic level: clausal.

Since the basic unit of machine translation is the sentence, our group like others has devoted considerable time to the structure of German clauses. Our earlier discussions led us to regard the finite verb as the key to machine translation rather than the capitalized noun. Hence our concern with the morphological classification of verb forms.

After the finite verb is identified, the subject in a clause must be determined. Accordingly identification of the subject has been dealt with at length by us; the second appended paper, prepared by Dr. Winter with the assistance of Mrs. Orme-Johnson and Mr. P. Pollard, presents the results of the analysis of a considerable body of texts for identification of the subject. Application of the findings will permit ready identification of the remainder of the sentence elements.

Objects will be identified partially by position, partially by negative criteria, partially by coding in the base forms of verbs.

Adverbial elements will generally be marked, as by prepositions, or they will be unambiguous, or they will be identified by negative criteria. Booth-Brandwood-Cleave have discussed many syntactic problems, but their findings may be atypical for technical materials, since they analyzed literary texts.

The semological hierarchy.

It was recognized early that giving the proper translation for words with multiple meanings would be one of the greatest problems for machine translation. Two procedures will assist in solving these problems; delimitation of subject matter of a particular text; delimitation of the context of a particular word.

For the first of these the device of setting up idioglossaries was proposed. V.A.Oswald, Jr. "The Rationale of the Indioglossary technique," <u>Georgetown University Monograph Series</u> 10.63-69 (1957), discusses the underlying theory and the procedures.

For the second, texts will have to be analyzed to determine the words which stand in the environment of each of the multiple meanings of a given word. Among the troublesome words in German will be prepositions. We have completed some analyses in an attempt to determine their proper translation. Booth-Brandwood-Cleave devote the last forty-three pages of their treatment of German, pp. 244-86, to the problem of multiple meaning.

Procedures for translation from German to English (linguistic).

After linguistic analysis of the source language, for us German, procedures will have to be devised to arrive at any English translation. The problems involved have been discussed with an outline for their solution by V.H. Yngve, "A Framework for Syntactic Translation," Mechanical Translation 4.59-65 (1957).

In the production of the English text a great deal of interest has been devoted to the recently developed transformational grammar (which was one of the chief topics discussed at the recent conference at The University of Texas). Though machine translation may be able to profit from some of the techniques of transformational grammar, it differs from the transformational approach in starting from a set of sentences in contrast with the formulae devised in transformational grammar. Transformational grammar will then need to have explicit rules for the treatment of the various elements in a language; machine translation will need such explicit rules only where two selected languages differ from one another.

3. Work in programming.

Machine translation programming may be expected to follow a common procedural sequence: system design and specification, program coding and assembly, system testing and revision, and system extension.

A. System design and specification.

Our preliminary planning has been especially influenced by two existent systems. The first is the Russian-to-English "SERNA System" described by Peter Toma, _Georgetown University Machine Translation Paper 1._ The second, A.F.R. Brown's "Manual for a Simulated Linguistic Computer," _Georgetown Occasional Papers on Machine Translation,_ No.1, explains a "direct coding" technique to be used by linguists in programming French-to-English translations. Incorporation of this feature in our German-to-English system is discussed below. Our tentative system design is as follows.

The entering German sentences will be processed through four separate programs, passing through the computer at each stage from an input magnetic tape to an output tape, The latter will be returned automatically as the input of the succeeding program. English sentences on the last output tape will be displayed by an off-line printer. Additional tapes will deliver programs and glossaries as they are needed by the system.

Pass one.

The first program will use graphemic clues to separate incoming German sentences into entities that may be looked up in the German glossary. These will be numbered sequentially before being sorted, in batches, into the order they would have on the glossary tape. The German glossary will supply all appropriate grammatical and semological codes for each located entity. Entities having multiple German usage will be given multiple codes. This ambiguous data, after being sorted into the original entity order, will be written on the output tape. Complete procedures for handling unlocated German entities have not been formulated.

Pass two.

Ambiguous grammatical and semological data from the first pass will be resolved by routines operating under the direction of an interpretative executive program. These form-recognition routines will be coded by linguists in a convenient macro-language, which an auxiliary program will then translate into the form interpreted by the computer. The routines will embody those results of analysis that can be used to decide between multiple codes, and to produce an unambiguous sequence of grammatical and semological data for pass three. When reliable recognition of form is not possible, the most probable codes will be chosen. This circumstance will cause a special mark, signalling ambiguity, to appear in the English translation. Alternate translations may also be given in specific cases.

### Pass three.

A second executive program will now act upon unambiguous semological and grammatical clues to perform code substitutions and transformations leading to the final sequence of English entity codes. No English alphabetic data will enter the computer during pass three. The program will again interpret the results of macro-language coding by linguists, whose insights will be made available to the computer as form-synthesis routines.

### Pass four.

The last program will search the English glossary to find alphabetic equivalents for the final codes written during pass three. As in the first program, the codes will be sorted to glossary order and later returned to original order as glossary data is assembled into English sentences. Graphemic criteria will generate printer controls to insure a readable output.

B. Program coding and assembly.

One appeal of direct linguistic coding is the reduction of faulty communication between linguistic and programming skills. A second consideration, in our case, is the necessity to program the system on two computers: the IBM 709 and the presently unavailable Army Signal Corps MOBIDIC field computer.

Programs which simulate MOBIDIC on the IBM 709 will soon open the possibility of programming immediately in MOBIDIC language. However, the large simulation time factor would greatly increase computer time required før system testing and revision.

In view of our present dependence upon partially empirical methods, immediate MOBIDIC simulation would probably be too costly.

If programs are originally written in IBM 709 language to conserve system testing time, our choice of macro-language coding will reduce the reprogramming necessary for eventual MOBIDIC simulation. The essential functions performed by our system will reside in the form-recognition and form-synthesis routines which record results of linguistic analysis in macro-language expressions. These expressions need not be rewritten for the MOBIDIC system. Reprogramming is confined to the two executive interpretive programs of passes two and three, and to the macro-language subroutines controlled by these programs. No glossary changes should be necessary. The first and last programs of the system are relatively simple in structure and may be rewritten completely in MOBIDIC language.

C. System testing and revision.

A third virtue of macro-language coding relates to a basic difficulty inherent in all of the systems known to us. It is evidently true that machine translation systems, to a degree surpassing other larger computer systems, can never be demonstratively finished. Our best strategy, therefore, should be to provide for easy revision. This consequence would seem even more pertinent to military machine translation systems for use in the field.

Unlike many other computer systems, machine translation systems must be geared to such a large number of possible inputs that conventional programming techniques cannot prepare for every eventuality. Machine translation system testing procedures illustrate this lack of finality. The system is first required to translate the source-language corpus upon which its design has been based. Successive adjustments and revisions then improve the quality of target-language output until previously unanalyzed source-language texts may be profitably tried in the system. From this moment, testing procedures merge steadily with actual use of the system as a translation tool. Revisions should occur with decreasing frequency, but will remain as a recurrent chore dependent upon the difficulty of source-language inputs and upon quality requirements for the target-language output.

D. System Extension.

The open-endedness of machine translation suggests the need for techniques which will let the computer assist in adjustments and revisions of the system. Macro-language coding is one route to this goal. Richard Robinson will undertake the development of our German-to-English macro-language using COMIT and other operative languages as his point of departure.

A second route to more automatic system revision is being explored by Ramon Faulk, whose work has entered the difficult area of "learning" or "self-organizing" systems. Here we are trying to express our desires to the computer in terms of example translations, in the hope that faulty translations may be corrected and given back to the machine as system revisions. In spite of the clearly formidable problems that are to be dealt with when traditional translation techniques are discarded, some evidence has accumulated to argue for the feasibility of a "self-organizing" translation system based on a process to be called "redundancy sorting." A more complete account of progress in this area will accompany our next report.

