JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics,
Vol. 66B, No. 2, April–June 1962

Hindsight Technique in Machine Translation of Natural Languages

Ida Rhodes and Franz L. Alt

(December 21, 1961)

In the proposed system for automatic syntactic analysis of Russian sentences developed at the National Bureau of Standards, the computer splits each Russian word into stem and ending and combines the information obtained from these two elements into a morphological description of the word, frequently containing several alternatives. The decision among such alternatives is normally made on the basis of "predictions" arising from preceding words of the same clause. There are, however, cases in which no prediction is available to account for a word, e.g., when the object of a verb occurs before the verb itself. In such a case, instead of the usual prediction of the object, we need "hindsight." Also, it may happen that more than one of the morphological alternatives of a word agrees with predictions; or that a single morphological alternative agrees with several predictions; or that only one of them agrees, yet there is a suspicion that the agreement is spurious; or that no agreement at all is found. It turns out that the alternating use of prediction and hindsight techniques overcomes most of these troubles.

1. Introduction

The present paper may be considered as a progress report on a research project aimed at Russian-to-English translation by means of digital computers, which has been conducted at the National Bureau of Standards since 1959. The project differs from half a dozen others concerned with automatic translation from the Russian in its emphasis on conventional grammar. Some of the other researchers in this area have been concerned with the compilation of automatic dictionaries, with statistical studies of the frequency of occurrence of given words or constructions, with semantic problems, etc.; still others, though concerned with grammatical problems, have stressed the development of new theories of linguistic structure. By contrast, we start from the traditional grammatical concepts, as taught in school, and modify them only occasionally where this seems expedient for computer coding. Thus our first aim is the automatic syntactic analysis ("parsing") of Russian sentences.

The main features of this approach have been described in an earlier paper [1].[1] The terminology and notation of that paper have been retained in the present one, with a few minor changes due to refinements adopted since the former was written. For the convenience of the reader we repeat here a few salient points of that paper, amplified by examples which lead up to the subjects of the present paper.

Reduced to simplest terms, our translation system rests on the following concepts. A Russian word, in general, is capable of several grammatical interpretations. For instance, the Russian word нации may be the genitive, dative or locative singular, or nominative or accusative plural, of the noun нация (nation). For another example, an adjective ending in —ым may be either in the dative plural, any gender, or in the instrumental singular, masculine or neuter gender. We call these alternatives "temporary choices" (TC). Furthermore, a word frequently calls for certain other words or forms of words. For instance, any form of the verb служить (to serve) is likely to be followed by a noun in the dative and/or by a noun in the instrumental case; similarly, a transitive verb predicts an object in the accusative case; many verbs, nouns, or adjectives predict instrumentals; at the start of each clause in a sentence we predict a subject and a predicate. Information of this kind is called "foresight predictions" (FP). The proposed translation scheme proceeds by choosing one among several alternative TC's for a word on the basis of previously recorded FP's; the chosen TC is labeled "selected choice" (SC). (In examining FP's, the most recent one is taken first; and the first agreement encountered is used as SC.) Thus, if the words служит нации occur in a sentence in this order, the latter word will be considered as dative singular, and will be translated accordingly (unless this decision is superseded by information obtained elsewhere in the sentence). In the normal course of events the computer scans a sentence from left to right, one word (or "occurrence") at a time, assigns an SC to each word and, based on it, forms additional FP's which are stored in the machine for use with subsequent words. The left-to-right scanning of a sentence may be iterated if necessary.

A number of complications can arise. For example, the object of a verb may occur before the verb itself; instead of the usual prediction of the object we need "hindsight" to explain its occurrence. In general, the appearance of a word before its prediction is quite frequent. Incorporating this feature into the machine code raises interesting problems. Another complication enters when more than one of the morphological alternatives of a word agrees with predictions; or when only one of them agrees, yet there is a suspicion that the agreement is spurious; or if no agreement at all is found. It turns out that the alternating use of prediction and hindsight techniques overcomes most of these troubles.

[1] Figures in brackets indicate the literature references at the end of this paper.

2. Definitions

We see then that the hindsight techniques are used in certain cases where the usual process of choosing a selected choice (SC) from among the temporary choices (TC) on the basis of foresight predictions (FP) runs into some kind of trouble, and where there is reason to expect that subsequent words of the same sentence will help to explain the situation. In such cases, rather than search for the explanation at once, the machine stores certain information in one of several memory areas set aside for this purpose. This information is called a hindsight entry (HE). Information obtained from subsequent words which sheds partial light on a particular HE is called a hindsight resolution (HR) and is stored alongside the HE. Whenever an occurrence completely resolves a difficulty which has caused an HE, this is called an explanation or decision. When this occurs, the HE and any previous HR's pertaining to it are erased.

