

Seminar Work Paper
MT - 73

Georgetown University
Institute of Languages & Linguistics
Research Project in
Machine Translation

REPORT OF GROUP II
OF THE GU PROJECT IN MT RESEARCH

Submitted
by

Paul L. Garvin, Dorita Lochak, Madeleine Mathiot, and
Christine Anne Montgomery.

Copyright, Georgetown University
Institute of Languages & Linguistics
1719 Massachusetts Ave., N.W.
Washington, D.C. U.S.A.
1957

INTRODUCTION

As a result of one year's working experience with machine translation, we have arrived at the following results and perspectives in regard to syntax in MT:

(1) We have developed and to a considerable extent verified a general theory of syntax in terms of machine translation;

(2) We have developed and refined a methodology for effecting translation, based on our theory of syntax;

(3) We have applied this methodology to what we consider the crucial structural features of the syntactic translation problem;

(4) We are now ready to complete our treatment of the syntactic translation problem by applying our method to the remaining details of the problem.

I. OUR THEORETICAL FRAMEWORK

(a) The "Fulcrum" Approach to Syntax

Our approach to syntactic problems in machine translation is based on our conception of language in terms of what we call "The Fulcrum Approach to Syntax", which is a general theory of syntax based on our previous knowledge of the syntactic problem area and on our experience on the MT Project.

We have chosen the term "fulcrum" because we metaphorically conceive of our MT rules as the lever by which the message is lifted from Russian into English. The fulcrum is that constituent of the sentence, clause, or lesser syntactic unit, which provides the best point of leverage for its translation.

Our syntactic research has thus taken its point of departure with the established syntactic categories of Russian grammar, and has led to a reexamination of these units in order to find their fulcra.

Our research has revealed that the fulcrum determines all the possible varieties of a syntactic unit of a given type (for instance, a clause or phrase), and thereby allows the precise recognition of its boundaries. Thus, once the fulcrum of a unit has been stated, only those questions have to be asked which relate to that particular type of unit as determined by

its fulcrum.

The portion of a syntactic unit which is not included in its fulcrum we call the "remainder" of the unit,

(b) The Fulcra of Syntactic Units

The types of syntactic units which we have dealt with so far for purposes of machine translation, are the following, in descending order of complexity:

sentence
 clause
 clause member (subject, predicate, object)
 construction (relative construction, see (5) below)
 phrase

Clause members fall into two basic types: predicates, and nominal blocks. The latter classification includes both subjects and objects.

Phrases of interest to our machine translation procedure include nominal phrases and prepositional phrases. We have so far not found it necessary to break down the predicate block into constituent phrases.

The following is a table of syntactic units and their fulcra:

Unit:	Fulcrum:
Sentence	no fulcrum
clause	predicate block
predicate block	verb or verb substitute
nominal block (subject or object)	noun or noun substitute
relative construction	relative pronoun or predicate block
nominal phrase	noun or noun substitute
prepositional phrase	preposition and noun or noun substitute

The above fulcra have been established on the basis of the following criteria:

(1) Sentence. We hold that the sentence as defined for MT purposes has no fulcrum qua sentence, since under MT sentences must be included units of such varied structure as individual words serving as headings, as well as elaborate multiclaue constructions. We therefore divide MT sentences into the following types:

clauseless sentences (those, such as headings, made up
of less than a clause)

clause sentences

single-clause sentences
multiclaue sentences

Each of the clauses included in a clause sentence will contain its own clause fulcrum (one in a single-clause sentence, several in a multiclaue sentence); a sentence thus may contain from zero to several clause fulcra, but no sentence fulcrum.

This is procedurally significant, because it allows us to treat clauseless sentences in a much simpler manner than clause sentences. That is, a unit between two functional periods, or between a functional period and the beginning and the end of the text. We use the term "functional period" to differentiate it from periods used in abbreviations or formulas: we intend it to include not only periods but all period-type punctuation marks.

ences; it further allows us to separate and delimit from each other the several clauses of a multiclausal sentence by using their several fulcra.