## 4. The work of other groups in machine translation.

Earlier in this report the work of other groups has been referred to when pertinent. Here a brief characterization of the pertinent work elsewhere will be given by groups, as a type of cross-indexing and a summary. Several comprehensive books on machine translation have now been published, and information can be extracted from these on work done before this year. Since in a developing science the last book is generally the most useful, the most important survey now is to be found in Émile Delavenay's *La machine à traduire* (Paris, 1959). With Booth, Brandwood and Cleave's *Mechanical Resolution of Linguistic Problems* and Professor Bar Hillel's recent *Report on the State of Machine Translation in the United States and Great Britain* (Jerusalem, 1959), and the surveys of current work in the journal Mechanical Translation (published at the Massachusetts Institute of Technology under the editorship of Victor Yngve), Delavenay's book provides an admirable introduction to the problems and achievements of machine translation. Brief notes on some of these as treated by various groups follow.

The Georgetown group has been the most successful in this country and abroad in achieving translation by machine. Professor Bar Hillel recommends that any new group in machine translation become acquainted with the activities at Georgetown. We have done so through contacts in Washington and here. The chief investigator of our group was invited to give a paper at the Georgetown Round Table meeting dealing with machine translation iņ 1957 and again this spring; the 1957 paper was published in the Georgetown Monograph Series No. 10; the paper this year is to be published. Moreover, the director of the Georgetown University project has discussed at length with our group the problems of machine translation, as have several of the leading workers in the Georgetown project.-- Of all groups, that at Georgetown exhibits the greatest diversity of approaches. Two recent large-scale tests, one on the translation of Russian, the other on the translation of French, demonstrate that the Georgetown procedures have been successful, at least to the extent one might wish for a new science. Translations were produced, but improvements are necessary. A problem which remains to be answered is whether the programs which brought initial and limited success can be modified readily to secure the high-quality translation which is now expected of machine translation.

### The Massachusetts Institute of Technology

group, under the leadership of Victor Yngve, is committed to a full-scale, detailed analysis of German and of English as a prerequisite for machine translation. Using the standard grammar of German in English (G. O. Curme, _A Grammar of the German Language_) as a basis and extending its statements by consultation with native informants, this group attempts to arrive at analyses, e.g. of the noun phrase, which will satisfy the most rigorous demands in all situations and degrees of grammaticalness. In addition to the linguistic analysis carried out by J. R. Applegate and G. H. Matthews, this group has devoted its attention to the elaboration of a notational system for the writing of routines by linguistic analysts (Victor H. Yngve, " A Programming Language for Mechanical Translation", _Mechanical Translation_ 5.(1958) 25-41). Whether this particular system is generally adopted by researchers in the field or not, it will have contributed a great deal toward an understanding of the kind of communication which is necessary and possible among linguists, programmers and machines. A further advance of the M.I.T. group has been the preparation of a program for converting linotype tape to machine input as a first step in the machine analysis of language. This device may well provide us with the extensive corpus needed for research without an expensive card-punch operation. (Cf. Rand Corporation Research Memorandum. _Studies in Machine Translation_ - 9 Bibliography of Russian Scientific Articles")

The publication of the journal _Mechanical Translation_ at M.I.T. has been a welcome forum and source of information about developments in the field.

The M.I.T. group has also contributed generously to the recruitment and training

The Rand Corporation group has devoted itself to the solution of procedural and mechanical problems of machine translation with special attention to the preparation of material for machine storage and analysis and to the matter of pre- and post- editing of Russian scientific texts. This group has also cooperated with the Georgetown project in standardizing card formats for interchange of material. Corpora used in the tests of methods of research at Georgetown have in part been prepared on cards by The Rand Corporation.

The Washington group seems at present to be working along lines sketched above for the Michigan group.

Groups abroad:

Groups in England have done considerable excellent work on machine translation. The fullest analysis of German to date is contained in the book of Booth, Brandwood and Cleave, which reports the work of the group at Birkbeck College, of the University of London.-- The group at Cambridge University has done considerable study on a thesaurus approach which remains to be thoroughly described (cf. Bar Hillel pp. 35-37).

The Soviet Union is now supporting by far more work in machine translation than is any other country. Fortunately the work is well summarized in Delavaney's book. Important papers are made available in translation by the National Science Foundation. In theory and linguistic analysis the Soviet groups seem to be the equal of any other. There is as yet no indication however of the extent to which they have tested their findings and produced acceptable translations.-- It may be useful to restate here the five basic principles which according to Delavenay, pp. 51-52, were still held to be fundamental by the Soviet linguists in October 1958.

1. To separate as much as possible the glossary from the translation program.

2. To divide the translation program into two independent parts: analysis of the sentence in the source language and synthesis of the corresponding target (Russian) sentence. The aim here is to use the same synthesis program for a number of source languages.

3. To store in the glossary all words in their basic form. The aim here is to use standard Russian grammars in the synthesis of Russian.

4. To store in the glossary all the invariant grammatical characteristics of words.

5. To determine the particular sense of a word with multiple meanings from context and its variable grammatical characteristics after analysis of the grammatical structure of the clause.

Conclusions:

The studies carried out indicate the procedures necessary for completing an analysis of German adequate for programming. In morphological analysis, identification of the finite verb is of primary importance for machine translation and this accordingly was given high priority, as indicated by the study included in the report. In syntactic analysis, identification of the subject is of primary importance, and again high priority was assigned to unambiguous means for achieving it. As indicated above, identification was possible without ambiguity in more than 99 % of the sentences; procedures are being investigated which will enable us to deal with the remainder.

Work on programming sketched the broad procedures to be used. A projection of the programming along conventional lines has been made. Further investigation has been carried out of analysis by machine.

## Appendix 1.

### GERMAN VERB CLASSES AND PARADIGMS

The verbs of German have been assigned to classes on the basis of stem selection and paradigmatic categories. As part of the glossary storage for each verb stem, this information serves as criteria for machine analysis and recognition of verb forms which are not stored as separate glossary entries. With a subtraction routine of five letters as a maximum, any verb form in German can be recognized and identified morphologically.

The classification also provides the basis for a generation procedure for verb forms when German is the target language.

Verbs with separable prefixes are not included in the list of German verb classes, but they can be provided in an extension of the system.

### A. Verb Stem Distribution by Classes.

STEM

1 - infinite stem:
stem of 1st, 2nd (Sie-form) and 3rd person singular
and plural present subjunctive 1;
stem of present participle;

STEM

2 = stem of 1st person singular present indicative;

3 = stem of 3rd person singular present indicative;

4 = stem of 1st and 3rd person plural present indicative;
stem of 2nd person (Sie-form) present indicative;

5 = stem of 1st, 2nd (Sie-form) and 3rd person singular and plural past tense indicative;

6 = stem of 1st and 3rd person singular and plural present subjunctive II;

7 = stem of past participle;

8 = infinitive functioning as past participle: past partic. 2 (occurring with few verbs only.)

omitted: 2nd person singular and plural familiar form, present and past tense, indicative, subjunctive and imperative mood.

imperative Sie - form is considered present subjunctive I.

CLASS

1) stem 1 = 1, 2, 3, 4, 5, 6, 7 ($ 00, S 01, S 02) verkaufen,
erwarten,
belagern

2) stem 1 = 1, 2, 3, 4, 5, 6 ($ 03, S 04, S 05) kaufen, warten
lagern

stem 2 = 7 ($ 06)

