[From: *Modern Language Forum*, vol. 36 (3-4), 1951, pp.1-24.]

PROPOSALS FOR THE MECHANICAL RESOLUTION OF GERMAN SYNTAX PATTERNS•

Victor A. Oswald, Jr., and Stuart L. Fletcher. Jr

ABSTRACT

Recent research has shown that a reinvestigation of word frequency must be undertaken before the attempt can be made to solve the lexicographical problems of mechanical translation. Meanwhile those interested in mechanical translation have encountered the objection that, even it lexicographical problems could be solved, syntactical complications would probably constitute an insurmountable barrier to success. This present study indicates that syntactical problems can be solved by using a numerical code to identify syntactical functions and by employing mechanical routines to resolve foreign syntax patterns into English syntax patterns.

I

The operations outlined in this report are intended as a con-tribution towards solving the broader problem of the mechanical translation of foreign-language texts into English. To be sure, the lexicographical difficulties involved in mechanical translation are formidable, but at least their nature is understood and the solution of them will depend largely upon the size and range of the vocabulary that can be made automatically available[1]. On the other hand, no attempt has previously been made to provide a rigidly mechanical solution of syntactical problems[2], although it seems obvious enough that mechanical translation will be impossible until some system has been devised that will interpret, in terms of English patterns, the relations to one another of words or groups of words in foreign language texts. These proposals are intended to provide the elements of such a system[3].

German was chosen as the language for investigation, chiefly because any system capable of resolving the syntactical complexities of German could readily be applied to languages of less syntactical complexity, such as French or Spanish, and could readily be extended to embrace languages of greater syntactical complexity, such as Russian. The practical advantage of our proposals should require no exposition, since the mass of undigested material in foreign languages, particularly in Russian, is at present, and will be for some time to come, greater than the available force of translators can cope with. We should,

---

• This exposition was prepared originally as a report of investigations conducted under the auspices of the Institute for Numerical Analysis of the National Bureau of Standards. It has been officially released for publication.

The authors are indebted to Dr Harry D. Huskey, Assistant Director of the Institute for Numerical Analysis, whose interest in the possibilities of mechanical translation inspired this project and whose guidance was invaluable at every stage of our investigation. We also owe a debt of gratitude to our University of California colleagues Dr. Harry Hoijer, Professor of Anthropology, and Dr. William E Bull. Assistant Professor of Spanish, for advice and criticism

[1] Cf. especially E. U. Condon, "Statistics of Vocabulary." *Science*, Vol. 67 (1928) pp. 300 f.; George K. Zipf, *Human Behavior and the Principle of Least Effort* (Cambridge. Mass.: Addison-Wesley Press. 1949); and William E. Bull "Natural Frequency and Word Counts," *Classical Journal*. Vol. 64, No 8 (May, 1949) pp. 469 ff.

[2] For an interesting quasi-mechanical treatment of syntactic problems cf. C,V Pollard, A Key to Rapid Translation of German, University of Texas, 1947.

[3] The words "proposals" and "elements" are meant to constitute a caveat to the reader. It may be well, however, to quote verbatim the remarks of a critic of the official report: "It should be made clear, preferably very early in the presentation and in a conspicuous way, that the rules of translation proposed here are not intended as a complete system but merely as examples of the kind of rules which would make up such a system. While the present rules are adequate in, perhaps, 80 percent (?) of all cases, they could be enlarged until they form an almost infallible system. It is easy to give examples in which the rules fail."

however, wish to make it perfectly clear that mechanical translation as envisaged at present is not intended to provide an elegant—nor even a stylistically satisfactory—English text. What a mechanical translation might do is to furnish, at a speed vastly greater than would otherwise be possible, an intelligible approximation in English of the content of the original foreign-language text, and thus to facilitate the digestion of masses of printed material, particularly in diverse scientific fields.

II

Our procedures were purely pragmatic. We began by re-examining the processes by which the reader analyzes the syntactical relationship to one another of words and word-blocks. Then, after working out a code for the identification of syntactical relationships, we proceeded to reduce the analytical processes to mechanical routines. Some of our routines actually reproduce processes of grammatical analysis; others substitute processes which use indicia to which the reader pays little attention (e.g., capitalization of the first word of a sentence), but which can constitute marks of recognition for a machine.

We found that the elements of the language in question and their functional relationship to each other could be treated most efficiently in terms of traditional descriptive grammar. In fact, we found that much of our investigation led to the rediscovery of things familiar but partially forgotten, which is indeed fortunate, for this means that some form of coding system can be applied without much process of refinement to any language whose syntax has been adequately analyzed.

III

We should like to introduce a few definitions and explanations.

The element of discourse with which we are concerned is the printed word. Every printed word regarded as a vehicle of communication can potentially perform the function of communication on two levels: on the level of its own specific meaning (lexicography) and on the level of its relation to other words or word-blocks (syntax). Our primary concern has been the syntactical function of words and word-blocks. Words that can be syntactically isolated are called in the context of this report *syntactic units*; word-blocks that can be syntactically isolated are called in the context of this report *syntactic blocks*.

In actuality, the languages with which we are likely to be concerned employ two devices to indicate the relation of syntatic units and syntactic blocks to one another: inflectional endings and sequence (word order). English, happily enough, depends almost exclusively on word order, so that it proved possible to set up and prescribe for English an optimum sequence of syntactic units and blocks. German, on the other hand, although it prescribes for written discourse certain inflexible patterns of word order, operates with a complex of inflection and sequence. Our problem was first to isolate the syntactic units and syntactic blocks of German and then to devise a system by which they could be rearranged in a sequence which would produce intelligibility in English.

