A PRELIMINARY STRUCTURAL TRANSFER SYSTEM*

by

WILLIAM D. FOUST
and
JULIA R. WALKLING
(The Computation Laboratory of Harvard University,
Cambridge, Massachusetts, U.S.A.)

INTRODUCTION

AN essential step in automatic language translation is the transfer of
syntactic structure from the source language to the target language. A
preliminary set of transformations which map Russian constructions into
English constructions is described in this paper. This set of transforma-
tions is to be applied to syntactically analyzed Russian sentences, and the
result of this operation must then be synthesised into English sentences
to complete the automatic translation process. The procedure presented
here follows rather closely the outline for syntactic translation pro-
posed by Yngve[1] and adapted by Oettinger.[2] It is also highly probable
that in parallel with the syntactic analysis, transfer, and synthesis,
there will be similar processes operating on semantic elements.

2. THE STRUCTURAL TRANSFER SYSTEM

A. *Introduction*

The system of structural transfer presented here is based on the pre-
dictive syntactic analysis system in experimental operation at Harvard.[3]
The idealized output of the syntactic analysis, consisting of well-analy-
zed Russian sentences, serves as input to the transformations. Each well-
analyzed Russian sentence is expressed as a string of basic syntactic
units (the so-called "preferred arguments" of the predictive syntactic
analysis).

These syntactic units are elements such as <u>indirect object</u>, <u>subject</u>,
or <u>predicate head</u>, which define the syntactic role of an item in the
structure of a sentence, rather than elements such as <u>noun</u>, <u>verb</u>, <u>parti-
ciple</u>, or <u>adverb</u>, which indicate only the possible grammatical character-
istics of the word itself. Thus, <u>noun</u> is not a suitable syntactic unit,
since it may fulfil such roles in the sentence as <u>subject</u>, <u>object</u>,

---

preposition complement (object of a preposition), or noun complement (the first word of a genitive noun phrase following a noun). Similarly, the string verb + noun + noun could be represented by such syntactic structures as predicate head + object + noun complement, predicate head + indirect object + object, or predicate head + subject + object. Precise definitions and descriptions of these units are given by Sherry.[3]

The representation of structures solely as strings of syntactic units is not, however, wholly unambiguous. For example, the representation subject + noun complement + noun complement does not indicate to which of the preceding units the second noun complement is related. It is, therefore, logically necessary to make the representation more sophisticated by adding linkages which indicate the connection among the individual units. These linkages, when obvious to the reader, are left implicit in the examples in this paper.

The basic operations of a structural transfer system, as indicated by Oettinger,[4] are insertion, deletion, modification, and permutation or re-arrangement. An example of a transformation involving insertion might be one which inserts a missing pronoun subject agreeing in person and number with the predicate head. A Russian particle such as и or же, for instance, which serves as an intensifier or emphatic, and has no corresponding structure in English, might be deleted by another transformation. Modification is involved in changing the tense of the predicate head in a dependent clause to create the proper sequence of tenses in English. The operation of permutation is utilized in the transformation from Russian to English word order within clauses. In addition, it should be noted that a given transformation may perform more than one of these operations. For example, both insertion and modification are needed to transform an infinitive predicate head in Russian into a subject plus the predicate head in the third person singular present tense. The transformations as they are introduced in the next section will be explained in terms of these operations.

The transformations in this system of structural transfer are not independent; that is, a transformation can produce a construction which can have one or more further transformations applied to it. The order in which the transformations are applied is therefore important, and they are presented in that order in the following section.

B. The *Transformations*

The following list is a provisional set of transformations which should form a basis for a complete structural transfer system. The transformations have been determined largely on the basis of (1) the system of predictive syntactic analysis, treated as a grammar, (2) syntactically analyzed texts (3) conventional grammar books, both Russian and English, and (4) the

intuition of the authors.

Adjective Predicate Head. In transforming a Russian adjective predicate head, the missing verb predicate head "be" is inserted and the adjective predicate head becomes an English predicate adjective (object of "be"). The inserted "be" is assigned present tense and assumes as grammatical characteristics the number and gender of the Russian adjective and the person of the subject, if it is expressed, or third person if the subject of the predicate head is not explicit. For example, the string основные елементы схемы выделены where выделены is an adjective predicate head, will eventually be translated as "fundamental elements of the circuit <u>are chosen</u>." A compound adjective predicate head, one connected to a previous predicate head by a co-ordinating conjunction, should be treated in the same manner. (In the statement of further transformations, it should be assumed that the compound form of a syntactic unit will be treated in the same manner as the unit with which it is compounded.)