3) 1, 2, 3, 4, 6 ($ 07) verbrennen
5, 7 ($ 10)

| CLASS | | |
|---|---|---|
| 4) 1, 2, 3, 4, 7 | ($ 11) | berufen |
| 5, 6 | ($ 12) | |
| 5) 1, 2, 3, 4 | ($ 13) | verbleiben |
| 5, 6, 7 | ($ 14) | |
| 6) 1, 3, 4, 5, 6, 7 | ($ 15) | versammeln |
| 2 | ($ 16) | |
| 7) 1, 2, 3, 4, 5, 6, 7 | | bestecken |
| 5 (a.) | ($ 17) | |
| 6 | ($ 20) | |
| 8) 1, 2, 3, 4, 6 | ($ 21) | brennen |
| 5 (b) | ($ 22) | |
| 7 | | |
| 9) 1, 2, 3, 4 | | bleiben |
| 5, 6 | | |
| 7 | | |
| 10) 1, 2, 3, 4 | | verfliegen |
| 5, 7 (b) | ($ 23) | |
| 11) 1, 2, 4, 5, 6, 7 | ($ 24) | erlöschen |
| 3 | ($ 25) | |
| 5, 7 | | |
| 12) 1, 3, 4, 5, 6 | | sammeln |
| 2 | ($ 26) | |
| 7 | | |
| 13) 1, 2, 3, 4, 5, 6 | | stecken |
| 5 | | |
| 6 | | |
| 7 | | |
| 14) 1, 2, 3, 4 | | beschwören |
| 5, 7 | | |
| 5 | | |
| 6 | | |

| CLASS | | |
|---|---|---|
| 15) 1, 2, 3, 4 | | fliegen |
| 5 | | |
| 6 | | |
| 7 | | |
| 16) 1, 2, 4, 6 | ($ 27) | gebären |
| 3 | | |
| 5 | | |
| 7 | | |
| 17) 1, 2, 4, 7 | ($ 30) | befahren |
| 3 | | |
| 5 | | |
| 6 | | |
| 18) 1, 2, 4 | ($ 31) | verfechten |
| 3 | | |
| 5, 7 | | |
| 6 | | |
| 19) 1, 2, 3, 4 | | schwören |
| 5 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 20) 1, 2, 3, 4 | | beginnen |
| 5 | | |
| 6 | | |
| 6 | | |
| 7 | | |
| 21) 1, 2, 4 | | fahren |
| 3 | | |
| 5 | | |
| 6 | | |
| 7 | | |

| CLASS | |
|---|---|
| 22) 1, 2, 4 | zerdreschen |
| 3 | |
| 5 | |
| 5, 7 | |
| 6 | |
| 6 | |
| 23) 1, 2, 4 | bersten |
| 3 | |
| 5 | |
| 6 | |
| 6 | |
| 7 | |
| 24) 1, 2, 4 | dreschen |
| 3 | |
| 5 | |
| 5 | |
| 6 | |
| 6 | |
| 7 | |

the following classes have only 1 or 2 members each:

| | | |
|---|---|---|
| 25) 1, 2, 3, 4, 5, 6, 8 | ($ 32) | sollen |
| 7 | | hören |
| 26) 1, 4, 5, 6, 8 | ($ 33) | wollen |
| 2, 3 | ($ 34) | |
| 7 | | |
| 27) 1, 4, 6, 8 | ($ 35) | müssen |
| 2, 3, 5 | ($ 36) | |
| 7 | | |
| 28) 1, 4, 6 | ($ 37) | bedürfen |
| 2, 3 | | |
| 5, 7 | | |

CLASS

| | | |
|---|---|---|
| 29) 1, 2, 4, 8 | ($ 40) | lassen |
| 3 | | |
| 5, 6 | | |
| 7 | | |
| 30) 1, 4, 6, 8 | | können |
| 2, 3 | | dürfen |
| 5 | | |
| 7 | | |
| 31) 1, 4 | ($ 41) | vermögen |
| 2, 3 | | |
| 5, 7 | | |
| 6 | | |
| 32) 1, 2, 4, 8 | | sehen |
| 3 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 33) 1, 4, 8 | ($ 42) | mögen |
| 2, 3 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 34) 1, 4 | | wissen |
| 2, 3 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 35) 1, 2, 4, 8 | | helfen |
| 3 | | |
| 5 | | |
| 6 | | |
| 6 | | |
| 7 | | |

| CLASS | | |
|---|---|---|
| 36) 1, 2, 4 | | werden |
| 3 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | ($ 43) | |
| 37) 1 | ($ 44) | sein |
| 2 | | |
| 3 | | |
| 4 | ($ 45) | |
| 5 | | |
| 6 | | |
| 7 | | |

## B. Verb Classes (Stems and Endings).

1.a. verkaufen:

| | |
|---|---|
| verkauf | -en |
| | -end |
| | -ende |
| | -enden |
| | -ender |
| | -endes |
| | -endem |
| | -e |
| | -t |
| | -te |
| | -ten |
| | -ter |
| | -tes |
| | -tem |

2.a. kaufen(cont.)

| | |
|---|---|
| | -e |
| | -t |
| | -te |
| | -ten |

2.b. warten:

| | | | |
|---|---|---|---|
| warte | -n | gewartet | -∅ |
| | -nd | | -e |
| | -nde | | -en |
| | -nden | | -er |
| | -nder | | -es |
| | -ndes | | -em |
| | -ndem | | |
| | -∅ | | |
| | -t | | |
| | -te | | |
| | -ten | | |

2.c. lagern:

| | | | |
|---|---|---|---|
| lager | -n | gelagert | -∅ |
| | -nd | | -e |
| | -nde | | -en |
| | -nden | | -er |
| | -nder | | -es |
| | -ndes | | -em |
| | -ndem | | |
| | -e | | |
| | -t | | |
| | -te | | |
| | -ten | | |

3. verbrennen:

| | | | |
|---|---|---|---|
| verbrenn | -en | verbrannt | -e |
| | -end | | -en |
| | -ende | | -∅ |
| | -enden | | -er |
| | -ender | | -es |
| | -endes | | -em |
| | -endem | | |

3. <u>verbrennen</u> (cont.):

| | | | |
|---|---|---|---|
| | -e | | |
| | -t | | |
| | -te | | |
| | -ten | | |

4. <u>berufen:</u>

| | | | |
|---|---|---|---|
| berufe | -n | berief | -∅ |
| | -nd | | -en |
| | -nde | | -e |
| | -nden | | |
| | -nder | | |
| | -ndes | | |
| | -ndem | | |
| | -∅ | | |
| | -t | | |
| | -ne | | |
| | -nen | | |
| | -ner | | |
| | -nes | | |
| | -nem | | |

5. <u>verbleiben:</u>

| | | | |
|---|---|---|---|
| verbleib | -en | verblieb | -∅ |
| | -end | | -en |
| | -ende | | -e |
| | -enden | | -ene |
| | -ender | | -enen |
| | -endes | | -ener |
| | -endem | | -enes |
| | -e | | -enem |
| | -t | | |

6. <u>versammeln:</u>

| | | | |
|---|---|---|---|
| versammel | -n | versammle | -∅ |
| | -nd | | |
| | -nde | | |
| | -nden | | |
| | -nder | | |
| | -ndes | | |
| | -ndem | | |

6. versammeln(cont.):

-t
-te
-ten
-ter
-tes
-tem

7. bestecken:

| | | | | | |
|---|---|---|---|---|---|
| besteck | -en | bestak | -Ø | bestaeke | -Ø |
| | -end | | -en | | -n |
| | -ende | | | | |
| | -enden | | | | |
| | -ender | | | | |
| | -endes | | | | |
| | -endem | | | | |
| | -e | | | | |
| | -t | | | | |
| | -te | | | | |
| | -ten | | | | |
| | -ter | | | | |
| | -tes | | | | |
| | -tem | | | | |

8. brennen:

| | | | | | |
|---|---|---|---|---|---|
| brenn | -en | brannte | -Ø | gebrannt | -Ø |
| | -end | | -n | | -e |
| | -ende | | | | -en |
| | -enden | | | | -er |
| | -ender | | | | -es |
| | -endes | | | | -em |
| | -endem | | | | |
| | -e | | | | |
| | -t | | | | |
| | -te | | | | |
| | -te | | | | |

9. bleiben: (1)

| | | | | | |
|---|---|---|---|---|---|
| bleib | -en | blieb | -∅ | geblieben | -∅ |
| | -end | | -en | | -e |
| | -ende | | -e | | -en |
| | -enden | | | | -er |
| | -ender | | | | -es |
| | -endes | | | | -em |
| | -endem | | | | |
| | -e | | | | |
| | -t | | | | |

10. a. verfliegen:

| | | | | | |
|---|---|---|---|---|---|
| verflieg | -en | verflog | -∅ | verfloege | -∅ |
| | -end | | -en | | -n |
| | -ende | | -ene | | |
| | -enden | | -enen | | |
| | -ender | | -ener | | |
| | -endes | | -enes | | |
| | -endem | | -enem | | |
| | -e | | | | |
| | -t | | | | |

