Some semantic aspects of the machine translation from Russian into Hungarian

by

Gy. Sipóczy

1. The morphological analysis has to determine the grammatical role of certain morphemes (concerning tense, case, etc.) in a given sentence. In some cases, however, the grammatical role of the morpheme cannot be found out only with the help of analysis of the word form under consideration. In such cases the analysis must also bear on the context. We may find, for instance, a lot of homographs among the substantival morphemes. The question which of the possible cases the given morpheme really represents in the analyzed sentence may usually be answered by the aid of the analysis of the foregoing adjectives and prepositions which are in agreement with the given word. E.g. точки = singular genitive, plural nominative and plural accusative; but на эти точки = plural accusative. Seeing that in such cases the analysis of the prepositions and the substantive was necessitated by the uncertainty of the given substantival morpheme, we may - with good reason - consider the analysis of the whole syntagma (nominal group) as a part of the morphological analysis.

If the role of a homographic morpheme cannot be cleared up by the aid of the analysis of other morphemes belonging to the same syntagma (either because of the presence of another homograph word form or because of the lack of further morphemes in the syntagma at all), an additional - syntactic or semantic - analysis is needed [5, 7, 8].

The analysis of verbal forms consisting of several morphemes also belongs - in the case of compound tenses - to the sphere of morphological analysis. The same holds true if we are to analyze nominal predicates, connected with verbs as есть,
являться, представлять собой. E.g. будет выражать; был создан; был предметом; является предметом; представляет собой машину.

In the case of syntagmas of the type "нужно производить, может быть выражен" it may be questionable - because of the modifying role of "нужно, можно" - whether their analysis is to be considered as a morphological or as a syntactical one. This question is, however, a purely theoretical one and has no great importance from the point of view of the practical work in MT. But it seems more suitable - because of the delimitation of the morphological analysis from the syntactical one - to consider the analysis of syntagmas of the abovementioned type as a part of the morphological analysis, too.

In this way, morphological analysis includes the analysis of separated morphemes, of so-called attributive constructions in which attributes are in agreement with substantives and the analysis of predicative constructions.*

2. The syntactic analysis has to clear up the role of the morphemes (sequences of morphemes) in the sentence, i.e. the interconnections among the different words (word-groups) of the sentence under consideration. As the result of that analysis we must obtain the structural formula, on the basis of which the corresponding sentence in the target-language may be built up. The syntactical analysis necessitates such a code-system which gives suitable information with regard to the matter, with what kind of elements each element of the sentence under consideration (the morphemes or morpheme-groups of the sentence) may form a syntagma, i.e. what kind of "government" they have. (For the present purposes we shall understand

* Work on MT in Hungary has so far included only morphological analysis of Russian. (Cf. Computational Linguistics I, Edited by the Computing Centre of the Hungarian Academy of Sciences, Budapest, 1963.)
here by "government" all the rules concerning the interconnec-
tions between morphemes.)

We shall take the verb as the starting point of the
syntactic analysis. The choice of the verb is motivated by
the fact that, on the one hand, the verb represents one of the
morphologically most identifiable elements of the sentence and
that most of the substantival morphemes (nominal groups) are
linked with the verb, on the other.

In Russian the morphemes linked with the verb may be de-
vided into two groups:

a) Nominal complements expressed by cases only (according
to the Russian grammatical terminology: подлежащее;
прямое и косвенное дополнения,

b) Prepositional constructions or adverbial complements.

This subdivision is necessitated by practical purposes.
The complements expressed by mere cases may be linked with par-
ticular verbs, i.e. in the cases of knowing the verb it may be
stated which complement may be expected to occur in the given
sentence and which cannot occur at all. At the same time the
complements of the type b) may occur more or less in connection
with each verb (regardless of the meaning of the verb).

A similar subdivision may be found in Tesnière's theory of
sentencehood. He calls the complements a) "actant" and the
complements b) "circonstant".*

* No doubt, some of Tesnière's ideas are theoretically very
disputable (e.g., that the subject is always subordinated
to the predicate), but his theory - as S. Károlyi has recently
pointed out - contains from a practical point of view many
useful points 5. The subdivision of the complements of
the verbs in "actant" and "circonstant" may be a useful ba-
sis for the compiling of a suitable code-system for the
syntactic analysis in MT. Otherwise, to make a difference
between the complements of the verb expressed by mere cases
and those expressed by prepositional constructions cannot be
considered as an arbitrary procedure only; it is the registra-
tion of the regularities concerning the case-system of the
2.1 With regard to the complements expressed by more cases we may subdivide the verbs in the following manner:

\[ N_{\text{Nom}} \rightarrow V \rightarrow N_{\text{Acc}} \]  