Infinitive Predicate Head In Russian, an infinitive may serve as a predicate head, as does пренебречь   in the clause если пренебречь предварительным уменьшением which is translated "if one neglects preliminary reduction". The indefinite pronoun "one" is inserted as subject, and the infinitive predicate head is modified to a verb predicate head in the third person singular present tense.

Predicate Head without Subject. A personal pronoun subject is inserted when the predicate head has no explicit subject. This pronoun must correspond in person, number, and gender with the predicate head. However, the correspondence in gender between Russian and English is not very precise, because the gender of the English third person singular pronoun actually depends less on the gender expressed in the Russian verb than on both the gender and the animate-inanimate distinction in the antecedent of the Russian pronoun. If the predicate head does not have any expressed gender, the gender of the pronoun must depend completely on the antecedent.

   A Russian impersonal construction such as ХОЛОДНО or очевидно will be treated by this rule, combined with the adjective predicate head transformation. For example, ХОЛОДНО will be syntactically analyzed as an adjective predicate head with grammatical specification neuter singular. The adjective predicate head transformation will then insert the predicate head "be" in the present tense third person singular neuter, and then a neuter singular personal pronoun subject will be inserted by the current transformation. English synthesis will eventually produce the English string "it is cold". In this case it should be noted that the pronoun has no antecedent and therefore there is no problem of gender correspondence.

Predicate Head With Numeral Subject Any singular predicate head which has a numeral other than ОДИН as subject should be pluralized.

Compound Future When a future is expressed in Russian as a compound tense by a future form of the verb быть and an infinitive, this string is syntactically analyzed as verb predicate head plus verb master. This construction is transformed into a Russian canonical construction by the deletion of the original verb predicate head and modification of the verb master to a verb predicate head with the grammatical information of the original predicate head.

-ся Verbs The Interpretation and translation of verbs ending in -ся or -сь(often called reflexive verbs) are discussed by Lynch.[5,6] The transformations expressed in the following paragraph are based on the salient features of her work.

The addition of -ся to a transitive verb can either produce a completely different meaning for the verb or modify the basic meaning of the verb to passive, reflexive, or reciprocal. We are here concerned not with the former, since it cannot be treated by a structural transformation but with the latter, since it involves a structural change. For example, интересоваться the -СЯ form of инересовать "interest," is best interpreted as passive and therefore translated as "be interested". Мыться , formed from мыть "wash", should be interpreted as "Wash oneself" (reflexive). Встречаться from встречать "meet," takes on the reciprocal meaning "meet each other". The transformations corresponding to these three cases are as follows: For passives, the Russian verb is transformed to "be", with the grammatical information of the original verb, followed by a verb complement with past participle form. The reflexive is transformed into verb plus the object "X-self," where X stands for a pronoun in the objective case having the gender and number of the subject. The reciprocal becomes verb plus the object "each other"'

Lynch has already commented that for reflexives and reciprocals insertion of the object is often optional, since it is not always necessary to make the object explicit in English. For example, бреет "he shaves" has the reflexive бреется "he shaves himself," which may more normally be translated simply as "he shaves".

Sequence of Tenses In indirect discourse in Russian, the predicate head of the dependent clause is expressed in the same tense as it would be if it were in direct quotation. Thus the time of the dependent clause is relative to the time of the main clause. In English, however, the tense of the predicate head in the dependent clause in indirect discourse is a function both of the tense of the main predicate head and the tense of the dependent predicate head as it would be in direct quotation. The following example illustrates the differences between the use of the tenses in both direct and indirect discourse.

```
          Russian                        English

Он сказал: "Она пишет письмо."   He said, "She writes the letter",
   (past)        (present)         (past)       (present)

Он сказал, что она пишет письмо.  He said that she wrote the letter.
   (past)        (present)           (past)        (past)
```

When the Russian main predicate head is present or future, the tense of the
English dependent predicate head is the same as the Russian dependent pred-
icate head; that is, no modification of the tense is needed. When, however,
the Russian main predicate head is past, the tense of the English dependent
predicate head is modified according to the following pattern: the present
is made past, the past is made past perfect, and the future auxiliary
"shall" or "will" is replaced by its past form "should" or "would."