10. b. verbringen:

| | | | | | |
|---|---|---|---|---|---|
| verbring | -en | verbracht | -e | verbraechte | -∅ |
| | -end | | -en | | -n |
| | -ende | | -∅ | | |
| | -enden | | -er | | |
| | -ender | | -es | | |
| | -endes | | -em | | |
| | -endem | | | | |
| | -e | | | | |
| | -t | | | | |

11. erlöschen:

| | | | | | |
|---|---|---|---|---|---|
| erloesch | -en | erlischt | -∅ | erlosch | -∅ |
| | -end | | | | -en |
| | -ende | | | | -ene |
| | -enden | | | | -enen |

11. erlöschen (cont.):

| | | | | | |
|---|---|---|---|---|---|
| | -ender | | | | -ener |
| | -endes | | | | -enes |
| | -endem | | | | -enem |
| | -e | | | | |
| | -te | | | | |
| | -ten | | | | |
| | -t | | | | |
| | -ter | | | | |
| | -tes | | | | |
| | -tem | | | | |

12. sammeln:

| | | | | | |
|---|---|---|---|---|---|
| sammel | -n | sammle | --Ø | versammelt | -Ø |
| | -nd | | | | -e |
| | -nde | | | | -en |
| | -nden | | | | -er |
| | -nder | | | | -es |
| | -ndes | | | | -em |
| | -ndem | | | | |
| | -t | | | | |
| | -te | | | | |
| | -ten | | | | |

13. stecken:

| | | | | | |
|---|---|---|---|---|---|
| steck | -en | stak | -Ø | staeke | -Ø |
| | -end | | -en | | -n |
| | -ende | | | | |
| | -enden | | | | |
| | -ender | | | | |
| | -endes | | | gesteckt | -Ø |
| | -endem | | | | -e |
| | -te | | | | -en |
| | -ten | | | | -er |
| | -e | | | | -es |
| | -t | | | | -em |

14. beschwoeren:

| | | | | | |
|---|---|---|---|---|---|
| beschwoer | -en | beschwor | -∅ | beschwur | -∅ |
| | -end | | -en | | -en |
| | -ende | | -ene | | |
| | -enden | | -enen | | |
| | -ender | | -ener | | |
| | -endes | | -enes | | |
| | -endem | | -enem | | |
| | -e | | | | |
| | -t | | | | |
| | | | | beschwuere | -∅ |
| | | | | | -n |

15. a. fliegen:

| | | | | | |
|---|---|---|---|---|---|
| flieg | -en | flog | -∅ | floege | -∅ |
| | -end | | -en | | -n |
| | -ende | | | | |
| | -enden | | | | |
| | -ender | | | | |
| | -endes | | | geflogen | -∅ |
| | -endem | | | | -e |
| | -e | | | | -en |
| | -t | | | | -er |
| | | | | | -es |
| | | | | | -em |

15. b. bringen:

| | | | | | |
|---|---|---|---|---|---|
| bring | -en | brachte | -∅ | braechte | -∅ |
| | -end | | -n | | -n |
| | -ende | | | | |
| | -enden | | | | |
| | -ender | | | | |
| | -endes | | | | |
| | -endem | | | gebracht | -∅ |
| | -e | | | | -e |
| | -t | | | | -en |
| | | | | | -er |
| | | | | | -es |
| | | | | | -em |

16. <u>gebaeren:</u>

| | | | | | |
|---|---|---|---|---|---|
| gebaere | -n | gebiert | -Ø | gebar | -Ø |
| | -nd | | | | -en |
| | -nde | | | | |
| | -nden | | | | |
| | -nder | | | | |
| | -ndes | | | | |
| | -ndem | | | geboren | -Ø |
| | -Ø | | | | -e |
| | | | | | -en |
| | | | | | -er |
| | | | | | -es |
| | | | | | -em |

17. <u>befahren:</u>

| | | | | | |
|---|---|---|---|---|---|
| befahre | -n | befaehrt | -Ø | befuhr | -Ø |
| | -nd | | | | -en |
| | -nde | | | | |
| | -nden | | | | |
| | -nder | | | | |
| | -ndes | | | | |
| | -ndem | | | befuehre | -Ø |
| | -ne | | | | -n |
| | -nen | | | | |
| | -ner | | | | |
| | -nes | | | | |
| | -nem | | | | |

18. <u>verfechten:</u>

| | | | | | |
|---|---|---|---|---|---|
| verfechte | -Ø | verficht | -Ø | verfocht | -Ø |
| | -n | | | | -en |
| | -nd | | | | -ene |
| | -nde | | | | -enen |
| | -nden | | | | -ener |
| | -nder | | | | -enes |
| | -ndes | | | | -enem |
| | -ndem | | | | |

19. schwoeren:

| | | | | | |
|---|---|---|---|---|---|
| schwoer | -en | schwur | -∅ | schwor | -∅ |
| | -end | | -en | | -en |
| | -ende | | | | |
| | -ender | | | | |
| | -enden | | | | |
| | -endes | | | | |
| | -endem | schwuere | -∅ | geschworen | -∅ |
| | -e | | -n | | -e |
| | -t | | | | -en |
| | | | | | -er |
| | | | | | -es |
| | | | | | -em |

20. beginnen:

| | | | | | |
|---|---|---|---|---|---|
| beginn | -en | begann | -∅ | begaenne | -∅ |
| | -end | | -en | | -n |
| | -ende | | | | |
| | -enden | | | | |
| | -ender | | | | |
| | -endes | begoenne | -∅ | begonnen | -∅ |
| | -endem | | -n | | -e |
| | -e | | | | -en |
| | -t | | | | -er |
| | | | | | -es |
| | | | | | -em |

21. fahren:

| | | | | | |
|---|---|---|---|---|---|
| fahre | -n | faehrt | -∅ | fuhr | -∅ |
| | -nd | | | | -en |
| | -nde | | | | |
| | -nden | fuehre | -∅ | gefahren | -∅ |
| | -nder | | -n | | -e |
| | -ndes | | | | -en |
| | -ndem | | | | -er |
| | -∅ | | | | -es |
| | | | | | -em |

22. zerdreschen (12):

| | | | | | |
|---|---|---|---|---|---|
| zerdresche | -n | zerdrischt | -∅ | zerdrasch | -∅ |
| | -nd | | | | -en |
| | -nde | | | | |
| | -nden | | | | |
| | -nder | | | | |
| | -ndes | | | | |
| | -ndem | | | | |
| | -∅ | zerdrosch | -∅ | | |
| | | | -en | | |
| | | | -ene | | |
| | | | -enen | | |
| | | | -ener | | |
| | | | -enes | | |
| | | | -enem | | |
| | | zerdraesche | -∅ | zerdroesche | -∅ |
| | | | -n | | -n |

23. bersten:

| | | | | | |
|---|---|---|---|---|---|
| berste | -n | birst | -∅ | barst | -∅ |
| | -nd | | | | -en |
| | -nde | | | | |
| | -nden | | | | |
| | -nder | baerste | -∅ | boerste | -∅ |
| | -ndes | | -en | | -n |
| | -ndem | | | | |
| | -∅ | | | | |
| | | geborsten | -∅ | | |
| | | | -e | | |
| | | | -en | | |
| | | | -er | | |
| | | | -es | | |
| | | | -em | | |

24. dreschen:

| | | | | | |
|---|---|---|---|---|---|
| dresche | -n | drischt | -∅ | drasch | -∅ |
| | -nd | | | | -en |
| | -nde | | | | |
| | | drosch | -∅ | | |
| | | | -en | | |

24. <u>dreschen</u> (cont.):