(1)

This group contains verbs having only a substantive in nominative (or a nominal group containing a substantive in nominative). This represents for the most part, the group of the intransitive verbs.

\[ N_{\text{Nom}} \rightarrow V \rightarrow N_{\text{Acc}} \rightarrow N_{\text{Dat}} \]  

(2)

i.e. besides the subject the verb governs an objective complement (besides the subject a substantive in accusative or a nominal group containing a substantive in accusative is to be expected). *\(^x\)/

\[ N_{\text{Nom}} \rightarrow V \rightarrow N_{\text{Acc}} \rightarrow N_{\text{Dat}} \]  

(3)

i.e. besides the subject an objective and a dative complement (a substantive or a nominal group) is to be expected. *\(^x\)/

\[ N_{\text{Nom}} \rightarrow V \rightarrow N_{\text{Acc}} \rightarrow N_{\text{Instr}} \]  

(4)

i.e. besides the subject an objective and an instrumental complement (a substantive or nominal group) is to be expected. *\(^x\)/

\(^x\)/ The occurrence of the abovementioned complements is not necessary, only expected; we obtain, of course, correct sentences also when some of the complements fail to appear. E.g. "Петр читает"; "Петр читает книгу".
2.2 With verbs having different governments in Hungarian the coding contains, besides the code-number referring to the government, also a device referring to the corresponding "translation" into Hungarian. E.g.

\[ \text{N}_{\text{Nom}} \rightarrow V \quad \text{N}_{\text{Nom}} \rightarrow V \]

\[ \text{N}_{\text{Nom}} \rightarrow V \quad \text{N}_{\text{Nom}} \rightarrow V \quad \text{N}_{\text{Nom}} \rightarrow V \]

2.3 The verbal complements expressed by a prepositional construction are not specific to the different verbs. The expression of the temporal, local etc. circumstances of any action or any event may occur in the sentence. In addition to adverbs, these circumstances may be expressed, by prepositional constructions being adverbial complements joined with the verb. A problem arises in connexion with the prepositional constructions when they stand behind the substantive or the substantival group, may also be the attributive complements of a substantive or a substantival group standing before it. In such cases further analysis is required. The decision of the attributive or adverbial role of the prepositional constructions standing behind the substantive or substantives demands a particular information-system based on the system of possible governments. In previous publication the present author dealt in some detail with a possible type of the coding of such constructions [7].
2.4 It is possible to decide whether the adverb in a given sentence is a complement of a verbal or adjectival morpheme without a special code-system only on the basis of special devices.

2.5 The complements of the substantives and adjectives having unverified governments may be found out by the aid of particular code-numbers (управление + Н_{Instr} равный + Н_{Dat}; характерный + для + Н_{Gen}).

2.6 The interconnection between the substantive in genitive and the substantive standing before it may be determined in the course of the morphological analysis. To decide whether a substantive in genitive (or a nominal group in genitive) represents really a possessive relation or only a quality (gen. qual.) is impossible without taking into account some semantic features, i.e. without semantic analysis, this problem goes, therefore, beyond the frame of syntactic analysis.

3. After the morphological and syntactical analysis roughly outlined here, it is obvious that there exist a great many cases when we cannot get the necessary information by means of a lower level analysis. In these cases the determination of the role of the different elements of the sentence is only possible if we also take into consideration the semantical features of these elements. That is what we call semantic analysis.

The most frequent cases in which semantic analysis is required are the following:

3.1 Substantives with endings of nominative and accusative (or nominal groups with such endings).
The nominative and the accusative of neuter substantives, of masculine inanimate substantives and of feminine substantives ending in 
are morphologically identical. In many cases the sentence turns out to have such a structure that the grammatical role of the substantive (nominal group) is self-evident without any analysis. So, for instance, if the verb of the sentence does not govern an object, the substantive (or the nominal group) stands doubtlessly in the nominative: ...явление происходит...

The same holds true, as a matter of course, if the given verb governs an object but, the verb stands in passive voice or is reflexive: ...деталь обрабатывается...

Further analysis is not needed also in such cases when between the two substantives (or nominal groups) belonging to the transitive verb only one turns out to be homographic:

...завод получил машину...
...машина вращает деталь...

In the abovementioned cases the determination whether the homograph substantive (or nominal group) stands in the nominative or accusative is only a question of programming.