We distinguish four types of hindsight, designated H_0, . . ., H_3. Each has its own reserved area in the machine's memory, its own entries, labeled H_0E, etc., and its own resolutions, labeled H_0R, etc.

Briefly and with some oversimplification, H_0 is used when *no* agreement is found between any TC of the current word and any FP on record; H_2 is used when *two* or more agreements are found; H_1 is used when there is an agreement which, however, is considered doubtful and likely to be superseded by information obtained later in the sentence. Finally, H_3 includes all those TC's which have served neither as SC nor as HE. However, in some cases we discard these left-over TC's without even recording them in H_3; namely, when the SC fits so well that there is no doubt about its being the correct choice. Examples of this are given in section 6.

The foregoing definitions of types H_0 to H_3 have to be modified for so-called "unpredictable" occurrences. This will be discussed in the following sections.

Resolutions may be obtained at several stages of the examination of an occurrence during syntactic analysis:

1. After the new foresight predictions are made, yielding H_iR_{FP} (i=0, 1);
2. after the TC's of an occurrence are compared with the FP's and the selected choice is made, yielding H_iR_{SC} (i=0, 1, 2);
3. at the completion of the processing of one occurrence, when the unselected TC's are examined for possible clues, yielding H_1R_{TC}.

Thus, the four types of hindsight differ both in respect to the situations from which they arise, and in respect to the circumstances which can contribute to their resolution. As we have just stated, a hindsight entry of type 0, H_0E, admits resolutions from FP and SC; an H_1E admits resolutions from FP, SC and TC, and H_2E only from SC. No resolutions are recorded for H_3 entries.[2] The reasons for these differences will become clearer in the next few sections.

[2] These entries are used only in subsequent iterations of the process.

It may happen, of course, that an agreement between foresight prediction and temporary choice is established uniquely and — seemingly — without doubt, and therefore no hindsight entry is made, yet the agreement is invalidated by some information occurring later in the sentence. In such a case, a better translation is obtained in a later iteration of the left-to-right scanning of the sentence. Thus, hindsight serves the purpose of reducing the number of iterations. It is an economical expedient, rather than a logical necessity.

In the following we shall discuss some features of the hindsight scheme in more detail. This discussion should be understood as a collection of illustrations, rather than as a complete treatment. It is based on small samples of text, and will undoubtedly undergo considerable revision in the course of time.

3. Hindsights of Type 0

If, in the analysis of a given word of a sentence, we find that none of its TC's satisfy any of the FP's on record, then in general we choose the first TC as SC, and make a record of the situation in H_0E. The listing of TC's in the machine is roughly in order of the frequency with which they occur, so that in the absence of other information the first one is probably the best choice.

Certain TC's are considered unpredictable; for instance, prepositions, conjunctions, adverbs, and certain idioms taking the place of adverbs (incidental expressions, parenthetic remarks.) Similarly, accusatives of nouns designating time, certain instrumentals, and the gerunds of certain verbs are considered unpredictable. If all TC's of one occurrence are unpredictable, we choose the first one but make no entry in H_0. (There would be no point to such an entry, since no resolution is either possible or needed.)

If some TC's of an occurrence are predictable while others are not, the predictable ones are examined first. If no FP matches any of them, the first unpredictable TC is chosen as SC and no entry in H_0 is made. However, the commonest of these cases, such as the ambiguity between adverb and short-form neuter adjective (adjectival stem with ending —o) are entered in H_1.

Inasmuch as an H_0E records the fact that an SC has been chosen without the required FP, it is the purpose of the resolutions to find an FP which explains either the SC or an alternative TC. There are certain kinds of TC which cannot possibly be governed by a subsequent occurrence, and such TC's may be deleted at the time H_0E is made. For instance, a locative case can only be called for by a preceding preposition (or as "master" by a preceding adjective which in turn is governed by a preposition). In the example given in the introduction, the word нации has the TC's: *gen., dat., loc. sing.;* and *nom., acc. plur.;* if this word occurs at a time when there is no prediction of a locative on record, the TC *loc.* may be deleted. See, however, section 7(c).

Subsequent to the recording of an H_0E, every new FP is compared with the TC's of the H_0E and, if satisfied by one of them, is recorded as a "partial resolution for H_0 at time FP," H_0R_{FP}. This is in addition to its recording in the foresight pool as usual. If a subsequent word has a TC which satisfies the same FP, this fact is marked alongside the H_0R_{FP}. This marking is designated H_0R_{SC}.