-nden
-nder
-ndes
-ndem
-Ø

| | | | |
|---|---|---|---|
| draesche | -Ø<br>-n | droesche | -Ø<br>-n |

gedroschen -Ø
-e
-en
-er
-es
-em

## C. Paradigm Classes by Stems.

| | | | |
|---|---|---|---|
| $00 | verkauf | -en | inf; 1, 2, 3 pl. pres ind;1, 2, 3 pl. pres subj I; |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -e | 1 sg pres ind; 1, 3 sg pres. subj I |
| | | -t | 3 sg pres ind; past part. |
| | | -te | 1, 3 sg past ind; 1, 3 sg pres subj II;past part infl. |
| | | -ten | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II; past part infl. |
| | | -ter | past part. infl. |
| | | -tes | |
| | | -tem | |
| $01 | erwarte | -n | inf; 1, 2, 3 pl. pres ind;1, 2, 3 pl. pres subj I; |
| | | -nd | pres. part. |
| | | -nde | pres. part infl. |
| | | -nden | |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |
| | | -∅ | 1 sg pres ind; 1, 3 sg pres. subj I |
| | | -t | 3 sg pres ind; past part. |
| | | -te | 1, 3 sg past ind; 1, 3 sg pres subj II; past part infl. |
| | | -ten | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II; past part |
| | | -ter | past. part. infl. |
| | | -tes | |
| | | -tem | |
| $02 | belager | -n | inf; 1, 2, 3 pl. pres ind; 1, 2, 3 pl. pres subj I; |
| | | -nd | pres. part. |
| | | -nde | pres. part. infl. |
| | | -nden | |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |

| | | | |
|---|---|---|---|
| $02 (cont.) | | | |
| | | -e | 1 sg pres ind; 1, 3 sg pres subj I |
| | | -t | 3 sg pres ind; past part |
| | | -te | 1, 3 sg past ind; 1, 3 sg pres subj II; past part infl. |
| | | -ten | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II; past part i. |
| | | -ter | past. part. infl. |
| | | -tes | |
| | | -tem | |
| $03 | kauf | -en | inf; 1, 2, 3 pl pres ind; 1, 2, 3 pl pres subj I; |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -e | 1 sg pres ind; 1, 3 sg pres subj I |
| | | -t | 3 sg pres ind |
| | | -te | 1, 3 sg past ind; 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II |
| $04 | warte | -n | 1, 2, 3 pl pres ind; 1, 2, 3 pl pres subj I; inf. |
| | | -nd | pres. part. |
| | | -nden | pres. part. infl. |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |
| | | - ∅ | 1 sg pres ind; 1, 3 sg pres subj I |
| | | -t | 3 sg pres ind |
| | | -te | 1, 3 sg past ind; 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II |
| $05 | lager | -n | inf; , 1, 2, 3 pl pres ind; 1, 2, 3 pl pres subj I; |
| | | -nd | pres. part. |
| | | -nde | pres. part. infl. |
| | | -nden | |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |

| | | | |
|---|---|---|---|
| $05 (cont.) | | | |
| | | -e | 1 sg pres ind; 1, 3 sg pres subj I |
| | | -t | 3 sg pres ind |
| | | -te | 1, 3 sg past ind; 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II |
| $06 | gekauft | -∅ | past part. |
| | | -e | past. part. infl. |
| | | -en | |
| | | -er | |
| | | -es | |
| | | -em | |
| $07 | verbrenn | -en | inf; , 1, 2, 3 pl. pres. ind;1, 2, 3 pl. pres. subj I |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -e | 1 sg pres. ind. ;1, 3 sg pres subj I |
| | | -t | 3 sg pres. ind. |
| | | -te | 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl. pres. subj II |
| $10 | verbrannt | -e | 1, 3 sg past ind; past part. i. |
| | | -en | 1, 2, 3 pl past ind; past part. i. |
| | | -∅ | past part |
| | | -er | past part. infl. |
| | | -es | |
| | | -em | |
| $11 | beruf | -en | inf; , 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I, past p. |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |

| | | | |
|---|---|---|---|
| $11 (cont.) | | | |
| | | -e | 1 sg pres ind; 1, 3 sg pres subj I |
| | | -t | 3 sg pres ind |
| | | -ene | past. part. infl. |
| | | -enen | |
| | | -ener | |
| | | -enes | |
| | | -enem | |
| $12 | berief | -ø | 1, 3 sg past ind; |
| | | -en | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II |
| | | -e | 1, 3 sg pres subj II |
| $13 | verbleib | -en | inf; 1, 2, 3 pl pres ind; 1, 2, 3 pl pres subj I |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -e | 1 sg pres ind;1, 3 sg pres subj I |
| | | -t | 3sg pres ind. |
| $14 | verblieb | -ø | 1, 3 sg past ind |
| | | -en | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II; past p. |
| | | -e | 1, 3 sg pres subj II |
| | | -ene | past. part. infl. |
| | | -enen | |
| | | -ener | |
| | | -enes | |
| | | -enem | |
| $15 | versammel | | |
| | | -n | inf. 1, 2, 3 pl pres ind; 1, 2, 3 pl pres subj I; |
| | | -nd | pres. part. |
| | | -nde | pres. part. infl. |
| | | -nden | |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |
| | | -t | 3 sg pres ind; past part. |
| | | -te | 1, 3 sg past ind;1, 3 sg pres subj II; past part. infl. |

| | | | |
|---|---|---|---|
| $15(cont.) | | | |
| | | -ten | 1, 2, 3 pl past ind;1, 2, 3 pl pres subj II; past. p.infl. |
| | | -ter | past-part. infl. |
| | | -tes | |
| | | -tem | |
| $16 | versammle | -Ø | 1 sg pres ind; 1, 3 sg pres subj I |
| $17 | flog | -Ø | 1, 3 past ind (bestak) |
| | | -en | 1, 2, 3 pl past ind |
| $20 | floege | -Ø | 1, 3 sg pres subj II (bestäke) |
| | | -n | 1, 2, 3 pl pres subj II |
| $21 | brenn | -en | inf; , 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -e | 1 sg pres ind;1, 3 sg pres subj I |
| | | -t | 3 sg pres ind |
| | | -te | 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl pres subj II |
| $22 | brannte | -Ø | 1, 3 sg past ind |
| | | -n | 1, 2, 3 past ind |
| $23 | verflog | -Ø | 1, 3 sg past ind |
| | | -en | 1, 2, 3 pl past ind;past. part. |
| | | -ene | past. part. infl. |
| | | -enen | |
| | | -ener | |
| | | -enes | |
| | | -enem | |

|  |  |  |  |
|---|---|---|---|
| $24 | erloesch | -en | inf; 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I; |
|  |  | -end | pres. part. |
|  |  | -ende | pres. part. infl. |
|  |  | -enden |  |
|  |  | -ender |  |
|  |  | -endes |  |
|  |  | -endem |  |
|  |  | -e | 1 sg pres ind; 1, 3 sg pres subj I |
|  |  | -te | 1, 3 sg pres subj II; past part. |
|  |  | -ten | 1, 2, 3 pl pres subj II;past part.infl. |
|  |  | -∅ | 3 rd sing pres ind; past.part. |
|  |  | -ter | past. part. infl. |
|  |  | -tes |  |
|  |  | -tem |  |
| $25 | erlischt | -∅ | 3 sg pres ind. |
| $26 | sammel | -n | 1, 3 pl pres ind; 1, 2, 3 pres subj I;infinitive |
|  |  | -nd | pres.part. |
|  |  | -nde | pres. part. infl. |
|  |  | -nden |  |
|  |  | -nder |  |
|  |  | -ndes |  |
|  |  | -ndem |  |
|  |  | -t | 3 sg pres ind |
|  |  | -te | 1, 3 sg past ind;1, 3 sg pres subj II |
|  |  | -ten | 1, 2, 3 pl past ind; 1, 2, 3 pl pres subj II |
| $27 | gebaer | -en | inf. 1, 3 pl pres ind; 1, 3 pl pres subj I; |
|  |  | -end | pres. part. |
|  |  | -ende | pres. part. infl. |
|  |  | -enden |  |
|  |  | -ender |  |
|  |  | -endes |  |
|  |  | -endem |  |
|  |  | -e | 1 sg pres ind; 1, 3 sg pres subj I |
|  |  | -te | 1, 3 sg pres subj II |
|  |  | -ten | 1, 2, 3 pl pres subj II |