IV

To put our operations within a fixed frame of reference we designed them for an automatic computer of the type of the National Bureau of Standards Western Automatic Computer (SWAC), and we tried whenever possible to devise routines which would be convertible into a series of "instructions" "flow diagrams) that could be interpreted and executed by such an automatic computer.

Actually, our operation assumes an automatic setup not entirely yet in being: a high-speed automatically-sequenced electronic digital computer, to which is attached one or more intermediate-speed auxiliary memories in the form of magnetic drums. (Although the computer proper of the SWAC has been completed, a magnetic drum has not yet been integrated into the SWAC computer system.) This combination will make possible the

following sort of operation:

1. The foreign-language text, in this case German, is thought of as being supplied to the high-speed computer either on tape by an especially adapted electric typewriter (this is at present feasible but would drastically limit the speed of operation), or through the use of one of the devices now in preparation that will make possible direct line-by-line scanning of an entire printed page.

2. Through a series of "instructions" each German word is then to be compared with the German words stored in the intermediate-speed memory. The intermediate-speed memory is intended to be used both as a repository for coded elements and. of course; as a "dictionary."

3. Each German word on the drum is to have attached to it a code number which will serve to indicate its function. The coded words that correspond to the words appearing in the text being translated, together with their English lexicographical equivalents, are to be transferred from the intermediate-speed memory to the high-speed memory of the computer, where they will be temporarily stored.

4. This process is to be continued until the input introduces any one of diverse predetermined stop signals (cf. below); whereupon the high-speed computer will rearrange the accumulated material into English patterns and will transfer the English context to an output unit, presumably an automatic typewriter.

V
BOUNDARY MARKS

It is requisite to set up a system of boundary marks, at any one of which the machine is to cease scanning and is to rearrange into English word order the material it has been storing. Whenever possible, we have tried to make the mechanical boundaries coincide with the boundaries of the grammatical clause (independent clause, dependent clause, infinitive clause).

1. Any period, semicolon, colon, question mark, or exclamation point is the terminal boundary of a grammatical clause.

2.1 The comma can be recognized as the initial boundary mark of a subordinate clause whenever it occurs immediately before a relative pronoun, subordinating conjunction, interrogative adverb, or interrogative pronoun; or before a preposition followed by any one of the aforementioned elements.

2.2 The comma can be recognized as the terminal boundary mark of a subordinate clause when it precedes either *so* or *dann*, or whenever it occurs after a finite verb in a clause introduced by a subordinating element.

2.3 The comma can be recognized as the terminal boundary mark of any infinitive clause when it occurs immediately after an infinitive preceded by *zu* or an infinitive with *zu* infixed.

3. Except in the environments outlined in 2.1, 2.2. and 2.3 the comma serves only as a mark of coordination, i.e., marks the linkage of like functional units, or the linkage of like partial clauses, or the linkage of complete independent clauses (cf. p. 18).

4. Any group of words isolated within parentheses, brackets, or dashes is to be treated as a special (partial) clause, the elements of which are to be arranged like those of an independent clause, unless they actually constitute a complete subordinate clause or infinitive clause.

VI
CODING

The table of coded syntactic units is to be found in Appendix I. The Arabic figure .which appears at the left of each listing is to be understood as the code number assigned to the respective unit or units. The combinatory distribution of syntactic units is elaborated in sections VIII to XIV inclusive.

It should be pointed out that because certain syntactic units have two or more functions, it was necessary to find some way in which these diverse functions could be differentiated. We have relied upon environmental indicia to solve this problem.

For example, a glance at the combinatory chart for nominal blocks (p. 8) will readily reveal the fact that in the case of the definite article there are many more functions than forms; *der*, as definite article, can mark either the masculine singular noun, the genitive or dative feminine noun, or the genitive of a plural noun. These functions can be differentiated only by exarnining the environment in which the form *der* occurs, i.e., by establishing the nature of the nominal form with which it is combined. When it is combined with a feminine noun we have the further problem of determining whether the function is genitive or dative. We have solved this by requiring that *der* be considered to mark the genitive whenever it follows any nominal unit or nominal block and that it be treated as a marker of the dative in any other environment—a routine which, thought quite arbitrary, resulted in accurate rearrangement in an overwhelming majority of the occurrences we examined.

All other problems raised by functional overlapping are solved by the same procedure or by some similar procedure.

VII

ENGLISH SEQUENCE OF SYNTACTIC UNITS AND BLOCKS

Syntactic units and blocks are to be sought for and arranged in the fixed sequences indicated in the tables below. If any one of the units predicated is missing, the operation is to proceed to the next predicated unit. When more than one element is predicated for the same position in any sequence, it is to be understood that these elements are mutually exclusive.

1. When the first grammatical clause encountered is an *independent clause*:

1	2	3	4	5	6	7	8	9
Prepositional Phrase (occurring before any nominal or pronominal unit)	Subject	Verb	Indirect Object (pronominal)	Absolute Adverb	Direct Object	Indirect Object (nominal)	Adverb Predicate	Unattached Prepositional Phrase

2. When the first grammatical clause encountered is an *infinitive clause*:

1	2	3	4	5	6	7	8
Infinitive Predicator	zu plus Infinitive	Indirect Object (pronominal)	Absolute Adverb	Direct Object Predicate Nominative	Indirect Object (nominal)	Adverb Predicate Adjective	Unattached Prepositional Phrase

3. When the first grammatical clause encountered is a *dependent clause:*

1	2	3	4	5	6	7	8	9	10	11
Preposition	Relative Pronoun	Prepositional Phrase (occurring before any nominal or pronominal unit)	Subject	Verb	Indirect Object (pro-nominal)	Absolute Adverb	Direct Object Predi-cate Nomi-native	Indirect Object (nominal)	Adverb Predicate	Unattached Prepos-itional Phrase

The clausal types are to be identified by the following routine:

1. If a German finite verb form (code numbers 39 through 59, incl.) occurs either immediately before the subject or immediately after it, the clause can be identified as an *independent clause*.