Third Person Imperative. The Russian expression

```
            пусть +     subject  +  predicate head
                   (third person)  (third person,
                                     present tense)
```

corresponds to an English imperative construction beginning with "let". The
particle пусть maps into the imperative "let"; the Russian subject into
the object of "let"; and the predicate head becomes a verb master in canoni-
cal form (i.e., the infinitive minus "to"). Thus, the Russian string

```
                      пусть он идет
```

is analyzed as

```
            пусть  +     subject   +       predicate head
                    (personal pronoun,  (third person,
                     third person,         singular,
                      singular,         present tense)
                       masculine)
```

which is transformed to

```
   predicate head   +      object    +   verb master
   (imperative,          (personal pronoun,  (canonical)
     "let")              third person,
                          singular,
                          masculine)
```

and can then be synthesised into the English string

```
                      "let him go".
```

Numeral Master Any singular noun or nominal pronoun which serves as
master of (i.e., is modified by) a numeral other than один should be

made plural. For instance, два газа in which газа is singular in form, should be translated "two gases".

Comparative Complement A Russian comparative complement is a genitive following a comparative adjective or adverb. If the чем ("than") is not expressed, it should be inserted preceding the genitive.

Negated Predicate Head. In English, a negated predicate head other than "be" or an auxiliary verb must be expressed by some auxiliary verb plus verb master. For example, the string "he walks" when negated becomes "he does not walk" or "he is not walking". The Russian construction negative (не) plus predicate head is transformed to predicate head (auxiliary) plus negative plus verb master in canonical form. If the original predicate head is present or past, the auxiliary inserted is "do," with grammatical information carried over from the original predicate head. If the original predicate head is future, the future auxiliary ("shall" or "will") is inserted, with person carried over from the original predicate head.

   When the negated predicate head is "be" or an auxiliary, the negative particle simply is moved to a position following the predicate head.

Interrogative Predicate Head. The main predicate head of an interrogative sentence must be transformed to auxiliary predicate head plus verb master in canonical form. The form of the auxiliary is the same as that used with a negated predicate head.

Adjective With Dependent Structure. In Russian, an adjective may have a dependent structure which separates it from its noun master. Typical dependent structures are prepositional phrases, objects, agents, or adverbs.

The following example is illustrative of such constructions.

Russian Word	Preferred Argument	First English Correspondent
в	Preposition	in
осуществленном	Preposition Complement	developed
нами	Agent (of participle)	(by) us
приборе	Preposition Complement master	device

   In English, a dependent structure may not separate an adjective from the noun which it modifies. No one transformation will always produce the desired result in English, since although the adjective is generally moved along with the dependent structure to follow the noun, it sometimes should not be moved. In the example,

Russian Word	Preferred Argument	First English Correspondent
измерить	Verb Master	(to) measure
среднюю	Object	average
за	Preposition	over
много	Preposition Complement	many
периодов	Preposition Complement master	period(s)
амплитуду	Object master	amplitude
этой	Noun complement	(of) this
частоты	Noun complement master	frequency

the prepositional phrase which modifies the adjective "average" must be
moved without the adjective. This may be due to the fact that numerals
and some other quantitative expressions must precede the noun which they
modify. A second complication illustrated by this example arises when the
noun has one or more additional dependent structures following it, such
as noun complement or relative clause.

Basic Ordering Within Clauses The major structural elements of a clause,
except in the case of relative clauses and main clauses of interrogative
sentences, should be ordered as follows:

1. Subject
2. Predicate Head
3. Object
4. Indirect Object
5. Agent or Instrument

Included with each of these elements are any constructions which are de-
pendent on the given element. Machine methods for grouping dependent
structures and re-ordering the major structural elements have been deve-
loped by Plath[8] in connection with a scheme for automatic sentence dia-
gramming. It is clear that this is not the only ordering possible in
English, but the adoption of such a canonical form leads to English which
is always syntactically correct and readable, if monotonous and stylisti-
cally inelegant.