| | | | |
|---|---|---|---|
| $30 | befahre | -n | inf 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I;past. part. |
| | | -∅ | 1 sg pres ind; 1, 3 sg pres subj I |
| | | -ne | past. part. infl. |
| | | -nen | |
| | | -ner | |
| | | -nes | |
| | | -nem | |
| | | -nd | pres. part. |
| | | -nde | pres. part. infl. |
| | | -nden | |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |
| $31 | verfechte | -n | inf;1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I |
| | | -nd | pres. part. |
| | | -nde | pres. part. infl. |
| | | -nden | |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |
| | | -∅ | 1 sg pres ind;1, 3 sg pres subj I |
| $32 | soll | -en | 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I; inf; past.part. |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -∅ | 1, 3 sg pres ind |
| | | -e | 1, 3 sg pres subj I |
| | | -te | 1, 3 sg past ind; 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl past ind;1, 2, 3 pl pres subj II |

| | | | |
|---|---|---|---|
| $33 | woll | -en | inf; 1, 2, 3 pl pres ind;1, 2, 3 pres subj I; past part. 2 |
| | | -end | pres. part. |
| | | -ende | pres. part.infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -e | 1, 3 sg pres subj I |
| | | -te | 1, 3 sg past ind; 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl past ind;1, 2, 3 pl pres subj II |
| $34 | will | -∅ | 1, 3 sg pres ind. |
| $35 | müss | -en | inf 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I; past. part. 2 |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -e | 1, 3 sg pres subj I |
| | | -te | 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl pres subj II |
| $36 | muss | -∅ | 1, 3 sg pres ind |
| | | -te | 1, 3 sg past ind |
| | | -ten | 1, 2, 3 pl past ind |
| $37 | beduerf | -en | inf; 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I; |
| | | -end | pres. part. |
| | | -ende | pres. part. infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -te | 1, 3 sg pres subj II |
| | | -ten | 1, 2, 3 pl pres subj II |

| | | | |
|---|---|---|---|
| $40 | lasse | -∅ | 1 sg pres ind;1, 3 sg pres subj I |
| | | -n | inf; 1, 3 pl pres ind;1, 3 pres subj I; past part. 2 |
| | | -nd | pres. part. |
| | | -nden | pres. part. infl. |
| | | -nder | |
| | | -ndes | |
| | | -endem | |
| $41 | vermöge | -n | inf; 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I; |
| | | -nd | pres. part. |
| | | -nden | pres. part. infl. |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |
| | | -∅ | 1, 3 sg pres subj. |
| $42 | möge | -n | inf; 1, 2, 3 pl pres ind;1, 2, 3 pl pres subj I; past part. 2 |
| | | -nd | pres. part. |
| | | -nden | pres. part. infl. |
| | | -nder | |
| | | -ndes | |
| | | -ndem | |
| | | -∅ | 1, 3 sg pres subj I |
| $43 | worden | -∅ | past part. 2 |
| $44 | sei | -n | inf. |
| | | -end | pres. part. |
| | | -ende | pres.part.infl. |
| | | -enden | |
| | | -ender | |
| | | -endes | |
| | | -endem | |
| | | -∅ | 1, 3 sg pres subj I |
| | | -en | 1, 2, 3 pl pres subj I |
| $45 | sind | -∅ | 1, 3 pl pres ind. |

## Appendix 2.

### On the resolution of subject-object ambiguity in German texts.

In present-day English declarative sentences, the distinction between noun as the subject and noun as the object is made entirely on the basis of position in respect to the verb: the subject always precedes the verb; what follows it, can never be the subject. In contrast to this situation, we find two German sequences to be acceptable renderings of an English sentence such as 'The dog bit the boy', viz., '<u>Der Hund biss den Jungen</u>', and '<u>Den Jungen biss der Hund</u>'. Either formulation is quite unambiguous in all contexts and in isolation; '<u>der Hund</u>' is marked as subject case, and as subject case only, '<u>den Jungen</u>', in combination with a transitive verb, can only be identified as a direct object. Such clear formal marking of the subject or object function of a noun is, however, restricted to masculines; feminine, neuter, and plural nouns show no formal distinction between subject and object case. Thus, the two possible translations of 'The cat saw the girl' coincide formally with those of

'The girl saw the cat'. Confronted with '*Das Mädchen sah die Katze*' or '*Die Katze sah das Mädchen*' without further context, a native speaker will probably first provide only one translation, but then point out that also the reverse would be a possible rendering. Given more context - or given the help of intonational features -, the native speaker will probably reject completely the possibility of an ambiguity; but clearly neither of these two solutions is of any immediate practical value for the purposes of machine translation.

The investigation covered by this report presents an attempt to determine the distribution of subject and object case with regard to the verb in large bodies of actual text, to ascertain the amount of actually occurring subject-object ambiguity, and to explore avenues for the eventual elimination of such ambiguities. The materials studied were chosen from a wide array of text types; to gain as representative a sample as possible, non-scientific texts were also included.

The attached table shows the most significant results of the frequency counts. The first point which deserves emphasis is, that the incidence of subject preceding verb by far exceeds that of object preceding verb: In 10,450 main clauses analyzed in Jung, Snell, Hauptmann, Deutschland heute, and Mann (Faustus), we find 6,167 occurrences of subject preceding the verb as against 329 instances of object in first position. This suggests very stongly the possibility of positing subject-verb-object as the normal sentence type, with object-verb-subject as the secondary modification. The incidence of the inverted sequence is, however, too high to permit neglecting it even in a preliminary and rough translation: an error rate of slightly more than 3 °/o is excessive in view of the fact that we do not know as of now the extent of other possible errors outside the subject-object area.

A further breakdown of the material shows, however, that the rate of error can be reduced mechanically. Of the 329 items listed as object on the basis of syntactic and semantic considerations, all but 67 can be positively identified as objects by non-semantic, morphological or syntactic criteria - such criteria being unambiguous object form of the preverbal item (e.g., ihn, den Mann, etc.), unambiguous subject form of the postverbal item (e.g., man, der Mann, etc.), number match of the verb with the postverbal item only (e.g., das Kind sahen die Frauen, etc.).

Such findings suggest the following procedure for subject identification:

[1] Search for unambiguous subject.

[2] Search for unambiguous object.

[3] Search for unambiguous noun-verb match.

[4] If [1] - [3] fail, identify preverbal potential subject as actual subject.

In our example, we would be left with 463 subject/objects unaccounted for after [3]. The application of [4] would lead to 396 correct and 67 wrong identifications. If an error of 0.6°/o (which is higher than for other sets of texts, cf. the table) is tolerable, the procedure can stand unamended; it does, however, seem preferable to have each result of [4] accompanied in its final form with a signal indicating that the reader may have to reverse the order of subject and object to get the right meaning. If, for instance, a signal S00S were printed out with every sentence on which [4] had been applied, the amount of complication to the reader would seem tolerable, while at the same time the programming of apparently highly complex subsidiary routines for a reduction of subject-object ambiguities could be reserved for later stages in the development of machine translation.

As far as can be detected, such routines would have to involve special procedures for certain fixed expressions (..gibt es, etc.) and distinction of animate vs. inanimate nouns in the glossary: it appears that if both an animate and an inanimate noun appear as potential subjects in a given clause, the animate item has much greater likelihood of being the subject. Such a distinction may have to be considered once the problem of translating German personal pronouns is decided; on the basis of the advantages in subject recognition alone a pervading reorganization of the proposed glossary would hardly seem feasible.

Even if the points just discussed were solved by subroutines, a small but not negligible residue of ambiguous pairs would remain. There is no point whatsoever in accommodating a sentence like '*Ein solches* (i. e. dogma) *kannte die griechische Religion nicht*' (Snell) 'Greek religion did not know such a dogma' by subroutine - '*Religion*' has no greater likelihood of reoccurring as subject than '*Dogma*' would have. In view of such formally unsolvable ambiguities, and in view of the considerable gain in simplicity of the program, it is suggested that the abbreviated procedure with introduction of the 'possible error' signal (S00S) be used at least for the time being.