2. If the first verbal form encountered in the German text is a prepositional infinitive (code numbers 62 or 73-40, 47, 49,53, 57), the clause can be identified as an *infinitive clause*.

3. If the appearance of the finite verbal form is postponed until after the occurrence of one or more of the following syntactic units or blocks: pronominal indirect object, absolute adverb, direct object or .predicate nominative, nominal indirect object, predicate adjective, unattached prepositional phrase, then the clause can be identified as a *dependent clause*.

Note that when the first elements encountered are part of an independent clause and a dependent clause or infinitive clause intervenes before the subject and the verb have both been located, the intervening clause must be passed over in the scanning until the elements of the independent clause have been located.

VIII
MECHANICAL IDENTIFICATION OF NOMINAL BLOCKS

Any syntactic unit can be automatically identified by its code number. Blocks which, like the verb, consist of two or more elements, but rarely more than three, can be identified without prescribing the possible combinations. The function of the noun, however, is so diverse and the number of elements that may be attached to a noun is so unpredictable, that nominal blocks must be identified in terms of possible combinations of first and last elements. The combinations, though complex, can be prescribed as follows:

Combinatory Chart for Nominal Blocks

	Subject Function	Possessive Function	Indirect Object Function	Direct Object Function
	Nominative	Genitive	Dative	Accusative
1 +	…20	…21; …23	…21	
2 +	…22	…24		
3 +			…20; …22	
4 +			…23	…20
5 +	…21; …23			…21; …23
6 +	…22; …20			…22
7 +		…24	…23	…20
8 +	…21; …23			…21; …23
10+	…20	…21; …23	…21	
11+	…22	…24		…22
12+			…20; …22	
13+		…24	…23	…20
14+	…21; …23			…21; …23
15+	…20	…21; …23	…21	
16+	…22	…24		…22
17+			…20; …22	
18+		…24	…23	…20
19+	…21; …23			…21; …23

In terms of this chart, a nominal block is said to be any predicated sequence of units, the first of which is a descriptive adjective or an article or a *dieser*-word or an *ein*-word or an adjectival participle (code numbers 1-19), and the last of which is a noun (code numbers 20-23), as, for example :

8 23
verschiedene Mengen
5 21
die Menge
5 21
diese Menge
5 21

eine Menge
19 23
ausgewählte Mengen

Predicable elements that can occur between the first and last element of a noun block are: numerals, descriptive adjectives, and adverbs which modify the adjectives occuring in the block.

5 66 7 18 23
die vier verschiedenen ausgewählten Mengen
5 77 19 21
die so gebildete Menge

Prepositional phrases are nominal blocks of the types predicated under the rubricks genitive, dative, accusative, preceded by a preposition (code numbers 70, 73, 74).

70 1 7 23
innerhalb der verschiedenen Mengen
70 4 18 23
von den ausgewählten Mengen

Any prepositional phrase or any genitive construction which immediately follows a noun or a nominal block is to be attached to the noun or nominal block.

5 23 70 23
einige Beispiele von Mengen
5 19 21 1 23
eine bestimmte Anzahl konkreter Gegenstände

The most complicated variety of nominal block is the so called participial construction, with which we include nominal blocks containing adjectives with extended adverbial modification. These blocks require special treatment. They can usually be identified by the intervention of a prepositional phrase, an adverb, or an object block between the initial element and the element immediately preceding the noun. Such constructions must be rearranged in the following sequence:

1	2	3	4	5	6	7	8	9
Preposition	Article etc.	Noun (plus immediately following genitive/ prepositional phrase)	Participal or Adjective immediately preceding the noun	Indirect Object (pronominal)	Direct Object Predicate Nominative	Indirect Object (nominal)	Adverb Predicate Adjective	Unattached Prepositional Phrase

70 3 70 28 18 20
aus einem vor uns stehenden Obstteller
70 3 20 18 70 28
aus einem Obstteller stehenden vor uns
1 70 80 18 23
der in ihr zusammengefassten Früchte
1 23 18 70 80
der Früchte zusammengefassten in ihr

IX

ROUTINES FOR THE IDENTIFICATION OF SUBJECT,
DIRECT OBJECT AND INDIRECT OBJECT.

1. Subject

There are certain units and blocks whose function is always that of the subject of the clause in which they occur. These are:

der (1)
any *dieser*-word in *-er* (1)
any *ein*-word in zero (6)
any adjectival form in *-er* (1,10,15)
ich. *wir*, *er*, *man* (25)
wer (32)

} masculine singular noun (20)

These we term absolute subjects.

There are certain units and blocks whose function is sometimes that of the subject of the clause in which they occur, sometimes that of the direct object. These are:

die (5)
any *dieser*-word in *-e* (5)
any *ein*-word in *-e* (5)
any adjectival form in *-e* (8,14.19)
} feminine singular noun (21) or plural noun (23)

das (6)
any *dieser*-word in *-es* (6)
any *ein*-word in zero (6)
any adjectival form in *-es* (6,11,16)
} neuter singular noun (22)

es (29)
sie (30)
was (36)
any noun (20-23)

These we term possible subject/direct-object blocks.