4. Hindsights of Type 1

As we indicated in section 2, a hindsight entry of Type 1 (H_1E) is made when, in comparing the TC's of one occurrence with the FP's recorded in the foresight pool, an agreement is found, and when this agreement is of a kind considered doubtful. We have a standard list of such doubtful cases of agreement. Examples are:

A word which could be either nominative or accusative chosen as subject.

A verb in the infinitive chosen as subject.

A short-form adjective, neuter singular, which could also be an adverb.

The word и chosen as a conjunction; it could also be an adverb or part of the pair и . . . и.

One of the words его, ее, их, which may be rendered in English either as pronouns or as possessive adjectives.

Since H_1E is based on a match between FP and TC, resolutions must provide alternative explanations both for the TC and for the FP; these can be furnished by subsequent FP's and TC's, respectively. Thus a subsequent FP which is satisfied by the TC of H_1E or by one of its alternate TC's, is recorded as a resolution H_1R_{FP}, while a subsequent TC which satisfies the FP of H_1E is recorded as H_1R_{TC}. In addition, if the FP of an H_1R_{FP} is also satisfied by a later TC, this fact is marked in an H_1R_{SC}. In the case of the FP "Subject" or "Predicate," the fact that subject and predicate must agree in person, gender and number may sometimes eliminate some of the competing matches.

Obviously, a complete resolution is unattainable unless at least one resolution by FP *and* one by TC are recorded.

5. Hindsights of Type 2

Sometimes the process of comparing the TC's of an occurrence with the FP's in the foresight pool results in more than one agreement. We may find one FP satisfied by several TC's, or one TC satisfying several FP's, or different TC's in agreement with different FP's. The last case is the most obvious one. Thus in the example given in the introduction, the form нации has five TC's (*gen., dat., loc. sing.*, and *nom., acc. plur.*). If this word occurs at a time when the foresight pool contains predictions of accusative and genitive complements, these two FP's are satisfied by two different TC's of the present occurrence. An instance of one FP satisfied by two TC's occurs when, say, an adjective ending in —ом, designating the locative singular of either masculine or neuter gender, is encountered at a time when the foresight pool contains a prediction of a complement in the locative case. Two FP's satisfied by the same TC might be predictions of a genitive complement and of an adjectival modifier in the genitive case; this case is discussed under (a) below.

Whenever we have several agreements between FP's and TC's, the first such agreement found is chosen as SC; the others are, in general, entered in H_2E. (Exceptions will be discussed in sec. 6). Numerous combinations of SC and alternative satisfiable FP's are possible. As was done for H_1, these possible combinations will be numbered in accordance with a standard list. At the present time only a few examples of items on this list can be given; other cases will be added to the list as they are encountered in text.

(a) SC=complement, genitive; alternative FP=modifier, genitive. This situation arises when a genitive noun is followed by an adjective which agrees with the noun: нахождение тела максимального объема. The noun тела, in accordance with the rules of grammar, predicts a possible complement in the genitive case, and a possible adjectival modifier. The following adjective максимального satisfies both predictions. In such a case we choose "complement" as SC and record "modifier" in H_2E. The adjective, in turn, predicts a "master," i.e., a noun or other declinable word agreeing in case, number, gender, and animation. (Incidentally, the adjective максимального has a third TC, namely *acc. sing. masc. anim.*; if there is no prediction of an accusative on record, this TC would merely be entered in H_3.) For resolutions to be used with this case, the following empirical rule appears to work. If the "master" prediction is satisfied by a subsequent occurrence, this constitutes a complete decision. Since the adjective has seemingly found a master, we assume it cannot also serve as modifier of the preceding noun. Thus the originally chosen SC is confirmed, and the H_2E is erased. If, on the other hand, no master is found, the situation is reversed and the SC for the adjective is considered to be the postpositive modifier (attribute), unless it is marked as the kind of adjective which is frequently used as a noun. The other alternative choice is left standing in H_2E. Barring the remote possibility that the ambiguity will somehow be resolved by a subsequent occurrence, both versions will be printed out. Note that adjectives used as nouns sometimes require a different English translation. Thus, in исчисление объема последнего окончено ("the computation of the volume of the latter is completed") the adjective последнего is first considered as complement, with the alternative "modifier" stored in H_2. The SC "complement" predicts a master. When no master is found, we consider interchanging SC and H_2E; since, however, последний is often used as a noun and is so marked in the dictionary, we decide to retain the SC "complement," leaving "modifier" in H_2, and assigning the noun meaning to it. As such it is translated "latter," while as an adjective the likely translation is "last."