(2) Clause. The fulcrum of the clause is its predicate block, because it is the type of the predicate which determines the possibilities of occurrence, position, and type of nominal blocks within the clause. Thus, for instance, nominal blocks serving as objects are possible only with transitive predicates; nominal blocks centered around plural nominal forms or a severality of singular forms, or a combination of both, may serve as subjects only with plural predicates.

(3) Predicate Block. The predicate block includes, in addition to the one or several verbs or verb substitutes*, the various possible dependent adverbs and particles (such as the negative particle HE), the occurrence of which is determined by the characteristics of the verb or verb substitute. The latter thus serves as the fulcrum of the predicate block.

(4) Nominal Block. By a nominal block we mean the one or

* By a verb substitute we mean, for instance, a predicative adjective, adverb, or participle serving as the center of the predicate block.

several nouns or noun substitutes* serving as subject or object, together with all modifiers in agreement with it or them, as well as all the nominal and prepositional phrases and relative constructions directly or indirectly governed by it or them.

"Directly governed" here refers to government by the noun or noun substitute itself. "Indirectly governed" refers to the fact that the directly governed unit may in turn govern additional units; the latter are those included under the classification "indirectly governed".

It is the directly governing noun(s) or noun substitute(s) which determine(s) the possibilities for the remainder of a nominal block and thus serve(s) as the fulcrum (fulcra): thus, for instance, the number and gender of modifying adjectives will depend on the noun or noun substitute; only nouns or noun substitutes governing the dative will be followed by dependent nominal phrases in the dative.

(5) Relative Construction. We prefer this term to the traditional term "relative clause" because, although the relative construction is clause-like in its internal structure, its external functioning with regard to the translation of the sentence as a whole is that of being a constituent of a nominal block. Thus, a different fulcrum will serve when the sentence

*By a noun substitute we mean, for instance, an adjective, or pronoun functioning as nouns.

as a whole is subject to a translation operation and the external functioning of the relative construction is relevant, and a different fulcrum when the relative construction itself is to be translated and its internal structure is relevant.

In terms of its external functioning, the fulcrum of the relative construction is the relative pronoun, since it is the necessary condition for the occurrence of the remainder of the construction.

In terms of its internal structure, the relative construction is treated like any clause, and its fulcrum is the predicate block.

This dual treatment of the relative construction is facilitated by the fact that its boundaries are in Russian orthography indicated by commas.

(6) Nominal Phrase. The fulcrum of the nominal phrase is the noun or noun substitute, since it will determine the agreement characteristics of the accompanying modifiers,

(7) Prepositional Phrase. This is the only unit having two equivalent fulcra in our analytic scheme so far, for this reason: on the one hand, the preposition will determine the possible case(s) of the nominal material governed by it, and hence serve as a fulcrum: on the other hand, the presence of a noun or noun substitute is the necessary condition for the occurrence of a preposition and from this standpoint, the noun or noun sub-

stitute serves as a fulcrum. Either fulcrum may be utilized in a given translation routine, depending on the direction of search,

(c) The Application of the Fulcrum Approach to Syntactic Searching

Our syntactic searching rules are constructed in such a way that they go into effect only if a clause fulcrum, that is, a predicate block, is present in the sentence - since syntactic searching is required for clause sentences only, not for clauseless sentences, for which a string or semi-syntactic operation* is sufficient.

Thus, the identification of the predicate block initiates the syntactic search.

Our rules next contain provisions for dividing a multiclausal sentence into its several constituent clauses by identifying their several fulcra (predicate blocks), and for handling each clause in terms of the possibilities admitted by its particular fulcrum.

Within each clause, the fulcrum of the subject block is first identified, and if necessary for a translation operation, the remainder of the block is identified by recognizing the fulcra of its constituent phrases and constructions; similarly, the fulcrum and remainder of an object block can, if necessary, be recognized and subjected to a required translation operation.

*See section II (a).