If a single possible subject/direct-object block occurs in a clause (other than an infinitive clause) which contains no absolute subject, this possible subject/direct-object block is to be dentified as the subject of that clause.

(……...Subject……....)
Die restlose Erforschung des Wassers in seinen festen
Zustandsformen ist noch nicht abgeschlossen.

If two possible subject/direct-object blocks occur in the same clause, the first-occurring possible subject/direct-object block is to be identified as the subject of that clause.

(Subject)
. . . dass die Menge eine Anzahl konkreter Gegenstände enthält . . .
(…….....Subject.. ....)
. . . dass die technischen Ausarbeitungen Früchte gezeitigt haben.

2. Direct Object

There are certain units and blocks whose function is always that of the direct object of any clause in which they occur. These are:

den (4)
any *dieser*-word in *-en* (4)
any *ein*-word in *-en* (4)
any adjectival form in *-en* (7,13,18)
} masculine singular noun (20)

mich (27), *ihn* (27)
wen (35)

These we term absolute direct object blocks.

There are certain units and blocks whose function is sometimes that of the direct object of the clause in which they occur, sometimes that of the subject. These are:

die (5)
any *dieser*-word in *-e* (5)
any *ein*-word in *-e*(5)
any adjectival form in *-e*
} feminine singular noun (21) or plural noun (23)

das (6)
any *dieser*-word in *-es* (6)
any *ein*-word in zero (6)
any adjectival form in *-es* (6.11,16)
} neuter singular noun (22)

es (29)
sie (30)
was (36)
any noun (20-23)

These we term possible subject/direct-object blocks (cf. above):

If a possible subject/direct-object block occurs in an infinitive clause, it is to be identified as the object of that clause.

(. .Object. .)
.... um das Problem zu lösen.

If any clause contains two possible subject/direct-object blocks, the possible subject/direct-object block rejected as the subject of that clause is to be identified as the object of that clause.

(Subject) (. . Object. .)
.... dass die Menge eine Anzahl konkreter Gegenstände enthält.

There are certain units which may function either as the direct object of the clause in which they occur, or as the indirect object. These are :

uns (28)
sich (31)

When there is no other unit or block identifiable as direct object, *uns* or *sich* is to be identified as direct object. When there is another unit or block identifiable as direct object, *uns* and *sich* are to be identifed as indirect object.

(Dir. Object)
Wir denken uns eine Anzahl konkreter Gegenstände. . .

When two possible direct objects occur in the same clause, the second-occurring possible direct object is to be identified as the direct object and the first-occurring unit or block as the indirect object.

(Ind. Object) (Dir. Object)
Ich habe Kriegskameraden öde Stunden verkürzen können. . .

3. Indirect Object

There are certain units and blocks which always have a dative function (usually indirect object) in any clause in which they occur. These are :

dem (3)
any *dieser*-word in *-em* (3)
any *ein*-word in *-em* (3)
any adjectival form in *-em* (3,12,17)
} any masculine singular noun (20), any neuter singular noun (22)

den (4)
any *dieser*-word in *-en* (4)
any *ein*-word in *-en* (4)
any adjectival form in *-en* (7,13.18)
} any plural noun (23)

mir, *ihm*, *ihnen* (26)
ihr (80) (when used pronominally, i.e., when not immediately preceding a masculine or neuter singular noun or the adjectival modifiers of a masculine singular or neuter singular noun)

These we term absolute datives.

There are units whose function is sometimes that of the direct object, sometimes that of the direct object in any clause in which they occur. These are :

uns (28)
sich (31)
(cf. above)

There are certain blocks whose function is sometimes dative, sometimes genitive in any clause in which they occur. They are:

der (1)
any *dieser*-word in *-er* (l)
any *ein*-word in *-er* (l)
any adjectival form in *-er* (1)
} any feminine singular noun (21)

These are termed possible genitive/dative blocks. Any possible genitive/dative block not previously identified as a genitive block is to be identified as a dative block (cf. pages 5, 6).

X

MECHANICAL IDENTIFICATION OF VERBAL BLOCKS

It is not necessary to prescribe the combinatory distribution of verbal units. Functionally the finite forms, the past participles. and the infinitives are mutually exclusive (except for the cases of homonymy discussed in Section XI), and the basic routine can be devised by instructions to seek out and rearrange the verbal units in a fixed sequence. A Chart of Coded Verbal Units is provided both to relieve the reader of laborious consultation of the Table of Coded Syntactic Units and to obviate recapitulation of the appropriate code numbers here and in Section XI.

Verbal units are to be sought for and rearranged in the following fixed sequence:

1. Finite verb forms.
2. Past participle.
3. Modal infinitive.
4. Any other infinitive.
5. Infinitive with *zu*.

If any of the elements listed above is missing in a given clause, the remaining elements are to retain their positions relative to each other.

Chart of Coded Verbal Units

	haben	sein	werden	modals	-ieren; weak verbs insep.pref.	sep. prefix	other
present singular	39	43	48	52	56	56	56
present plural	40	44	49	53	57	57	57
past singular	41	45	50	54	58	58	58
past plural	42	46	51	55	59	59	59
infinitive	40	47	49	53	57	57	57
infinitive with infix						62	
past participle	60	60	60	60	56	60	60

XI

ROUTINES FOR THE IDENTIFICATION OF VERBAL FUNCTIONS

The major problem of identifying verb functions is caused by the homonymy in all verbs of the infinitive and 1st and 3rd plurals, present tense. There is, however, also homonymy between the past participle of verbs whose infinitive ends in *-ieren* and the 3rd singular present tense forms of these verbs. There is a similar homonymy between the past participle of weak-verbs with inseparable prefixes and their 3rd singular present tense forms. Finally, a further problem is raised by the homonymy of the infinitive and the 1st and 3rd plurals of the present tense of modals with the infinitive-participle form of these verbs in the so-called "double infinitive" construction.