(b) An adjective which can be genitive singular for masculine or neuter gender, animate or inanimate; or else accusative singular for masculine animate (i.e., adjective ending in —oro or —ero). If such an adjective occurs at a time when there are extant predictions both for accusative and genitive complement, one of the cases (the first one found to give agreement) is chosen as SC, the other goes to H_2E. This entry may be resolved in more than one way. If the adjective is followed by a "master" which is inanimate, then it could not have been the accusative; this would be considered a "decision." If a subsequent occurrence of the same clause, other than the master, is either clearly an accusative or clearly a genitive, it can be used to satisfy one of the two competing predictions, leaving the other one to be assigned to the ambiguous adjective. Such a resolution is entered as H_2R_{SC}. For example: Мы читали на курсах знаменитого учителя старую книгу, изданную в 1753 ("we read in the class of the famous teacher an old book, published in 1753"). By the time we arrive at знаменитого ("famous"), there are predictions of an accusative object, governed by the verb "read," and of a genitive complement, governed by the noun "class." The adjective знаменитого can be either *gen.* or *acc.* The machine assigns the former as SC; the latter is stored in H_2E. Next, this adjective predicts a master. If the following noun were an inanimate genitive, it would indicate that the adjective was also genitive. Since, however, учителя is animate, it can be either genitive or accusative, and thus throws no light on the ambiguity. If the sentence ended after учителя, the SC "*gen.*" and the H_2E "*acc.*" would both stand, and two translations would be printed; "we read in the class of the famous teacher" and "we read in the class the famous teacher." (If instead of читали we had a verb which *must* be accompanied by an accusative object, such as увидеть, the first version would be omitted.) Since, however, the sentence does not end here and the following words старую книгу are clearly accusative, this fact is entered as H_2R_{SC}, to serve as partial explanation. If, as is the case here, the remainder of the sentence throws no further light on the ambiguity, only the genitive version of the translation will be printed.

6. Omission of Hindsight Entries

In a number of cases we deviate from the general rules for recording hindsight entries, laid down in the preceding sections. This is done primarily if the selected choice appears secure beyond any doubt, but also when an alternative, if it were entered in hindsight, could not possibly lead to a resolution.

(a) The SC is undoubted:

(a–1) The SC is a prepositional complement. In this case no entry is made either in H_2 or in H_3. (Of course, items for entry in H_0 or H_1 do not even arise in this case.) In Russian, unlike English, a preposition must be followed by its complement, and nothing can intervene between the two except adverbs or certain incidental expressions. There are, however, cases where we have two FP's of prepositional complements, or where one preposition governs several complements; in such cases Rule (a–1) must be modified. This will be discussed in the next section.

(a–2) The SC is the "master" of a preceding adjective.

(b) No resolution possible or necessary:

(b–1) A TC other than the SC is in the locative case. There is no need to enter this in any hindsight. The reason is that the locative can occur only as a prepositional complement (or as a master of adjectival modifier of another preceding locative). However, the case of multiple prepositional complements, to be discussed below, presents some complications.

(b–2) An unpredictable TC. Cases of unpredictable TC's are enumerated in section 3 above. There is no point in making an H_0 entry, since the only purpose of such an entry is to allow for a subsequent prediction. The case cannot arise for H_2, since this would presuppose an existing prediction. Some frequent ambiguities involving unpredictable TC's are among the standard types of H_1, such as that of the short-adjective/adverb ending in —o. If another TC has been chosen as SC, and if the case is not in H_1, then the unpredictable TC is entered in H_3.

7. Multiple Prepositional Complements

Special treatment is required in certain cases where a preposition can govern different cases, and thus generates several predictions of prepositional complement, or where a preposition governs a list of nouns or other declinables, all in the same case but possibly widely separated.

(a) The case where several predictions of prepositional complements are satisfied by the same occurrence happens most frequently with adjectives. For instance, the adjectival feminine ending —ой can be genitive, dative, instrumental, or locative, singular. Should a preposition such as с require either of a pair of these cases, we should have no way of determining which of the TC's is actually to be taken as the SC. In such a case, we shall choose *both* for the SC, and wait for the master to resolve the situation.

(b) The positional prepositions (requiring either the accusative or locative) may cause even greater trouble because there are some nouns whose endings do not distinguish between the locative singular and the accusative plural. An attempt is made to resolve this difficulty by storing a flag to indicate whether a previous occurrence governs a positional preposition and which of the cases would then be required. For example: the word основанный requires a на with a *locative*, whereas обращать внимание requires на with the accusative. Should either of these locutions be followed by a word, such as теории, which is either locative singular or accusative plural (among other cases), the above signals will indicate which is to be chosen.