The details of the above procedure will be illustrated in section III further below.

II. OUR TRANSLATION RULES

(a) The Hierarchy of Rules

Our translation rules form a hierarchic logical system based on the properties of the Russian language and the requirements of English equivalence.

We have three levels of rules, which are formulated to go into effect consecutively as required, with the subsequent levels presupposing the completion of the preceding ones.

(1) String rules, the search span of which is limited to the immediate environment. These are used to solve certain morphological and lexical ambiguities.*

(2) Semi-syntactic rules, the search span of which extends beyond the immediate environment, but does not include an entire clause. These are used to solve, among others, certain syntactic problems of detail such as the rearrangement of pre-

* Since our string operation has been presented in detail in the MT seminar and in Garvin's paper, *Linguistic Analysis and Translation Analysis* at the 8th Annual RTM on Linguistics and Language study, we shall not go into it here.

ceding participle and governing noun or noun substitute.*

(3) Syntactic rules, the search span of which potentially covers the entire sentence. These are designed to identify the total structure of a clause sentence, and to effect all those intricate translation operations which have not been solved at the preceding levels of the string and semi-syntactic rules.

Each successive level of the hierarchy goes into effect only when required by the structure of the Russian sentence to be translated: thus, clauseless sentences not containing special structures (such as participles preceding governing nouns) are handled by the first-level string rules alone; clauseless sentences which do contain special structures are handled by the first and second levels of string and semi-syntactic rules; only clause sentences require the entire hierarchy of all three levels, including the third level of syntactic rules.

(b) The Structure of Syntactic Rules

Just as our overall system of rules is hierarchically structured, so each of our syntactic rules has its own hierarchic structure.

* For an example of a semi-syntactic rule, see MT Work Paper # 60.

In constructing our rules, we use the following operational units:

rules, which consist of
sub-rules, which consist of
step sequences, which consist of
steps.

Each step of a rule is of the nature of a binary question, requiring a yes/no answer.

Any given step of a rule may through one of its answers require branching out into standardized repeatable step sequences, after completion of which a subsequent step of the rule is resumed. These special branching-out sequences are constituted by the following operational units:

loops, which may consist of
sub-loops.

Certain frequently repeated steps or very brief step sequences have been standardized as routines, and are incorporated into rules, sub-rules, loops and sub-loops as required.

(c) The Properties of Our System of Rules

1. Our system is an open system, in three senses:

(i) Our hierarchy of rules can, if required, be expanded by the introduction of new levels (for instance, semi-string rules) without necessitating a revision of the system as a whole.

(ii) Additional operational units, such as sub-routines or sub-sequences, may be devised as required and inserted without disturbing the system.

(iii) Any sequence of steps, whether part of a rule, sub-rule, loop, or sub-loop, may be interrupted at any point for the insertion of additional steps, without necessitating the revision of the entire sequence.

2. Our system is compressible, in two senses:

(i) The number of rules, sub-rules, loops, and sub-loops can be reduced by combining two or more rules into a single rule, two or more sub-rules into a single sub-rule, etc.

(ii) The length of any rule, sub-rule, loop or sub-loop can be reduced by compressing the sequences of steps contained in each.

III. THE APPLICATION OF OUR SYSTEM - ANALYSIS OF A SAMPLE RULE

(a) The Scope of Our Rule

To illustrate how the fulcrum approach to syntax is combined with our system of hierarchic operational units for the construction of a translation rule, we have chosen one of our set of syntactic searching rules, rule 4110.

This rule is designed to effect syntactic searching and to complete the translation of sentences containing clauses with plural intransitive predicates.

We consider this type of clause among the most complex translationally, because:

(1) having a plural predicate, it allows the occurrence of nominal subject blocks containing either plural fulcra or several singular fulcra or varied combinations of both;

(2) having an intransitive predicate, the occurrence of a subject block containing, in its fulcrum, and class of nouns or noun substitutes, is possible either preceding the predicate block or following the predicate block, in the latter case requiring rearrangement.