We propose a solution of these problems along the following lines:

In any clause in which the only verb form present is an infinitive/plural form, this

form is to be identified as a plural:

(3rd pl.)
Die Elemente bilden eine Merge.

In any independent clause in which the only verb forms are two infinitive/plural forms, the first-occurring infinitive/plural form is to be identified as a plural, the second-occurring form is to be identified as an infinitive.

(3rd pl.) (inf.)
Wir können die Menge aus diesen Elementen bilden.

If two infinitive/plural forms occur in the same clause and a form of the verb *werden* also occurs, the first-occurring infinitive/plural form is to be identified as the dependent infinitive, the second-occurring infinitive/plural form is to be identified as the modal infinitive. The form of the verb *werden* is to be identified as the finite verb.

(finite) (dep. inf.) (modal inf.)
Wir werden die Menge aus diesen Elementen bilden können.

If two infinitive/plural forms occur in the same clause with some form of the verb *haben*, the first-occurring infinitive/plural form is to be identified as the dependent infinitive, the second-occurring infinitive/plural form as the past participle. The form of the verb *haben* is to be identified as the finite verb.

(finite) (dep. inf.) (past part.)
Wir haben die Menge aus diesen Elementen bilden können.

In a dependent clause in which the only verb forms occurring are two infinitive/plural forms, the first-occurring infinitive/plural form is to be identified as a dependent infinitive, the second-occurring infinitive/plural form is to be identified as the plural.

(inf.) (3rd pl.)
.... weil wir die Menge aus diesen Elementen bilden können. . .

If a present/past participle verb form is the only verb form occurring in a clause, it is to be identified as a present singular.

(3rd sing)
Cantor definiert folgendermassen den Begriff der Menge:

In any clause in which a present/past particple form occurs with any form of the verb *haben* or *sein*, the present/past participle form is to be identified as a past participle, the form of *haben* or *sein* is to be identified as the finite verb.

(past part.)
Cantor hat den Begriff der Menge folgenderaiassen definiert

(past part.)
Er ist seit zwei Wochen verreist.

XII

ADVERBS

The routine for adverbs is relatively simple. Adverbs of quantity (66) are to be attached to the word immediately following them. Adverb/adjective forms (65) and absolute adverbs (81) are to be attached to any adjective, article, or other adverb that they immediately precede. Otherwise they are to be placed in the position prescribed for them in the sequence of functional units.

(link to *fünf*)
. . . aus einem vor uns stehenden Obstteller etwa fünf Aepfel. . .

(link to *billigen*)
So verbirgt sich hier hinter dem scheinbar billigen Effekt eine feinere Absicht.

(move *gewissermassen* to pre-final position)
Wir können die Elemente gewissermassen in einen Sack geworfen. . . denken.

XIII

The procedures previously outlined make possible a rough syntactic interpretation of the principal "parts of speech": nouns, pronouns, verbs, adverbs, adjectives, conjunctions-, prepositions. In the application of these procedures, however, we not infrequently encountered problems of functional overlapping for which no provision has been made. We found it possible to solve every such problem by setting up routines which provided for the mechanical scanning of the environment in which the problematic element occurred. The following routines are intended to serve as a representative sampling.

1. Relative Pronouns. Except for special genitive and dative forms the relative pronoun is represented by forms identical with those of the definite article *der* and the interrogative *welcher*. When these forms occur as relative pronouns, they are always separated by a comma from the word immediately preceding them, or are preceded by a preposition which in turn is separated by a comma from the word immediately preceding it. Relative pronouns can be identified by scanning beyond every such pattern of occurrence to see whether a predicated nominal unit follows. If no predicated nominal unit follows, the form is to be identified as a relative pronoun.

5 5 23
.... welche die Elemente der Menge genannt werden.
(5-5-23 is not a predicated nominal sequence)

5 70 5 21
.... die wir uns in eine Reihe angeordnet denken.
(5. . . 70-5-21 is not a predicated nominal sequence)

2. Many of the unattached verbal prefix forms are identical with those of prepositions. When such a form (70, 73, 74) occurs immediately before any punctuation mark or before a coordinating conjunction, it is to be prefixed to the finite verb of the clause.

70
Der Eisbrecher führt besondere Manöver aus. . .
(ausführt)

73
Er drückte den Deckel zu.
(zudrückte)

74
Er arbeitete sein früheres Werk um.
(umarbeitete)

3. *um*, *ohne*, *anstatt*. These words function both as prepositions and as introductory elements to the infinitive clause. In their latter function, they can be identified by the occurence of *zu* plus infinitive, or of an infinitive with *zu* infix, in the clause of which they form the initial elements.

73 57
Um das Problem numerisch zu lösen. . .

62
ohne sich in die Gedankengänge der Mengenlehre hineinzufinden. . .

62
Anstatt die neue Methode einzuschliessen
(In all of these examples the appearance of 62 or of the combination 73-57 makes identification possible.)

4. *zu* functions as a preposition, as an adverb, and as the constant element of the prepositional infinitive. In every case it is to be attached to the element that immediately follows it. *Zu* plus infinitive is always to be rearranged as the initial unit of the clause in which it occurs.