Present data cover only main declarative clauses: the investigation of subordinate clauses is under way and should provide comparable data in the near future. Also, a study of main clauses with adverbial in first position is being made with the purpose of determining the feasibility of mechanical subject/object identification in other contexts than the one described here.

Team: Nanette Orme-Johnson
Patrick Pollard
Werner Winter (reporter)

## Analysis of preverbial material in Declarative Main Clauses

| FIELD AND SOURCE: | NUMBER OF MC ANALYZED: | SUBJECT INCL. PREDICATE NOUN: | SUBJECT FORMALLY AMBIGUOUS: | DIRECT OBJECT: | OBJECT FORMALLY AMBIGUOUS: | PREPOSITIONAL PHRASE AND ADVERB: |
|---|---|---|---|---|---|---|
| PHYSICS:<br>Heisenberg,<br>Physik der Atomkerne | 2195 | 1227 = 55.9% | 60 = 2.7% | 71 = 3.2% | 8 = 0.4% | 815 = 37.1% |
| v. Weizsäcker<br>Atomenergie und<br>Atomzeitalter | 2080 | 1306 = 63.4% 1) | 102= 4.9% | 88 = 4.2% | 33 = 1.6% | 667 = 32.0% |
| POLITICAL SCIENCE:<br>Deutschland heute | 2021 | 1134 = 56.1% 2) | 124= 6.1% | 40 = 2.0% | 12 = 0.6% | 790 = 39.0% |
| PSYCHOLOGY:<br>Jung, Symbolik des<br>Geistes | 2119 | 1309 = 61.8% 3) | 88= 4.2% | 25 = 1.2% | 8 = 0.4% | 740 = 34.9% |
| JOURNALISM:<br>Schnabel, Anne Frank | 3233 | 2422 = 75.5% | 67= 2.1% | 92 = 2.8% | 11 = 0.3% | 659 = 20.4% |
| PHILOSOPHY:<br>Snell, Entdeckung<br>des Geistes | 2039 | 1104 = 54.2% 4) | 61= 3.0% | 94 = 4.6% | 30 = 1,5% | 760 = 37.3% |
| PROSE DRAMA:<br>Hauptmann, Vor<br>Sonnenuntergang | 2243 | 1680 = 74.9% 5) | 62= 2.8% | 109 = 4.9% | 8 = 0.4% | 417 = 18.6% |
| NOVEL:<br>Mann, Königliche<br>Hoheit | 4531 | 3361 = 74.2% | 149= 3.3% | 114 = 2.5% | 11 = 0.2% | 923 = 20.4% |
| Mann, Doktor Faustus | 2028 | 1230 = 60.6% | 61= 3.0% | 61 = 3.0% | 9 = 0.4% | 670 = 33.0% |
|  | 22,489 | 14,773 = 65.7% | 774= 3.4% | 694 = 3.0% | 130 = 0.6% | 6,441 = 28.6% |

Sample figures for Predicate Noun in preverbal position: 1)15 = 0.7%; 2)25 = 1.2%; 3)30 = 1.4%; 4)12 = 0.6%; 5)4 = 0.2%.

Planning for next three months:

Work on the analysis of German will be carried on, with concentration on those elements that have not been thoroughly analyzed:

a) morphological classes of nouns;

b) types of adverbial constructions;

c) reference to elements prior in the sentence and in earlier sentence, with study of the interrelationships of sentences;

d) verbal phrases and their constructions;

e) types of sentences;

f) semological analysis.

Conventional programming will be undertaken, in accordance with a tentative plan to produce a translation in four passes through the machine:

pass 1 will enter text and perform dictionary look-up;

pass 2 will carry out form-recognition;

pass 3 will transfer German to English categories;

pass 4 will perform English look-up.

Identification of personnel:

Linguists: Dr. Werner Winter
(10 hrs per week, May 1 to June 15)
(30 hrs per week, June 16 to July 31)

Assistants: Mr. L. Frye
(30 hrs per week, June 16 to July 31)
Mrs. J. Frye
(20 hrs per week, June 16 to July 31)
Mrs. N. Orme-Johnson
(40 hrs per week, June 1 to July 31)
Mr. P. Pollard
(20 hrs per week, June 1 to July 31)

Dr. S.N. Werbow
(10 hrs per week, May 1 to June 15)
(30 hrs per week, June 16 to July 31)

Assistants: Mr. K. Johanson
(20 hrs per week, June 4 to July 31)
Mr. J. Simons
(20 hrs per week, July 1 to July 31)
Mrs. L. Thomas
(20 hrs per week, June 22 to July 31)

Dr. J. L. Parker
(30 hrs per week, June 4 to July 31)

Assistants: Mr. M. E. Gottschalk
(20 hrs per week, June 22 to July 31)
Miss H.H. Jeddeloh
(20 hrs per week, June 8 to July 31)

Programmers:

Mr. Eugene Pendergraft
(40 hrs per week, July 1 to July 31)
Mr. Raymond Faulk
(40 hrs per week, June 1 to July 31)
Mr. Richard Robinson
(40 hrs per week, June 22 to July 31)

| | |
|---|---|
| Secretary: | Mrs. Elfriede Sessions<br>(40 hrs per week, June 1 to July 31) |
| Chief Investigator: | Dr. W. P. Lehmann<br>(10 hrs per week, May 1 to June 15)<br>(40 hrs per week, June 16 to July 31) |

Brief description of background information of key personnel:

Dr. Werner Winter is Associate Professor of Germanic Languages at the University of Texas. He is in charge of the University's program of Russian teaching. He has published extensively in Indo-European linguistics.

Dr. Stanley N. Werbow is Assistant Professor of Germanic Languages at the University of Texas. He has been in charge of the University's elementary German program. He has published extensively in Germanic linguistics.

Dr. John L. Parker is Associate Professor of Germanic Languages at Texas Christian University, and is in charge of the German program there. He has had extensive experience with a wide variety of languages, including Hungarian.

Mr. Eugene Pendergraft is Computer Programmer III at the University of Texas. He has worked with various groups, including the Sage project at the Massachusetts Institute of Technology.

Dr. W. P. Lehmann is Professor of Germanic Languages at the University of Texas, and Chairman of the Department of Germanic Languages. He has published widely particularly in Indo-European, but also in general linguistics and machine translation.

The University of Texas, Austin, Texas

MACHINE LANGUAGE TRANSLATION STUDY - W. P. Lehmann,
Chief Investigator

First Quarterly Progress Report, 1 May to 31 July 1959. 93 pp.
Signal Corps Contract DA36-039 SC-38911 File No 18678-PM-59-91-91
(6909), Unclassified Report.

The report summarises previous work in machine translation at the University of Texas in reference to work performed by other groups. It describes in detail the procedures which are necessary, particularly in linguistic analysis, and it reviews the techniques and findings of structural linguistics which are of interest to machine translation. Work by the University of Texas group is based on complete linguistic analysis of source and target languages at various levels, with an indication of areas that have been studied and others that still require considerable analysis. We hold that the linguistic analysis carried out for machine translation will have to be based on formal principles and will have to view language in much the same way as does structural linguistics. Two papers on specific problems illustrate in detail the procedures of the Texas group, one dealing with the classification of the German verb, the other with the determination of subjects.—The section of the report dealing with programming concerns itself primarily with surveying the work elsewhere, which we describe as conventional. Work done at the University of Texas on the use of machines for textual analysis is briefly summarized.

Unclassified

1. Machine Language Translation Study
2. Signal Corps Contract DA36-039-SC-78911

The University of Texas, Austin, Texas

MACHINE LANGUAGE TRANSLATION STUDY - W. P. Lehmann,
Chief Investigator

First Quarterly Progress Report, 1 May to 31 July 1959. 93 pp.
Signal Corps Contract DA36-039 SC-38911 File No 18678-PM-59-91-91
(6909), Unclassified Report.

