Arbeitspapier Nr. 15

Februar 1971

PROBLEMS OF ENGLISH-GERMAN AUTOMATIC TRANSLATION

Paul Otto Samuelsdorff

Problems of English-German Automatic Translation*

1. Aims and method

The two aims of the project described in this paper are
1. the practical aim of giving a mechanical procedure
for finding an equivalent expression in German for any
English input sentence and 2. the theoretical aim of
finding the corresponding structures of the two languages.
With both of these aims in mind, the empirical method was
adopted.

2. Structure of the dictionary

Two dictionaries are needed: an idiomatic dictionary
and a lexical dictionary. Both have next to each English
entry the German equivalent (or equivalents) and the
syntactic and semantic information needed for producing
a correct German sentence. The lexical dictionary may,
for technical reasons, be subdivided into a dictionary
for words of high frequency and one for words of low
frequency.

3. Syntactic Problems

3.1. Morphological problems: choosing the correct German
suffixes.

3.2. Lexicological problems: choosing the correct German
article, pronoun, relative pronoun or preposition.

3.3. Word order: finding the correct structure of German
word-groups and sentences for the corresponding
English structure.

4. Semantic Problems

4.1. Solvable problems: the correct German equivalent
of an ambiguous English word may be found by
testing the environment.

4.2. Partially solvable problems: the correct German
equivalent can only be found by approximation.
Possible solutions will be suggested.

5. Examples

Examples will be given for all the above problems.

1. Aims and method

The aim of any Automatic Translation project is to give a
mechanical procedure for finding an equivalent expression
in the target language for any sentence in the source
language. The aim of any linguistic translation project
is to find the corresponding structures of the languages
dealt with. The two main problems that have to be solved
by such a project are the difference of word order between
the source language and the target language and the ambi-
guous words of the source language for which the appropri-
ate word in the target language has to be chosen. The
firs problem is of major linguistic interest: once the
project has been worked out, it will give us the parallell
sentence structures for the two languages in question.
Since there is no complett analysis of any language that
could be used for the purpose of automatic translation,
we decided to build up our project sentence by sentence.
The rules which are needed for translating each sentence
will have to be included in the complete program anyway,
and the translation may be checked and corrected immediate-
ly. The program is split up into subroutines for each
word-class, so that a correction of the program in case of
an unsatisfactory translation does not complicate the pro-
gram unnecessarily.

2. Structure of the dictionary

The most important part of an automatic translation pro-
ject is an efficient machine dictionary. For setting up
such a dictionary we have to know what information it has
to contain. Since this information is best obtained after
experience with automatic translation, we started the pro-
ject by assuming that the dictionary lookup had already
taken place. The English sentences to be translated were
read into the computer word by word with the German trans-
lation and the grammatical information necessary for ob-
taining a correct German translation of the sentence.

With the articles and relative pronouns only the grammatical information is given, but no translation is stored. With some of the other words that have several possibilities of translation only the one relevant to the sentence in question is stored, while with some ambiguous words the ambiguity is solved by the program. After the machine dictionary has been set up, all the ambiguities have to be solved by the program. Idioms are placed in a separate dictionary, since without such a dictionary a complicated routine would be needed for testing the environment of a word that may appear in an idiom in order to translate it correctly. Apart form the solution of ambiguities, the setting up of the dictionary is merely a technical problem. All the other problems are solved by the program that translated the simulated dictionary output.

The structure of the dictionary is partly dependent on the type of computer used. At present the grammatical information is stored in two machine-words of 32 bits each, called DECO and SECO, as shown in figure 1. When setting up the complete dictionary it will probably be advantageous to have a few hundeed most frequent words stored in the core storage and the rest of the dictionary on disks. Another problem is the size of the dictionary. The program recognizes as a word a string of letters between blanks or a string of letters between a blank and a punctuation mark. It would be inefficient to have all possible words in the dictionary. Thus the number of nouns in the dictionary may be halved by listing only the singular nouns and letting the program deal with words that end in 's' and don't appear in the dictionary. The same applies to verb endings. It will have to be kept in mind that the smaller the size of the dictionary the larger the program and vice versa.