73 65
Die Anzahl ist schon zu gross.
(73-65 identifies *zu* as adverb)
73 4 23
Zu solchen Pflanzen gehören Rübsen nnd Flachs.
(73-4-23 identifies zu as preposition)
73 57
Um das Problem numerisch zu lösen. . . .
(73-57 identifies *zu* as part of prepositional infinitive)

5. *als* has at least three diverse functions: subordinating conjunction, comparative conjunction, and to link appositives. In the first of these functions it can be identified by the fact that it will either be capitalized or will be preceded by a comma. The second function can be identified only by the occurence of an adjective of comparative degree (9-14) in the functional unit immediately preceding its occurrence. In its third function it can be identified because it will always link like elements or functional units.

75
.... als wir das Problem zum erstenmal betrachteten.
(comma identifies als as conjunction)
9 75
Die eine Summe ist grösser als die andere.
(9-75 identifies als as comparative conjunction)
75
Ich als Mathematiker finde es schwer. . .
(absence of conditions for subordinating or comparative conjunction makes possible identification as appositive link)

6. *so*. All functions of *so* are adverbial, and it should always be attached to the element that immediately follows it except when it occurs as the first element of an independent clause preceded by a dependent clause. In the latter case it is to be retained as the first unit of the rearranged independent clause

77
Die Anzabl der so gebildeten Mengen. . .
(Link to the participial adjective)
77 53 25 6 22 57
.... so können wir unser Ziel erreichen.
77 25 53 57 6 22
.... so wir können erreichen unser Ziel

7. *ihr* performs two functions: that of a dative feminine pronoun and that of the neuter and masculine singular of the possessive adjective. In its pronominal function it can be identified by the fact that it does not immediately precede a neuter singular or masculine singular noun or the modifiers of these nouns. In its adjectival function it must always precede a masculine singular or neuter singular noun, or its modifiers.

80 22
Ihr Problem ist schwieriger als man hatte erwarten können.
(80-22 identifies ihr as possessive adjective)
80 14 23
Wenn wir die Menge gebildet haben, konnen wir ihr weitere Elemente hinzufügen.
(Absence of conditions for identification of possessive adjectives indicates pronominal function)

XIV

COORDINATION : *und*, *oder*, *aber*, AND THE COMMA

The problem of coordination proved to be the most corn-plicated of those with which we were confronted, and the solutions we propose are to be regarded as tentative. We have

not attempted to work out this problem conclusively since the solution of it will differ from language to language (variance in systems of punctuation will be critical) and since there is no pressing need at the present time for a final solution of the problem in terms of German.

When one of these coordinating elements occurs, the first operation is to scan beyond it to see whether the next unit in sequence is of the same grammatical category as the one immediately preceding it. If it belongs to the same category, we may assume a linkage of like elements.

66
. . . aus einem vor uns stehenden Obstteller etwa 5
23 66 23 66 23
Aepfel, 2 Birnen, und 1 Aprikose
(The apposition of the sequences 66-23, 66-23, 66-23 indicates linkage of like elements)

If the elements linked are nominal, a check must be made to see whether a preceding nominal unit with attached genitive or prepositional phrase is linked to a following nominal unit having the same function.

8 21 21
Die Anekdote weist Zuge auf, die wir oft in der Volksdichtung finden: episodenhafte Kürze, Betonung
1 21 21 70 6 22
der Handlung, Beschränkung auf das Notwendige.
(The linkage indicated is 8-21+21-1-21 + 21-70-6-22)

If the elements linked are verbal, then the following code number combinations are the only ones which can be considered to indicate a linkage of like elements.

56 + coordinating element + 56
57 + coordinating element + 57
58 + coordinating element + 58
59 + coordinating element + 59
60 + coordinating element + 60
56 + coordinating clement + 60

56 56
Er sucht und findet eine algebraische Lösung.
57 57
Wir suchen und finden algebraische Lösungen.
58 58
Er suchte und fand algebraische Lösungen.
59 59
Wir suchten und fanden algebraische Lösungen.
60 60
Wir haben algebraische Lösungen gesucht und gefunden.
56 60
Er hat das Problem studiert und gelöst

No other combination is to be interpreted as a linkage of like verbal elements. When any other combination occurs, the coordinating elements can be identified as the clausal boundary.

42 59
Als wir das erste Problem gelöst hatten, entdeckten wir, dass ....
(42-59 cannot indicate linkage of like elements)

If no linkage of like units is indicated, we must proceed to scan for a complete clause of the same type as that preceding the coordinating element.

XV

Obviously, the list of elements that require special identification is incomplete, and the series of special routines would have to be extended considerably if completeness were to be sought. Nevertheless this system, for all its tentative nature, will produce a certain

intelligibility; how great a degree of intelligibility the reader can decide for himself by examining Appendixes II, III, and IV.

Appendix II reproduces a section (pp. 4-5) of Adolf Fraenkel's *Einleitung in die Mengenlehre* (New York: Dover Publications, 1946) with our code numbers attached to each syntactic unit.

Appendix III indicates how this text would be mechanically rearranged by the application of our system.

Appendix IV is a word-by-word transverbalization of the rearranged text. The reader is to be cautioned that this "translation" had to be performed with deliberate disregard for the lexicographical problems of mechanical translation. It is intended solely to illustrate the application of our system of resolving German syntax patterns into the patterns of English.[4]

XVI

Although we are fully aware that our proposals are in no sense definitive, they intimate that problems of syntax and grammar can be solved mechanically, and that, syntax therefore does not constitute, as had been thought by some, a barrier to mechanical translations. However, before an elaboration of these proposals could serve any useful purpose, it would be necessary to have much more exact information than is at present available about the lexicographical aspect of the problem of mechanical translation. We should suggest that the following three step must be taken to make mechanical translation feasible.