From the point of view of analysis only those cases pose a problem in which both substantives (or nominal groups), belonging to the transitive verb, are homographs. Since the word order in Russian is not fixed to the same extent as, for instance, in English, on the basis of formal features any one of the two substantives may be considered as subject or object. E.g.

...основное уравнение динамики получает такой вид в том случае...

...такой вид получает основное уравнение динамики в том случае...

or:

...завод выпустил такой продукт...
...такой продукт выпустил завод...
Further investigations are required to decide whether it is possible to analyze mechanically (i.e. for the purposes of machine translation) such constructions by the aid of a suitable system of semantic information or whether another type of information is needed.

3.2 Possessive relations.

In cases of possessive relations the question to be answered is whether the substantive in genitive is a possessive adjective or a descriptive adjective of the substantive standing before it. (We shall leave aside the cases in which it is possible to translate the Russian possessive relation into Hungarian with a compound word or it is expressed in Hungarian by the aid of an adjectival ending, because in these cases the translation by a possessive relation does not result in ambiguities. E.g.:

девятнадцать пар; вопросы технологии.

In Russian the possessive relations expressing a descriptive adjective have two distinctive features:

1) The substantive in genitive is always an adjective;

2) the substantive in genitive concerns quantity, quality, measure etc. The quality expressed by the possessive word (the substantive in genitive) is connected with the possession: if we change the role of the two morphemes, we get a real possessive relation. E.g.:

сталь высокого качества; качество стали; пути одинаковой длины; длина пути.

Thus, the solution of our problem is as follows:
1) We have to analyze further those cases only, in which the possessive word has an adjective.

2) It has to be determined whether the possessive word denotes quality and whether this quality may be the quality of the possession.

The solution of these problems requires such a code-system on the basis of which it may be stated which substantives may form a real possessive relation and with which other substantives. The inversion of the role of the two substantives (possession ➔ possessive word and vice versa) enables us to clear up the interconnection between the given possession and the quality expressed by the given possessive word. (But this is not an extra operation for the machine, it means only the comparison of the corresponding code-numbers.)

If the abovementioned conditions are not fulfilled (i.e. the possessive word has not an adjective; the possessive word does not denote quality; the possessive word does denote quality but a quality which cannot be attached to a possession), the construction must be translated as a possessive relation even if according to the code-system a real possessive relation is not possible. E.g. in the abovementioned "продукт завода" construction the substantives may occur also in a reversed order: ...завод этого продукта...". This construction does not express a descriptive adjectival relation (the word "продукт" does not denote quality), thus, although the code-system regards only the "продукт завода" possessive construction as suitable, nevertheless it must be translated by a possessive adjectival construction. If we should succeed in compiling such a code-system which would refer also to these "reciprocal" possessive relations, this part of devices could be omitted.

The question could be solved on the basis of such a semantic information-system which would reveal the different qualities the given substantives can have. In possession of such a system, the algorithm for the possessive relations could be simplified to a great extent.
2.3 Verbs with the ending -cь

The translation of constructions containing as predicate a verbal morpheme with the ending -cь (or in a participial construction a participle with the ending cь) can only be executed, if the passive or reflexive role of the verbal morpheme is cleared up and if the morpheme fails to be an expression for impersonal subject. This is necessitated by the problem of interpretation of the verbal form and of the substantive in instrumental case (or of a corresponding nominal group) occurring in the construction. Namely, the substantive in instrumental designs the agent of the verb in a passive construction (since Hungarian has no passive constructions, this will be the subject of the Hungarian sentence), but it is an adverbial modifier of means or an adverb of manner in a reflexive construction or in a construction with an impersonal subject. E.g.: ...перевозка материалов обеспечивается применением машин...; ...механическая работа обеспечивается силовыми машинами....

There are among the abovementioned constructions a great many of which the passive or the reflexive character can be found out by suitable programming without any semantic analysis:

1) If the verb is a verb with perfective aspect and the construction is a reflexive one.

2) If the verb cannot be connected with an adverbial modifier of means (see point 2.1) and the substantive in instrumental in the construction is inanimate and the construction is a reflexive one. The substantive in instrumental case (or the corresponding nominal group) denotes in such cases an adverb of manner.

3) If the substantive in instrumental case in the given construction is animate and the construction is a passive one.
4) If the verb governs - contrary to Hungarian - the instrumental case; the establishment of the grammatical role of the substantive in the instrumental case (or of the corresponding nominal group) can be made on the basis of particular code-numbers (see point 2,2). Such verbs are: называться, являться, оставаться.