(c) Occasionally, a locative may occur in a list of locatives governed by the same preposition (unrepeated). We believe that the profiling scheme [2] would in some cases be able to detect this situation, but we have not yet found a formula which would be successful in all situations. For example: Мы говорили о теории Фадеевой, очень интересной части высшей алгебры. (We were speaking about the theory of Fadeeva, a very interesting part of higher algebra.) The adjective интересной (interesting) is here locative and, together with the noun теории (theory) is governed by the preposition о (about), but it has many other TC's, some of which may agree with extant predictions. During the process of profiling, this adjective will be marked as possible head of an adjectival phrase beginning at the comma, and will be given a "backward flag," a machine indication of the possibility that it may be a modifier of an earlier noun. But this does not enable us to decide whether the adjective интересной agrees with the locative noun теории or with the genitive noun Фадеевой.

We remark parenthetically that examples like the last one can give rise to still further difficulties. The word Фадеевой may be interpreted as an adjective, modifying the preceding noun теории, rather than as a noun serving as genitive complement to теории (translated "the Fadeevian theory" or "the Fadeev theory"). If so, the ambiguity in интересной disappears, since there is now only one noun (теории) which it can modify. Furthermore the noun части (part) may be construed either as "master" of the preceding adjective интересной, agreeing with it in case, number and gender (an interesting part); or as a dative complement to the same adjective ("interesting to a part"). Also, the adjective высшей ("higher") may be construed as modifying either the noun preceding it or the one following it, and this alternative may be combined with any of the uses already enumerated. In addition, the noun части, instead of being construed as either the master or the dative complement of интересной, may be understood as a genitive complement to that preceding noun which also governs интересной. Another possibility is that the noun алгебры is the genitive of comparison after the comparative adjective высшей (giving the translation "higher than algebra"). Finally, there are several other minor types of unlikely translations. One such type involves a dependent genitive construction *preceding* its governor, as in "about the Fadeeva of the theory," "a higher algebra of a very interesting part," "a very much higher algebra of an interesting part," etc. Another such type is concerned with the fact that almost all Russian adjectives can act as nouns, requiring the insertion of the English word "one," and can then take a dependent genitive construction, as in "a higher one of algebra," "the theory of the Fadeevan one," etc. Furthermore, each noun, or adjective used as noun, can be considered to be an appositive to a preceding noun in the same case, yield ng constructions translated as "the theory Fadeeva," "the very interesting one, a part," etc. A small number of the several hundred possible syntactic interpretations of the sentence will be found below. A resolution of some of these ambiguities will be possible only after the formidable problems of semantics have been attacked.

Our scheme will print out the following as the most likely translation: We were speaking about the theory of Fadeeva, a very interesting part of higher algebra.

Unfortunately, our syntactic analysis will be incorrect, even though this is not shown in the translation, because English is not sensitive to case distinctions in nouns and adjectives. Here, our scheme will connect the word интересной appositively with the closest preceding noun which agrees in case with it, namely with Фадеевой, rather than with the correct noun теории.

Some further possible syntactic interpretations of the sentence are given below. Explanatory words and punctuation marks have been added to show the syntactic structure.

We were speaking about:

The theory of Fadeeva, who is a very interesting part of higher algebra.

The Fadeevan theory, which is a very interesting part of higher algebra.

The Fadeevan theory, who is a very interesting part of higher algebra.

The theory of the Fadeevan one, which is very interesting to a part of higher algebra.

The theory of the Fadeevan one, who is very interesting-to-a-part higher one than algebra.

The theory, Fadeeva, which is a very interesting one, a part of a higher one, i.e., of an algebra.

The Fadeevan theory, which is very interesting to a part higher than algebra.

The Fadeevan one of the theory, who is a very interesting higher-than-a-part algebra.

The Fadeeva of the theory, who is an algebra of a higher one of a part of a very interesting one.

The theory of Fadeeva, which is a very interesting one of a part higher than algebra.

8. References

[1] Rhodes, Ida, A new approach to the mechanical syntactic analysis of Russian, MT, Vol. **6**, No. 1.

[2] Alt, Franz L., and Rhodes, Ida, Recognition of clauses and phrases in machine translation of languages, Proc. Internat. Conf. on Machine Translation, Teddington, 1961.

[3] Alt, Franz L., The outlook for machine translation, Proceedings of the Western Joint Computer Conference, Vol. **17**, San Francisco, May 1960.

(Paper 66B2–71)