1. An analysis of the number of items of each part of speech required to sustain discourse. That is, we should need to know how many noun forms, how many pronoun forms, how many verb forms, how many adjectives, adverbs, conjunctions, prepositions, particles, etc. are encountered in any relatively long sample of discourse, and the ratio of each of these classes required to arrive at any desired degree of understanding of the context. This information, which will be critical in determining the optimum size of the machine vocabulary, is not available. All previous investigations of word frequency have simply counted lexicographical items.

2. When the optimum size of the overall machine vocabulary has been worked out, we should then wish to determine what lexicographical items would be needed to translate various kinds of specific discourse.

3. When both the optimum size and the specific content of the machine vocabularies have been determined, we should then, and only then, wish to return to the question of how best to solve mechanically the syntactical problems of any specific language. It would then also be appropriate to examine simultaneously the question of whether the lexicographical items could not be stored in the mechanical memory by some system which would take advantage of the fact that the languages with which we are likely to be concerned employ a more or less regular pattern of endings which are attached to the roots of words. The problem of storage by root and ending would impinge upon the problem of coding for syntactical interpretation.

---

[4] In the transverbalization of the rearranged text for our official report we made the mistake of assuming that some of the lexicographical problems were insoluble. For instance, *bei* in line twelve of the text has the meaning "despite," probably the least frequent English equivalent, and so we gave it one of its high-frequency equivalents, "along-with." *Etwa*, which has as its high frequency equivalents English "perchance" and "approximately", has in line nine the rather rare equivalent "let us say," and so we chose to use one of the equivalents of higher frequency. The result of this leaning-over-backward, however, was confusion and, ironically enough, openly expressed doubt of out competence as translators. We have perforce had to assume that we can transfer our competence in toto to a mechanical translator.

Appendix 1

TABLE OF CODED SYNTACTIC UNITS

1. *der*; *dieser*-words in -*er*; *ein*-words in -*er*; descriptive adjectives in -*er*
2. *des*; *dieser*-words in -*es*; *ein*-words in -*es*
3. *dem*: *dieser*-words in -*em*; *ein*-words in -*em*; descriptive adjectives in -*em*
4. *den*; *dieser* words in -*en*; *ein*-words in -*em*
5. *die*; *dieser*-words in -*e*; *ein*-words in -*e*
6. *das*; *ein*-words in zero: descriptive adjectives in -*es*
7. descriptive adjectives in -*en*
8. descriptive adjectives in -*e*
9. comparative of descriptive adjectives in zero
10. comparative of descriptive adjectives in -*er*
11. comparative of descriptive adjectives in -*es*
12. comparative of descriptive adjectives in -*em*
13. comparative of descriptive adjectives in -*en*
14. comparative of descriptive adjectives in-*e*
15. present/past participle in -*er*
16. present/past participle in -*es*
17. present/past participle in -*em*
18. present/past participle in -*en*
19. present/past participle in -*e*
20. masculine noun
21. feminine noun
22. neuter noun
23. plural noun
24. genitive singular of masculine and neuter nouns
25. *ich; wir; er; man*
26. *mir; ihm; ihnen*
27. *mich: ihn*
28. *uns*
29. *es*
30. *sie*
31. *sich*
32. *wer*
33. *wessen*
34. *wem*
35. *wen*
36. *was*
37. *deren; dessen*
38. *denen*
39. *haben*, present singular
40. *haben*, present plural; infinitive
41. *haben*, past singular
42. *haben*, past plural
43. *sein*, present singular
44. *sein*, present plural
45. *sein*, past singular
46. *sein*, past plural
47. *sein*, infinitive
48. *werden*, present singular
49. *werden*, present plural; infinitive
50. *werden*, past singular
5l. *werden*, past plural
52. modal verbs, present singular
53. modal verbs, present plural; infinitive
54. modal verbs, past singular
55. modal verbs, past plural
56. other verbs, present singular; past participle of verbs in -ieren; past participle of weak verbs with inseparable prefix
57. other verbs, present plural; infinitive
58. other verbs, past singular
59. other verbs, past plural
60. past participle in zero
61. present participle in zero
62. infinitive with -*zu*- infix
63. *worden*
64. non-prepositional prefix, unattached (to be attached to finite verb of clause in which it occurs)
65. adverb; descriptive adjective in zero
66. adverbs of quantity; numerals (to be attached to word immediately following)
67. *denn; sondern*
68. *und; oder; aber; auch*
69. subordinating conjunction
70. preposition
71. preposition compounded with *da*-
72. interrogative adverb
73. *zu*
74. *um; ohne: anstatt*
75. *als*
76. *wie*
77. *so; dann*
78. *nicht*
79. contraction of preposition with article
80. *ihr*
81. absolute adverbs
82. explanatory conjunction

Appendix II

	39	4	20	1	21	8l	56
CANTOR	hat	den	Begriff	der	Menge	folgendermassen	definiert:

5	21	43	5	21	1	1	23	1
Eine	Menge	ist	eine	Zusammenfassung	bestimmter	wohlunterschiedener	Objekte	unserer

21	68	2	24		5	5	23	1	21	60	49
Anschauung	oder	unseres	Denkens	—	welche	die	Elemente	die	Menge	genannt	werden