5) In the light of the above cases further analysis is only required if the verb in the construction is a verb with a progressive aspect and ends in -ся and the substantive in the instrumental case is inanimate.

The following considerations may render help to compile the semantic information required for further analysis:

1) Extra coding must be carried out for substantives in the instrumental case (or for the corresponding nominal groups) which are always used in the same form as adverbs of manner. E.g.: полнотою, следующим образом, путём; ...подведённая работа используется полнотою....

2) Extra code-numbers may be assigned to the verbs, of which the forms ending in -ся express an impersonal subject. E.g.: называться, выражаться, получаться.

3) If the substantive in the instrumental case of the construction denotes an abstract notion, we do not regard the construction as a passive construction. E.g.: ...перевозка материалов обеспечивается применением машин....

4) If the substantive in the instrumental case has a concrete meaning, the question must first be answered, whether the verb of the construction occurs among the constructions for expressing the impersonal subject. If the answer is affirmative, we may consider the construction as one with an impersonal subject and
translate the adverbial modifier of means into Hungarian with a substantive with the adverbial ending -val, -vel. E.g.: ...механическая работа выражается формулой...; ...работа...получается интегрированием отдельных элементарных работ...

If the verb is not registered on the list of the verbs expressing an impersonal subject, we may consider the construction as a passive construction. E.g.: ...сопротивление развивается поднятием грузом...; ...механическая работа обеспечивается силовыми машинами...

5) The solution is also feasible by comparison \( ^\# \) of the two substantives of the construction (in accusative and instrumental case, respectively), taking into account, as a matter of course, the list of verbs expressing an impersonal subject. The basis of the comparison may be the possessive relation in the same way as we used it to clear up the grammatical role of the possessive relation (see point 3,2). If the substantive in the instrumental case may occur as the possessive word of the substantive in nominative, we consider the construction as a passive construction. Taking the above example we have: ...механическая работа обеспечивается силовыми машинами...

The possessive relation of the two substantives: работа машины. ...сопротивление развивается поднятием грузом...

The possessive relation: сопротивление груза.

The possessive relations such as "машина работы" and "груз сопротивления" do not occur among the possible ones.

\( ^\# \) The expression "обеспечивать работу" must be coded extra as an idiom.
in our code-system, thus, according to the above we consider them as passive constructions in the translation.

If the possessive relation of the two substantives is only correct in reverse order (namely, the possessive word is the substantive in the nominative), the construction is a reflexive one or expresses an impersonal subject. E.g.: ...механическая работа выражается формулой...

The possessive relation: формула работы.

The verb occurs on the list of the verbs which express an impersonal subject, the sentence can be translated as follows: "A mechanikai munkát képlettel fejezzük ki". The translation is not incorrect even though we consider the construction - merely on the basis of the possessive relation of the substantives - as a passive construction. We receive: "A mechanikai munkát képlett fejezi ki". This kind of translation renders unnecessary the extra coding of the verbs which may express an impersonal subject. E.g.: ...перевозка материалов обеспечивается применением машин...

The possessive relation: применение перевозки.

The above example may also occur in the following manner: ...перевозка материалов обеспечивается машинами...

According to the above the construction may be regarded on the basis of the code-system, which allows the possibility for the two substantives of forming a possessive relation (перевозка машины ), as a passive construction. The translation remains correct: "Az anyagok szállítását gépek végzik (biztosítják)".

It remains to be proved whether it is possible to determine the grammatical role of two substantives in a construction by means of the comparison based on the possessive relation in general. Up to the present we may only say that a great amount of examples shows the abovementioned regularities. From a pure-
ly practical point of view, it seems best, therefore, to use
the above proposed solution all the more because the code-system
and the devices applied for the solution of the problems in
point 3.2 may also be used here in the same manner.

With respect to the translation of the verbal forms
ending in -ca we have limited ourselves to such cases as are
important from the point of view of semantic information. L.Dezső
has recently tackled the question in detail (2,3). He also pre-
sumes in his algorithm semantic information.

3.4 Prepositional constructions.

The governments enable us to clear up whether the prepo-
sitional constructions standing after the substantive (or
substantives) play an attributive or an adverbial role in the
construction under consideration. This procedure is, however, not
enough in itself: quite frequently semantical information is
also required. E.g. the constructions ...установили печь на
новом заводе... and ...установили печь на твёрдом топливе...
are morphologically identical, it is impossible even on the
basis of government to differentiate between them: the role of
"на новом заводе" as an adverbial modifier of place and the
attributive role of "на твёрдом топливе" can only be stated
on the basis of the meaning of the verb and the substantive,
respectively of the substantive standing before the preposition-
al construction and of the substantive in the prepositional
construction: установить + на + a substantive denoting p.e. a
construction with an adverbial modifier of place; печь + на +
concrete noun: an attributive construction.