In clauses introduced by a relative pronoun, the ordering is the same,
except that all elements preceding and including the relative should not
be re-ordered. The position of the relative must be retained because, al-
though it is linked to the subordinate clause in some essential function,
it must still be linked to its antecedent in the main clause. In the
Russian string shown below, which is translated "the book, which I gave
to him", the relative pronoun is not moved to the normal object position,
but the rest of the dependent clause is re-ordered.

Russian Word	*Preferred Argument*	*First English Correspondent*
книга	Subject	book
,	Comma (introducing clause)	,
которую	Object (relative)	which
я	Subject	I
ему	Indirect Object	(to) him
дал	Verb predicate head	give (past)

The main clause of an interrogative sentence has the subject placed between the auxiliary predicate head (which has been inserted by the interrogative predicate head transformation) and the verb master. The remainder of the elements are arranged in normal clause order.

Agent, Instrument, Indirect Object, and Noun Complement. The four syntactic units agent, instrument, indirect object, and noun complement are transformed by the insertion of an appropriate preposition and the modification of the given Russian element to a preposition complement. In addition, any elements which serve as master of any of the elements modified must themselves be modified to preposition complement master. The preposition inserted for agent is "by"; for instrument, "with"; for indirect object, "to;" and for noun complement, "of".

3. GENERAL DISCUSSION OF THE SYSTEM

Several general characteristics of this structural transfer system become apparent upon examination; these will be discussed in the following paragraphs.

The ordering of the transformations could be based on two factors: the construction on which a given transformation operates, and the type of operation performed by a transformation. As an example of the former approach, one might first do all transformations on the subject of a sentence, and then deal with the predicate head. On the other hand, one might perform all deletion operations first, then modification, and so forth. It is obviously impossible to use either one of these types of ordering exclusively, since, as has already been indicated, many of the transformations affect more than one syntactic unit and perform more than one type of operation. In the ordering of the set of transformations presented in the previous section, it seemed desirable to perform the operation of insertion before making any modifications, since we may wish to modify items which have been inserted. For example, an inserted subject of a previously subjectless predicate head may later be modified to object by the third person imperative transformation. Since re-arrangement depends on certain basic units which may have been modified in the

course of the transformations, the operation of re-arrangement must follow modification. Thus, the major order in which the transformations have been presented and are to be applied is insertion, modification, and re-arrangement. Within this outline of ordering, some attempt has also been made to group together operations affecting the same syntactic units. A further narrow category involving both insertion and modification has for purposes of convenience in re-arrangement been ordered following the re-arrangement. In this type of transformation, a preposition is inserted preceding a noun, and the syntactic unit of the noun is modified to preposition complement. The units <u>agent</u>, <u>instrument</u>, and <u>indirect object</u>, which fall into this category, are lost by the modification. Since these units are important in the transformation of the sentence into canonical order, they must be retained until after the re-arrangement is accomplished.

The final step in conversion to English structure is the inflection[7] of the nouns and verbs of the constructions created by the transformations. This inflection cannot strictly be said to form a part of the syntactic structural transfer system, since it does not operate solely on the basis of the syntactic units, the "preferred arguments" of the syntactic analysis. It has, however, two characteristics which make it similar to the structural transformations. First, in certain cases, notably those involving the future and the past gerund of the verb, structural elements are added. Second, the Russian morphological information on which the inflection is based can be said to be a part of the syntactic unit in that its ambiguity is reduced in the determination of the "preferred argument" by the syntactic analysis.

The structural transfer system leads from Russian constructions expressed in Russian syntactic units to English constructions expressed in English syntactic units. There is some option in the form which the intermediate expressions may take. The adjective predicate head transformation, for instance, may be viewed simply as a transformation of the Russian structure into the canonical Russian construction of a verb predicate head plus a verb complement. It may, on the other hand, be viewed as a transfer to the construction predicate head plus predicate adjective. In some instances, a purely English construction must be introduced, as in the third person imperative transformation, where "let" is supplied.

Since there is such wide variety in the nature of the transformations in the system, it is difficult to determine a convenient notation for use in the statement of the transformations. A rough attempt at a suitable notation appears in the previous section in the description of the transformation for the third person imperative. If an adequate notation could be developed, it would constitute an intermediate language, as is indicated by Oettinger:[9]

> The development of rigorous, explicit notational systems
> for describing properties of strings and relations among
> these suggests to many that such systems may eventually
> serve the role of synthetic *intermediate languages,* to
> which properties of many different natural languages may
> be related.