73 3 22
— zu einem Ganzen.
23 69 25 5 21 79 22 57 53 25 5
1. Beispiele. Bevor wir diese Definition im Einzelnen zergliedern, wollen wir einige
23 70 23 57 5 28 6 22 79 22 1
Beispiele von Mengen betrachten, die uns anschauliches Material zum Verständnis der
21 57 53
Definition liefern sollen.
25 57 28 5 8 21 1 23 82 70 3 70 28
1. Wir denken uns eine bestimmte Anzahl konkreter Gegenstände, z.B. aus einem vor uns
18 20 66 66 23 66 23 68 66 21 1 20 1 66 23
stehenden Obstteller etwa 5 Aepfel, 2 Birnen, und 1 Aprikose; der Inbegriff dieser 8 Dinge
56 5 21 64 5 23 1 77 18 21 44 5 7 23
stellt eine Menge dar. Die Elemente der so gebildeten Menge sind die einzelnen Früchte;
70 4 70 1 21 1 23 81 7 20 1
durch den bei aller Handgreiflichkeit dieser Elemente doch gedanklichen Akt ihrer
21 73 3 22 40 25 5 21 1 66 23 60 5 21
Zusammenfassung zu einem Ganzen haben wir die Menge der 8 Früchte gebildte. Die Menge
56 66 81 8 23 5 25 28 70 5 21 60
enthält 8 untereinander verschiedene Elemente, die wir uns in eine Reihe angeordnet
57 82 6 1 20 6 1 20 82 5 5 21 5 5
denken (z. B.: ein erster Apfel, ein zweiter Apfel, usw., die eine Birne, die andere
21 65 81 5 21 57 25 70 1 7 21 1
Birne, endlich zuletzt die Aprikose). Sehen wir von der besonderen Natur der
7 23 64 77 56 28 5 21 81 81 6 22 64 70 3
einzelnen Elemente ab, so stellt uns die Menge nur mehr ein Ordnungsschema dar mit dem
20 81 81 81 65 53 25 70 70 1 21 1
Inhalt: erstens, zweitens, ... achtens. Endlich können wir ausser von der Natur der
23 68 81 70 1 21 57 5 23 81 70 4
Elemente auch noch von ihrer Anordnung absehen, die Elemente gewissermassen in einen
20 60 68 60 57 77 56 28
Sack geworfen und durcheinandergeschüttelt denken; dann vermittelt uns
5 21 75 7 20 81 81 5 21 1 70 80 19
die Menge als einzigen Inhalt nur mehr die Anzahl der in ihr zusammengefassten
23 65 5 21 66
Früchte, nämlich die Anzahl 8.

Appendix III

Cantor hat definiert folgendennassen den Begriff der Menge:

Eine Menge ist eine Zusammenfassung bestimmter wohlunterschiedener Objekte unserer Anschauung oder unseres Denkens — welche werden genannt die Elemente der Menge — zu einem Ganzen.

1. Beispiele. Bevor wir zergliedern diese Definition im Einzelnen, wir wollen betrachten einige Beispiele von Mengen, die sollen liefern uns anschauliches Material zum Verständnis der Definition.

1. Wir denken uns eine bestimmte Anzahl konkreter Gegenstände, z. B. aus einem Obstteller stehenden vor uns etwa 5 Aepfel. 2 Birnen, and l Aprikose; der Inbegriff dieser 8 Dinge darstellt eine Menge. Die Elemente der Menge so gebildeten sind die einreinen Früchte; durch den Akt ihrer Zusammenfsasung zu einem Ganzen doch gedanklichen bei aller Handgreiflichkeit dieser Elemente wir haben gebildet die Menge der 8 Früchte. Die Menge enthält 8 untereinander verschiedene Elemente, die wir denken angeordnet uns in eine Reihe (z.B. ein erster Apfel, ein zweiter Apfel, usw., die eine Birne, die andere Birne, zuletzt die Aprikose endlich.) Wir absehen von der Natur der einzelnen Elemente so die Menge darstellt uns nur mehr ein Ordnungsschema mit dem Inhalt: erstens, zweitens, . . . achtens. Wir können absehen ausser von der Natur auch von ihrer Anordnung noch, denken geworfen die Elemente in einen Sack gewissermassen und durcheinandergeschüttelt; die Menge vermittelt uns dann nur mehr die Anzahl der Früchte zusammengefassten in ihr als einzigen Inhalt, nämlich die Anzahl 8.

Appendix IV

Cantor has defined as follows the concept of the set:

A set is a collection of definite well-distinguished[5] objects of-our perception or of-our thought — which will be called the elements of the set — to a whole.

1. Examples. Before we analyze this definition in detail, we want-to regard some examples of sets. which shall furnish us perceptible material for-the understanding of-the definition

1. We think to-ourselves a definite number of-concrete objects. for example out-of a fruit-plate standing before us let-us-say 5 apples, 2 pears. and l apricot; the sum of-these 8 things represents a set. The elements of the set so formed are the single fruits; through the act of-their collection to a whole still mental despite all palpability of-these elements we have formed the set of-the 8 fruits. The set contains 8 among-one-another different elements which we think ordered to-ourselves in a -series (for example. a first apple. a second apple, etc., the one pear, the other pear, last the apricot finally.) We take-no-account of the nature of the individual elements, so the set represents to-us only a scheme-of-order with the content: first. second, . . . eighth. We can take-no-account besides of the nature also of their order still, think tossed the elements into a sack as-it-were and shaken-about: the set conveys to-us then only the number of-the fruits comprised in it as sole content, namely the number 8.

Victor A. Oswald, Jr.,
and
Stuart L. Fletcher. Jr

University of California, Los Angeles,
and National Bureau of Standards

[5] Hyphenation indicates that two or more English words are required to represent one German word