The subject block may, in addition to the plural and singular forms serving as its fulcrum, contain a variety of dependent and appositional structures, made up of all possible

types of nominal and prepositional phrases as well as relative constructions. Thus, by covering the subject block of this type of clause we have been able to come close to exhausting the inventory of Russian nominal structures.

In order to isolate the problem of nominal structural types, we selected as our starting point that plural intransitive predicate block which has the simplest internal structure (while, of course, allowing the same complexity in the remainder of the clause as the more intricate predicate blocks of the same type). Such an internally simple, but functionally complex one-word fulcrum is the plural intransitive verb, without accompanying adverbs or particles.

(b) The Linearity Requirement

In accomplishing the syntactic search, **our** rule - like any other MT rule - has to have provisions for the linear sequence of items in the input, and for the fact that grammatically relevant items may be separated by intervening structures of varying length and complexity. Thus, an elaborate system of readdressing and step repetition has to be built into the rule.

This is accomplished by two devices:

(1) referral steps - steps requiring the repetition of a previous step at a new location;

(2) loops - standardized search step sequences applicable in a great variety of locations.

(c) The Need for This Type of Rule

The immediate translation objective of rule 4110 can be stated very simply; to determine whether the Russian verb should be translated by an English verb preceded or not preceded by a pronoun, and - when necessary - to effect rearrangement of subject and verb.

We at first tried to meet this objective by incorporating a few simple "tricks" into our string operation. We found, however, that a genuinely generalized solution can only be obtained by identifying the subject block of the verb, no matter where within the clause this block may be located.

In attempting to identify the subject blocks of different varieties of clause sentences with plural intransitive predicates, we discovered that this identification is possible only if an exhaustive recognition procedure is held in readiness for those cases where the subject block is not immediately accessible to identification.

We also found that in sentences where the subject block precedes the Russian verb, identification of the fulcrum of the subject block is sufficient to effect the choice between the presence and the absence of a pronoun in the English translation.

However, in sentences in which the subject block follows the predicate, the above choice is only one of two translation requirements. The second requirement is the rearrangement of the order of subject block and predicate, since in English the subject precedes the predicate. To effect this rearrangement, it is necessary to identify not only the fulcrum of the subject block, but all of its remaining constituents as well, since the translation of the entire block (and not merely that of its fulcrum) has to be shifted into a position preceding that of the translation of the verb.

Thus, by solving an apparently simple translation problem in a generalized way, we have succeeded in working out a syntactic search procedure capable of covering a most complex variety of nominal material within the search span of an entire sentence.

Our procedure now allows us to recognize and handle the structure of any constituent of a nominal block; by exhausting all the possible types of clause fulcra we shall in addition be able to cover all possible combinations of nominal blocks and of constituents of nominal blocks.

We are able to accomplish this, because our procedure operates with a number of individually regroupable operational units (rules, sub-rules, step sequences, steps, loops, sub-loops, routines). A certain grouping of these units constitutes our

rule 4110; other groupings constitute further rules that we have constructed. Additional groupings will constitute the rules required to solve the remaining varieties of the syntactic search problem.

(d) The Design of Our Rule*

Questions

1. Is a clause fulcrum present?

To ascertain whether a clause sentence of the required type is present, a decision point diacritic P 4110 is assigned to plural intransitive verbs, thus serving to recognize the clause fulcrum. This diacritic will initiate rule 4110 whenever such a clause fulcrum is present.

A. Search of Span Preceding the Verb (Sub-rules 4111,4112)

2. Is the clause fulcrum a one-word fulcrum?

Only if the verb is unaccompanied by adverbs or particles will there be a one-word clause fulcrum. The presence or absence of adverbs of particles before the verb is checked by step (1) of sub-rule 4111.