Rather than developing an original notational system, it may be possible to incorporate notational features from COMIT,[10],[11] a programming language designed for use in automatic translation.

Proper re-arrangement within a sentence may be made difficult when more than one construction depends on (i.e., is linked to) a given construction, In the example,регистрация сигнала , которая характеризуется временем ΔT("recording of signal, which is characterized by time Δt"), both the relative pronoun которая and the noun complement сигнала are linked to the noun регистрация. In this case, although not necessarily in general, this string can be made unambiguous in translation by making the noun complement a prepositional modifier to the noun, thus permitting the relative to follow its antecedent immediately, i.e., "signal recording, which is ... ." In this and many similar cases, Russian uses morphological information (in this instance, gender), whereas English must rely on word order, which may introduce ambiguity.

As can be observed from the example in the preceding paragraph, in certain situations some information may be lost in translation from Russian to English. This may involve not only Russian morphology, as in the case just cited, but other characteristics of the language as well. The complexities of the verb system of Russian, such as aspect, cannot be fully expressed in English without introducing an awkwardness of style which may not be wholly warranted by the accuracy gained. Russian, in addition, makes use of word order to lend emphasis to a given part of a sentence. This emphasis is lost in the re-arrangement procedure which converts sentences into canonical form, but it might be regained by using some device such as italics, if some satisfactory method of determining the emphasis by means of the syntax could be developed.

Many constructions have not been included in the present discussion of structural transfer. Some are reasonably important in an idealized structural transfer system, but have so far proved difficult to treat. For instance, there are the problems of transfer of punctuation; position of adverbs in each language; addition of implicit objects, predicate heads, and other elements in elliptical constructions; double negatives; and addition of articles in English. Many minor or low-frequency constructions have been omitted as unimportant for a preliminary exposition. These include imperatives, contrary-to-fact conditionals, constructions involving individual lexical peculiarities, and various minor uses of the oblique cases.

4. CONCLUSION

The set of structural transformations described in this paper constitutes a preliminary approximation to a complete structural transfer system. Further application of the transformations to syntactically analyzed Russian text will hopefully lead to successively closer approximations through the detection of errors and the realization of new transformations. It is felt that the present system includes the major transformations essential to structural transfer. This provisional set of transformations may form an embryonic comparative grammar of Russian and English, which when fully developed should make possible a translation of considerably better quality than a word-by-word translation.

REFERENCES

1. YNGVE, V.H., "A Framework for Syntactic Translation," *Mechanical Translation* 1957 4 No. 3. 59-65.

2. OETTINGER, A.G., *Automatic Language Translation,* Harvard University Press, Cambridge, Mass, 1960, p. 119.

3. SHERRY, M.E., *Syntactic Analysis in Automatic Translation,* Doctoral Thesis, Harvard University, 1960, Chapter 5.

4. OETTINGER, A. G., Op. cit., p. 125.

5. LYNCH, I.B.-M., "A Proposed Treatment of the Russian —ся Verbs for Purposes of Automatic Translation, *Mathematical Linguistics and Automatic Translation,* Report No. NSF-4, Harvard Computation Laboratory, 1960, Section XIV.

6. LYNCH, I.B.-M., "Russian -ся Verbs, Impersonally Used Verbs, and Subject/Object Ambiguities," *Mathematical Linguistics and Automatic Translation,* 1961, Report No. NSF-6, Harvard Computation Laboratory.

7. WALKLING, J.R., "The English Inflector", *Mathematical Linguistics and Automatic Translation,* 1960, Report No. NSF-4, Harvard Computation Laboratory, Section XII.

8. PLATH, W.J., "Automatic Sentence Diagramming", *Mathematical Linguistics and Automatic Translation,* 1961, Report No. NSF-6, Harvard Computation Laboratory.

9. OETTINGER, A.G., op. cit., p. 120.

10. YNGVE, V.H., "A Programming Language for Mechanical Translation", *Mechanical Translation,* 1958. **5**, No. 1, 25-41.

11. COMIT - *Programmers Reference Manual,* Massachusetts Institute of Technology, 1960, unpublished manuscript.