D E C O	S E C O
1. definite article	1. after definite article
2. indefinite article	2. material
3. noun	3. after indefinite article
4. preposition	4. part of verb to end of clause
5. adjective	5. suffix ED / ANY
6. verb	6. suffix ING
7. relative pronoun	7. verb after auxiliary / adverb without change of word order
8. numeral	8. end of adverbial group
9. comma	9. TOO,AND
10.THIS,THESE	10.quantity
11.conjunction	11.OF
12.pronoun	12.TO
13.infinitive	13.FROM
14.adverb	14.ABOUT
15.verb with 2 objects(give)	15.verb after which 'to'not transl.
16.intransitive verb	16.verb followed by 'to'
17.modal or auxiliary verb	17.verb after relative pronoun
18.BE	18.OF-group aftee noun
19.dative plural EN	19.THAN
20.dative plural N	20.compound noun with N
21.	21. " " " S
22.genitive	22. " " " EN
23.dative	23. " " " ES
24.accusative masc.	24.
25.MAKE	25.
26.masc.	26. verbs after which nouns are not compounded (call)
27.fem.	27. THERE ARE
28.neuter	28.
29.plural	29.
30.genitive sing. ES	30.
31. " " EN	31.
32. " " N	32. not found in dictionary

FIGURE II

3. Syntactic problems

3.1. Since German is more inflectional than English, a
subroutine had to be written that adds the German suffixes
to adjectives or to the indefinite article according to
the gender, number and case of the folloing noun or to
the noun itself. Gender and number are found in the dict-
ionary output, while the case is determined by the posi-
tion of the noun in the sentnce (before the verb, after
the verb or after a preposition). Examples may be found
in every sentence. One grammatical item is stored for the
following sentence. The gender of the last noun in the
sentence is memorized, in case the following sentence
starts with 'this' not followed by a noun. We are
aware of the fact that this is not a hundred per cent
solution.(Sentences 11 and 12) If 'more' or 'most' are
followed by an adjective or adverb in English, the suffix
'er' or 'est' is added to the German adjective.(Sent.39)

3.2. The German definite article is chosen in a simalar
way as the suffix of the adjective and the indefinite
article, namely according to the gender, number and case
of the following noun. (Examples may be found in almost
every sentence.) The German article is inserted if an
English singular noun is not preceded by an article, un-
less this noun is classified as 'material'. (Sent. 3)
The relative pronoun is chosen according to gender and
number of the preceding noun. This choice is in most
cases correct, but is still unsatisfactory. (Sent.5,23,39)
The case of the pronoun is determined similarly to that
of the noun and the dictionary output is altered accord-
ingly.(Sent.32) If an English finite verb is translated
by more than one lexical item in German, the last of these
will go to the end of the clause(Sent.39). Two nouns fol-
lowing each other will be translated by a compound noun
in German, unless they follow a verb like 'call' which
permits such a constellation. The necessary German infix
will be inserted.(Sent.53)

3.3. German word order is relatively free and the English
word order will in many cases not be altered. But the
German text will become fairly unreadable if the syntactic
rules that exist are not adhered to. It is also one of
the aims of our project to obtain a comparative structure
of English and German. That is why our main effort from
the start has been to obtain the correct German word
order for each sentence. This concerns mainly the posi-
tion of the verb. When a sentence starts with an adverb
or an adverbial group, it has to be followed by the verb
(Sent.5,32,33). In relative clauses the verb stands at
the end of the clause (Sent.5,23,39) and in case there
is an auxiliary or nodal verb, it follows the main verb
(Sent. 5). In the main clause the infinitive stands at
the end of the clause (Sent. 32,33,53).

4. Semantic problems

4.1. One of the most difficult problems in automatic
translation is the translation of words with multiple
meaning. In some cases the solution may be found by test-
ing the environment. Thus the word 'too' will be trans-
lated by 'zu' if it precedes an adjective or an adverb
and by 'auch' in all other cases. It will then be in-
serted after the verb (Sent. 25). The correct transla-
tion is easily found when the word belongs to different
word-classes. If the different translations belong to
the same wordclass they may often be found by checking
the environment. A verb may, for instance, have several
meanings depending on the preposition which follows it.
In some cases the semantic information with the preceding
or following word may determine the translation of a noun
or a verb. Prepositions may in general be translated
correctly by checking the preceding verb and/or follow-
ing noun (Sent. 11)

4.2. In cases where there is no exact way of determin-
ing the translation, several approximate solutions are
possible. One solution is having separate dictionaries
for various kinds of texts to be translated. Thus in
a genaral dictionary the word 'bush' will be translated
by 'Busch', while in a dictionary for the translation
of technical texts it will be translated by 'Buechse'.
A second solution is the statistical solution. The most
frequent translation may be chosen as the dictionary
entry. In case this leads to misunderstandings, the
second alternative may be given in parantheses. The
correct gender for pronouns may sometimes be difficult
to find, because of their ambiguous relation to pre-
ceding nouns (Sent.25). Here the statistical solution
may be the only feasible one.

5. Examples

The following examples are ten out of sixty-four
sentences that have so far been translated. They are
taken from A.J.Herbert, The Structure of Technical
English, Longmans, Green and Co. Ltd., London, 1968.
The numbers behind the German translation refer to
the bits in DECO and SECO that are set to '1'. If
there is no hyphen, SECO contains only zeros, other-
wise the SECO entries may be recognized by the numbers
following the hyphen.