The report summarizes previous work in machine translation at the University of Texas in reference to work performed by other groups. It describes in detail the procedures which are necessary, particularly in linguistic analysis, and it reviews the techniques and findings of structural linguistics which are of interest to machine translation. Work by the University of Texas group is based on complete linguistic analysis of source and target languages at various levels, with an indication of areas that have been studied and others that still require considerable analysis. We hold that the linguistic analysis carried out for machine translation will have to be based on formal principles and will have to view language in much the same way as does structural linguistics. Two papers on specific problems illustrate in detail the procedures of the Texas group, one dealing with the classification of the German verb, the other with the determination of subjects.—The section of the report dealing with programming concerns itself primarily with surveying the work elsewhere, which we describe as conventional. Work done at the University of Texas on the use of machines for textual analysis is briefly summarized.

Unclassified

1. Machine Language Translation Study
2. Signal Corps Contract DA36-039-SC-78911

The University of Texas, Austin, Texas

MACHINE LANGUAGE TRANSLATION STUDY - W. P. Lehmann,
Chief Investigator

First Quarterly Progress Report, 1 May to 31 July 1959. 93 pp.
Signal Corps Contract DA36-039 SC-38911 File No 18678-PM-59-91-91
(6909), Unclassified Report.

The report summarises previous work in machine translation at the University of Texas in reference to work performed by other groups. It describes in detail the procedures which are necessary, particularly in linguistic analysis, and it reviews the techniques and findings of structural linguistics which are of interest to machine translation. Work by the University of Texas group is based on complete linguistic analysis of source and target languages at various levels, with an indication of areas that have been studied and others that still require considerable analysis. We hold that the linguistic analysis carried out for machine translation will have to be based on formal principles and will have to view language in much the same way as does structural linguistics. Two papers on specific problems illustrate in detail the procedures of the Texas group, one dealing with the classification of the German verb, the other with the determination of subjects.—The section of the report dealing with programming concerns itself primarily with surveying the work elsewhere, which we describe as conventional. Work done at the University of Texas on the use of machines for textual analysis is briefly summarized.

Unclassified

1. Machine Language Translation Study
2. Signal Corps Contract DA36-039-SC-78911

The University of Texas, Austin, Texas

MACHINE LANGUAGE TRANSLATION STUDY - W. P. Lehmann,
Chief Investigator

First Quarterly Progress Report, 1 May to 31 July 1959. 93 pp.
Signal Corps Contract DA36-039 SC-38911 File No 18678-PM-59-91-91
(6909), Unclassified Report.

The report summarises previous work in machine translation at the University of Texas in reference to work performed by other groups. It describes in detail the procedures which are necessary, particularly in linguistic analysis, and it reviews the techniques and findings of structural linguistics which are of interest to machine translation. Work by the University of Texas group is based on complete linguistic analysis of source and target languages at various levels, with an indication of areas that have been studied and others that still require considerable analysis. We hold that the linguistic analysis carried out for machine translation will have to be based on formal principles and will have to view language in much the same way as does structural linguistics. Two papers on specific problems illustrate in detail the procedures of the Texas group, one dealing with the classification of the German verb, the other with the determination of subjects.—The section of the report dealing with programming concerns itself primarily with surveying the work elsewhere, which we describe as conventional. Work done at the University of Texas on the use of machines for textual analysis is briefly summarized.

Unclassified

1. Machine Language Translation Study
2. Signal Corps Contract DA36-039-SC-78911

The University of Texas, Austin, Texas

MACHINE LANGUAGE TRANSLATION STUDY - W. P. Lehmann,
Chief Investigator

First Quarterly Progress Report, 1 May to 31 July 1959. 93 pp.
Signal Corps Contract DA36-039 SC-38911 File No 18678-PM-59-91-91
(6909), Unclassified Report.

The report summarises previous work in machine translation at the University of Texas in reference to work performed by other groups. It describes in detail the procedures which are necessary, particularly in linguistic analysis, and it reviews the techniques and findings of structural linguistics which are of interest to machine translation. Work by the University of Texas group is based on complete linguistic analysis of source and target languages at various levels, with an indication of areas that have been studied and others that still require considerable analysis. We hold that the linguistic analysis carried out for machine translation will have to be based on formal principles and will have to view language in much the same way as does structural linguistics. Two papers on specific problems illustrate in detail the procedures of the Texas group, one dealing with the classification of the German verb, the other with the determination of subjects.—The section of the report dealing with programming concerns itself primarily with surveying the work elsewhere, which we describe as conventional. Work done at the University of Texas on the use of machines for textual analysis is briefly summarized.

Unclassified

1. Machine Language Translation Study
2. Signal Corps Contract DA36-039-SC-78911

The University of Texas, Austin, Texas

MACHINE LANGUAGE TRANSLATION STUDY - W. P. Lehmann,
Chief Investigator

First Quarterly Progress Report, 1 May to 31 July 1959. 93 pp.
Signal Corps Contract DA36-039 SC-38911 File No 18678-PM-59-91-91
(6909), Unclassified Report.

The report summarizes previous work in machine translation at the University of Texas in reference to work performed by other groups. It describes in detail the procedures which are necessary, particularly in linguistic analysis, and it reviews the techniques and findings of structural linguistics which are of interest to machine translation. Work by the University of Texas group is based on complete linguistic analysis of source and target languages at various levels, with an indication of areas that have been studied and others that still require considerable analysis. We hold that the linguistic analysis carried out for machine translation will have to be based on formal principles and will have to view language in much the same way as does structural linguistics. Two papers on specific problems illustrate in detail the procedures of the Texas group, one dealing with the classification of the German verb, the other with the determination of subjects.—The section of the report dealing with programming concerns itself primarily with surveying the work elsewhere, which we describe as conventional. Work done at the University of Texas on the use of machines for textual analysis is briefly summarized.

Unclassified

1. Machine Language Translation Study
2. Signal Corps Contract DA36-039-SC-78911

The University of Texas, Austin, Texas

MACHINE LANGUAGE TRANSLATION STUDY - W. P. Lehmann,
Chief Investigator

First Quarterly Progress Report, 1 May to 31 July 1959. 93 pp.
Signal Corps Contract DA36-039 SC-38911 File No 18678-PM-59-91-91
(6909), Unclassified Report.

The report summarises previous work in machine translation at the University of Texas in reference to work performed by other groups. It describes in detail the procedures which are necessary, particularly in linguistic analysis, and it reviews the techniques and findings of structural linguistics which are of interest to machine translation. Work by the University of Texas group is based on complete linguistic analysis of source and target languages at various levels, with an indication of areas that have been studied and others that still require considerable analysis. We hold that the linguistic analysis carried out for machine translation will have to be based on formal principles and will have to view language in much the same way as does structural linguistics. Two papers on specific problems illustrate in detail the procedures of the Texas group, one dealing with the classification of the German verb, the other with the determination of subjects.—The section of the report dealing with programming concerns itself primarily with surveying the work elsewhere, which we describe as conventional. Work done at the University of Texas on the use of machines for textual analysis is briefly summarized.

Unclassified

1. Machine Language Translation Study
2. Signal Corps Contract DA36-039-SC-78911

The University of Texas, Austin, Texas

MACHINE LANGUAGE TRANSLATION STUDY - W. P. Lehmann,
Chief Investigator

First Quarterly Progress Report, 1 May to 31 July 1959. 93 pp.
Signal Corps Contract DA36-039 SC-38911 File No 18678-PM-59-91-91
(6909), Unclassified Report.

The report summarizes previous work in machine translation at the University of Texas in reference to work performed by other groups. It describes in detail the procedures which are necessary, particularly in linguistic analysis, and it reviews the techniques and findings of structural linguistics which are of interest to machine translation. Work by the University of Texas group is based on complete linguistic analysis of source and target languages at various levels, with an indication of areas that have been studied and others that still require considerable analysis. We hold that the linguistic analysis carried out for machine translation will have to be based on formal principles and will have to view language in much the same way as does structural linguistics. Two papers on specific problems illustrate in detail the procedures of the Texas group, one dealing with the classification of the German verb, the other with the determination of subjects.—The section of the report dealing with programming concerns itself primarily with surveying the work elsewhere, which we describe as conventional. Work done at the University of Texas on the use of machines for textual analysis is briefly summarized.

Unclassified

1. Machine Language Translation Study
2. Signal Corps Contract DA36-039-SC-78911