3. Is our fulcrum the fulcrum of a real clause or that of a relative construction?

Since relative constructions are clause-like in their intern-

* A preliminary version of rule 4110 (which has since been somewhat revised, especially in regard to step sequence (2) (a) of sub-rule 4111) is available in GU MT Seminar Work Papers # 61, #52.

al structure, part of the role of step sequence (2) (a) [together with referral step (5)] is to ascertain whether the verb carrying P4110 is a genuine clause fulcrum or the internal structure fulcrum of a relative construction [see section II (b) (5)]. In the latter case, sub-rule 4112 will go into effect to recover the relative pronoun and exhaust the relative construction.

4. Does the real clause have a potential subject?

If, as per the above, P 4110 has been found to mark the fulcrum of a real clause, step sequence (2) (a) will assume the role of searching for the potential fulcrum of a subject block (that is, a noun).

5. Is our potential subject-block fulcrum the fulcrum of a real subject block? (= Is the potential subject a real subject?)

If stems (2) (a) (iv) and (2) (a) (v) establish the presence of a potential subject-block fulcrum (that is, of a noun), step sequences (2) (b) - (f), (3) and (4) will unequivocally establish whether the potential fulcrum is a real fulcrum (that is, a noun in the nominative plural), or one of several real fulcra (that is, a noun in the nominative singular); in the latter case, the remaining necessary fulcrum (that is, additional noun in the nominative, linked to the first noun by conjunction or comma) will be established by

Loop I through steps (4) (e) (i) and (4) (e) (ii).

Referral steps (5) - (11) serve to carry this identification process through from the verb to the beginning of the sentence. During this procedure, a return to steps (2) (a) (ii) and (2) (a) (iii) via the referral steps allows us to by-pass an intervening relative construction (which, however, is translated in terms of question 3. above) or comparable intervening structure, by recognizing the commas marking its boundaries.

B. Search of Span Following the Verb (Sub-rule 4113)

6. Is a potential subject-block fulcrum present?

This is checked by step (1) (a).

7. Is our potential subject-block fulcrum the fulcrum of a real subject block?

Step (1) (b) establishes the type of search required to check whether our potential fulcrum is a real fulcrum, by referring to step sequences (2), (3), (4), respectively, designed to test for various types of fulcra. Step sequence (2) serves to establish a singular fulcrum, step sequences (3) and (4) serve to establish the possible kinds of plural fulcra.

Referral step (13) is designed to carry this search from the verb to the end of the sentence if required.

8. What additional structures are associated with the subject-block fulcrum established under 7.?

Step sequence (5) - (12), which includes the utilization of Loop II [with Sub-Loop (i)], is designed to cover all the nominal structures contained in the subject block up to and including a possible conjunction or comma, serving as a link to a second fulcrum, or to another clause.

9. Is there a second potential subject-block fulcrum present?

Step (14) checks this.

10. Is the second potential subject-block fulcrum part of our subject block, or is it part of the subject block of a subsequent clause?

Step (15), by searching for an additional verb, establishes whether or not there is present a subsequent clause, to which the subject-block fulcrum found under 9. would have to be assigned, thus establishing a clause boundary within the sentence.

11. If it is part of our subject block, is the potential fulcrum one of its real fulcra?

If step (15) does not find a second clause, then it refers back to step (1) (b) to check whether the potential subject-block fulcrum is a real fulcrum.

12. Is there more left to include within the subject block?

The second go-round of step sequence (5) - (12) will cover all the structures associated with the second fulcrum up to and including the next possible conjunction or comma, then the procedure outlined under Questions 9. - 12. above can be repeated as often as required, in case a third, fourth, etc., fulcra are present.

Commands

1. If the search under A of span preceding the verb establishes the presence of a subject block by finding its fulcrum no rearrangement is required, and the choice of translation will be: no pronoun before the English verb.

2. If the search under A is negative, and only the search under B of span following the verb establishes a subject block, then the choice of English translation will be the same as above. In addition, however, all the constituents of the subject block, having been identified by the search, are now combined into a transfer block and their translations are shifted into the position immediately preceding, the translation of the verb.

3. If the searches under both A and B are negative, the choice of translation will be: pronoun "they" before the English verb.