Input for Sentence 3

MODERN	MODERN	5
INDUSTRY	INDUSTRIE	3,27
NEEDS	BRAUCHT	6
CONSIDERABLE	BEDEUTEND	5
QUANTITIES	MENGEN	3,27,29
OF	VON	4,23 - 11
THIS	DIES	10
METAL	METALL	3,20,28 - 2
,	,	9
EITHER	ENTWEDER	11
IN	IN	4,23
THE		1
FORM	FORM	3,27
OF	VON	4,23 - 11
IRON	EISEN	3,28 - 2
OR	ODER	11
IN	IN	4,23
THE		1
FORM	FORM	3,27
OF	VON	4,23 - 11
STEEL	STAHL	3,28 - 2

Translation

DIE MODERNE INDUSTRIE BRAUCHT BEDEUTENDE MENGEN VON DIESEM
METALL , ENTWEDER IN DER FORM VON EISEN ODER IN DER FORM
VON STAHL..

Input for Sentence 5

MOREOVER	FERNER	14
,	,	9
IRON	EISEN	3,23 - 2
POSSESSES	BESITZT	6
MAGNETIC	MAGNETISCH	5
PROPERTIES	EIGENSCHAFTEN	3,27,29
WHICH		7
HAVE	HABEN	6,17
MADE	GEMACHT	6,25
THE		1
DEVELOPMENT	ENTWICKLUNG	3,27
OF	VON	4,23 - 11
ELECTRIC	ELEKTRISCH	5
POWER	ENERGIE	3,27 - 2
POSSIBLE	MOEGLICH	5
.	.	

Translation

FERNER BESITZT EISEN MAGNETISCHE EIGENSCHAFTEN, DIE DIE
ENTWICKLUNG VON ELEKTRISCHER ENERGIE MOEGLICH GEMACHT
HABEN .

Input for Sentence 11

THE		1
NON-METALLIC	NICHT-METALLISCH	5
CONSTITUENTS	BESTANDTEILE	3,29
OF	VON	4,23 - 11
THE		1
ORE	ERZ	3,28,30 - 2
COMBINE	VERBINDEN SICH	6
WITH	MIT	4,23
THE		1
LIMESTONE	KALK	3,26 - 2
TO	ZU	4,23 - 12
FORM	BILDEN	6
A	EIN	2
LIQUID	FLUESSIG	5
SLAG	SCHLACKE	3,27

Translation

DIE NICHT-METALLISCHEN BESTANDTEILE DES ERZES VERBINDEN
SICH MIT DEM KALK, UM EINE FLUESSIGE SCHLACKE ZU BILDEN.

Input for Sentence 12

THIS	DIES	10
FLOATS	SCHWIMMT	6
ON TOP OF	AUF	4,23
THE		1
MOLTEN	GESCHMOLZEN	5
IRON	EISEN	3,28 - 2
,	,	9
AND	UND	11 - 9
PASSES	KOMMT	6
OUT OF	AUS	4,23
THE		1
FURNACE	HOCHOFEN	3,26
THROUGH	DURCH	4,24
A	EIN	2
TAP	HAHN	3,26,30
.	.	

Translation

DIESE SCHWIMMT AUF DEM GESCHMOLZENEN EISEN UND KOMMT AUS
DEM HOCHOFEN DURCH EINEN HAHN .

Input for Sentence 23

IRON	EISEN	3,28
WHICH		7
CONTAINS	ENTHAELT	6
A	EIN	2
NEGLIGIBLE	GERINGFUEGIG	5
QUANTITY	MENGE	3,27 - 10
OF	VON	4,23 - 11
CARBON	KOHLENSTOFF	3,26 - 2
,	,	9
FOR EXAMPLE	ZUM BEISPIEL	14
WROUGHT-IRON	SCHMIEDEEISEN	3,28 - 2
,	,	9
BEHAVES	VERHAELT SICH	6
DIFFERENT FROM	ANDERS ALS	14 - 7
IRON	EISEN	3,28 - 2
WHICH		7
CONTAINS	ENTHAELT	6
A	EIN	2
LOT	MENGE	3,27 - 10
OF	VON	4,23 - 11
CARBON	KOHLENSTOFF	3,26 - 2
.	.	

Translation

EISEN, DAS EINE GERINGFUEGIGE MENGE KOHLENSTOFF ENTHAELT,
ZUM BEISPIEL SCHMIEDEEISEN, VERHAELT SICH ANDERS ALS EISEN,
DAS EINE MENGE KOHLENSTOFF ENTHAELT.