The solution of this problem requires a code-system which
gives information on the lexical meaning of the given morpheme.
4. For the solution of the problems mentioned in point 3, the vocabulary of the machine has to be completed with code-numbers concerning the suitable semantic information. The code-system giving semantical information supposes the subdivision of the vocabulary according to the categories of meaning. The different morphemes belong to certain sets (groups) according to their lexical meaning. The semantic information is therefore such a code-number which indicates in which group the given morpheme belongs.

The possibilities of interconnections between the different groups must be a part of the program. The programs which serve to clear up the syntactic role of the morphemes must take into consideration in the syntactic analysis, beyond the mentioned possibilities of government, the semantical possibilities, too. The program containing the semantic information plays an important role, besides in the determination of the syntactic role of the morpheme, also in finding out the meaning of the given morpheme. (E.g.: \ldots решить задачу за определённое время...="meghatározott időre". Here the meaning of -ra, -re of the preposition "sa" comes from the meaning of "time" of the substantive in the construction.

The possible possessive relations of the substantival morphemes may be compiled in a rather complicated table. This table, however, demands a great amount of capacity, it is, therefore, more suitable to build these constructions in to the program directly.

The number of groups necessary for the semantic information is determined by the size of the machine vocabulary, the length of the programs utilizing the semantic information is fixed by the requirements concerning the correctness of the translation. The more words the vocabulary contains and the more precise translation we can obtain, the more detailed the semantic information system is and the longer program are required.
5. The code-system consists of the following parts:
   a) the code referring to the form class (C.I.)²,
   b) the code referring to morphological features (C.II.)²/,
   c) the code referring to the governments, necessary for the syntactic analysis (C.III.),
   d) the code referring to semantic information (C.IV.).

C.I., C.III. and C.IV. refer to permanent features of the morpheme, C.II. must be determined during the program run [4,6].

The scheme of the mechanical analysis

```
  morphological analysis
     |__________________________|
     |                           |
     v                           v
  syntactic analysis
     |__________________________|
     |                           |
     v                           v
  semantic analysis
```

The scheme shows the "structure" of the program required for the analysis. The program begins with the morphological analysis of the word forms of a given text. In the case of homographic forms also syntactic and semantic relations are taken into consideration. As a result C.II. will be stored.

The second step is the determination of the syntactical role of the different morphemes by the aid of C.III. In many cases governments fail to give the required information, thus

² The code-system C.I. and C.II. has been elaborated by Gy. Helc [4].
semantic information must also be taken into account (C.IV.). The result of the syntactic analysis or of the semantic one, if any, must be stored extra: this will be the structural formula of the Russian sentence under consideration, which enables the synthesis of the corresponding Hungarian sentence.

The proposed semantic information-system is by no means complete, it aims only at the clearing up of some syntactic constructions. A completely unambiguous and adequate Hungarian translation would require such a semantic information-system, on the basis of which it would be possible to state mechanically all the possible semantic interconnections of morphemes. This would make possible to build up not only the syntactic structure of the given sentence of the source language but also its semantic structure, i.e. its "semantic" formula. The whole analysis would deliver two ranks of information, a grammatical and a semantical one. The synthesis on the basis of their cross-checks would provide an unambiguous and quite correct sentence of the target-language.

6. Since both the compiling of the government-system necessary for the syntactic analysis and the word-sets as the base for the semantic analysis can be carried out only with respect to a given vocabulary, the practical work may begin only after the vocabulary has been compiled. But it can already be stated in advance that only some particular verbal and nominal morphemes subdivided according to the government, may belong to the same group, and the same holds true with respect to the subdivision according to the meaning of the morphemes for the purposes of semantic analysis. The comparison of these two groupings may lead to interesting results of great linguistic relevance. It would be interesting to inquire into the size of the different groups, to state with what kind of other morphemes a given morpheme will belong to the same group, to clear up the differences and coincidences between the two classes of
the given morpheme. From the result received in this way we should be able to draw some conclusions regarding the semantical content of the words. Thus, this comparison may give information about the interconnections between the lexical meaning of the given morpheme and its syntactic function [1].
REFERENCES