Input for Sentence 25

THIS	DIES	10
IS	IST	6,18
A	EIN	2
VERY	SEHR	14 — 7
HARD	HART	5
SUBSTANCE	SUBSTANZ	3,27
,	,	9
AND	UND	11 — 9
IT	ES	12
MAKES	MACHT	6,25
THE		1
IRON	EISEN	3,28 — 2
HARD	HART	5
TOO	AUCH	14 — 9

Translation

DIES IST EINE SEHR HARTE SUBSTANZ , UND ES MACHT AUCH
DAS EISEN HART .

Input for Sentence 32

SECONDLY	ZWEITENS	14
,	,	9
WE	WIR	12
CAN	KOENNEN	6,17
HEAT	ERHITZEN	6
THE		1
STEEL	STAHL	3,26 - 2
ABOVE	UEBER	4,24
A	EIN	2
CERTAIN	GEWISS	5
CRITICAL	KRITISCH	5
TEMPERATURE	TEMPERATUR	3,27
,	,	9
AND	UND	11 - 9
THEN	DANN	14
ALLOW	GESTATTEN	6,23 - 16
IT	ES	12
TO	ZU	4 - 12
COOL	KUEHL WERDEN	6
AT	BEI	4,23
DIFFERENT	VERSCHIEDEN	5
RATES	GESCHWINDIGKEITEN	3,26,29
.	.	

Translation

ZWEITENS KOENNEN WIR DEN STAHL UEBER EINE GEWISSE KRITISCHE
TEMPERATUR ERHITZEN , UND GESTATTEN IHM DANN, BEI VERSCHIE-
DENEN GESCHINDIGKEITEN KUEHL ZU WERDEN

Input for Sentence 33

AT	BEI	4,23
THIS	DIES	10
CRITICAL	KRITISCH	5
TEMPERATURE	TEMPERATUR	3,27
,	,	9
CHANGES	VERAENDERUNGEN	3,27,29
BEGIN	BEGINNEN	6,16
TO	ZU	4,23 - 12
TAKE PLACE	STATT FINDEN	6
IN	IN	4,23
THE	,	1
MOLECULAR	MOLEKULAR	5
STRUCTURE	STRUKTUR	3,27
OF	VON	4 - 11
THE		1
METAL	METALL	3,26
.	.	

Translation

BEI DIESER KRITISCHEN TEMPERATUR BEGINNEN VERAENDERUNGEN
IN DER MOLEKULAREN STRUKTUR DES METALLS STATT ZU FINDEN .

Input for Sentence 39

METAL	METALL	3,28 - 2
WHICH		7
WE	WIR	12
CAUSE	VERANLASSEN	6,16
TO	ZU	4,23,12
COOL	KUEHLEN	6,13
RAPIDLY	SCHNELL	14
CONTRACTS	ZIEHT SICH ZUSAMMEN	6 - 4,7
MORE	MEHR	14
RAPIDLY	SCHNELL	14
ON THE OUTSIDE	AUSSEN	14
THAN	ALS	11 - 19
ON THE INSIDE	INNEN	14

Translation

METALL, DAS WIR SCHNELL ZU KUEHLEN VERANLASSEN, ZIEHT SICH SCHNELLER AUSSEN ALS INNEN ZUSAMMEN .

Input for Sentence 53

THE		1
ROLL	WALZE	3,27 - 20
PRESSURES	DRUCK	3,26
MUST	MUSS	6,17
BE	SEIN	6,13,18
MUCH	VIEL	14 - 7
GREATER	GROESSER	5
FOR	FUER	4,24
COLD	KALT	5
ROLLING	WALZEN	3,28 - 2,6
THAN	ALS	14,7,19
FOR	FUER	4,24
HOT	HEISS	5
ROLLING	WALZEN	3,28 - 2,6
,	,	9
BUT	ABER	11
COLD	KALT	5
ROLLING	WALZEN	3,28 - 2,7
ENABLES	ERMOEGLICHT	6.23 - 16
THE		1
OPERATOR	OPERATOREN	3,29
TO	ZU	4,23 - 12
PRODUCE	PRODUZIEREN	6
ROLLS	WALZEN	3,27,29
OF	VON	4,23 - 11
GREAT	GROSS	5
ACCURACY	GENAUIGKEIT	3,27 - 2
AND	UND	11 - 9
UNIFORMITY	EINFOERMIGKEIT	3,27 - 2
,	,	9
AND	UND	11 - 9
WITH	MIT	4,23
A	EIN	2
BETTER	BESSER	5
SURFACE	OBERFLAECHE	3,27 - 20
FINISH	FINISCH	3,26

Translation

DER WALZENDRUCK MUSS VIEL GROESSER FUER KALTES WALZEN ALS FUER
HEISSES WALZEN SEIN, ABER KALTES WALZEN ERMOEGLICHT DEN OPE-
RATOREN WALZEN VON GROSSER GENAUIGKEIT UND EINFOERMIGKEIT ZU
PRODUZIEREN UND MIT EINEM BESSEREN OBERFLAECHENFINISCH .