Linguistic Analysis And Translation Analysis

Paul L. Garvin
Institute of Languages and Linguistics

I have stated my theoretical position on MT in a recent paper in *For Roman Jakobson*.[1] I think that at the present stage of MT no useful purpose would be served by a further elaboration of my general ideas. What I would like to do instead is to discuss in detail some workaday problems of methodology on the basis of experience gained, rather than in terms of ideas as yet untested.

Concretely, I shall present a single Russian example and attempt to show how, and to what extent, certain morphological features and syntactic relations have been utilized to work out a translation procedure which, within its admittedly narrow limits, I consider to be sufficiently rigorous to be pragmatically valid, and yet sufficiently flexible to allow expansion beyond the limits as presently defined.

In doing this, I shall draw upon the procedures and results developed by the Experimental Group of the Georgetown University Project in MT Research[2] with whom I have been working.

In dealing with morphological features, these procedures are based on the assumption, which by now I consider validated, that the translation operation must proceed from the written morphs in the source language directly to the translation in the target language. In this, the assignment of the morph to its appropriate grammatical category is an essential preparatory step, but not one which has to be incorporated into the code by which the translation is effected, unless it is relevant translationally as well as grammatically.

The syntactic relations which have been considered in our procedures so far are those of agreement, government, and dependence. I shall attempt to show the extent to which these relations have been found to be relevant to, and have been utilized for, translation into English, within the limits of the present state of our research.

To define the relations of agreement and government for the present purpose, I shall utilize the traditional concept of grammatical

---

[1] Some Linguistic Problems in Machine Translation in: *For Roman Jakobson* (The Hague, 1956), pp. 180-6.

[2] Consisting of Dorita Lochak, Madeleine Mathiot, and Jane A. Pyne.

category (for Russian, the categories of case, number, gender, and person).

It can then be said that two or more units are in a relation of agreement if they share one or more grammatical categories (that is, agree as to number, gender, case and/or person).

Two units are in a relation of government if, irrespective of the grammatical category of unit A, unit B has to be assignable to a certain grammatical category. Unit A is then said to govern the grammatical category of unit B (thus, a preposition would govern the case of the following noun.)

The relation of dependence is a relation of presupposition: A is dependent on B, if A presupposes B for its occurrence, or for some other functional property.

The translation design under development by the Experimental Group[3] first envisions a sensing and matching routine in which paradigmatic suffixes are separated from the preceding bases whenever translationally convenient; then undivided words, as well as base and suffix partials of subdivided words, are matched against corresponding entries in a stored glossary. The glossary is intended to contain the Russian items to be matched, together with their single or multiple translations, accompanied by a set of code symbols serving to effect the choice from among several alternatives in case of multiple translation, as well as other required translation steps.

The diacritic code thus carries the major translation load, since it is intended to handle those cases in which the translation is not one-to-one, that is, where it is not procedurally trivial. The function of the code is dual: it permits the recognition of decision points— items in the input which, because of possible multiple translation, require translation decisions; and it initiates a set of instructions designed to find the cue for the appropriate decision and to implement it.

---

[3] In this, we follow the basic pattern of the Georgetown-IBM experiment in 1954. (See L. E. Dostert, The Georgetown-IBM Experiment, in: William N. Locke and A. Donald Booth, eds., Machine Translation of Languages [The Technology Press, M.I.T., John Wiley and Sons, New York: Chapman and Hall, Ltd., London, 1955], pp. 124-35; and Jacob Ornstein, Mechanical Translation, New Challenge to Communication, Science 122. 745-8 [October 21, 1955]).

Let me now discuss my example.[4]

пирокатехиновые эфиры триарилметилфосфиновой кислоты

In the type of structure exemplified here, a relation of agreement can be said to exist within each of the two modifier-head constructions: пирокатехпновые agrees with эфиры, триарилметил— фосфиновой agrees with кислоты. A relation of dependence can in addition be said to exist within each of these constructions: the adjective is agreement-dependent on the noun, in the well-known sense that the grammatical categories of the modifying adjective vary with those of the modified nouns, and not vice versa. In addition, a relation of government can be said to exist between the two constructions, in that the first noun construction governs the genitive of the second.

All of these relations are of equal significance from the standpoint of generating Russian utterances, such as is the case in speaking or writing, or translating into, Russian.

For the translation from Russian into English, the general design will have to be implemented as follows:

The sensing-and-matching routine will separate suffixes from bases and match the following items: пирокатвхинов—, — ые, эфир—, —ы, триарилметилфосфинов—, —ой, кислот—, —ы.

Of the above, the following will be translationally unambiguous;

пирокатехинов—   - *pyrocatechol*

—ые   - *zero* (Russian nominative/
                              accusative suffixes require no English translation; since English
                              modifiers do not agree
                              with their heads, the
                              grammatical plurality of
                              the Russian suffix requires no English
                              translation either.)

_____

[4] Taken from the chemical corpus of the Georgetown Project, Журнал Общей Химии, vol. XXII (1952).

триарилметилфосфинов—   . *triarylmethylphosphinic*

The following right partials will require a choice from among several translational possibilities:

—ы, a grammatically ambiguous morph: assignable to a noun suffix for either genitive singular, or nominative or accusative plural;

—ой, a grammatically ambiguous morph: assignable either to an adjective suffix for nominative, genitive, dative, accusative, or instrumental singular, or to a noun suffix for instrumental singular.

Of the two remaining left partials, эфир— and кислот—, the former clearly requires a lexical choice from among two possibilities, *ether* or *ester;* the latter presents no lexical ambiguity: it translates as *acid.* Grammatically, they are both unambiguous: the first is a masculine noun base, the second a feminine noun base.

A grammatical ambiguity is introduced for these two bases, however, by the requirement of rendering the number meaning of the nouns as wholes. It has been found procedurally simpler to assign the number meaning to the translation of the base, and to retain, for the translation of the suffix, only the remaining elements of meaning, in order to reduce the coding and output load of the suffix entries in the glossary.

Thus, the glossary entry for эфир— will, in addition to the lexical ambiguity, contain a procedurally induced grammatical number ambiguity, requiring two sets of choices: *ether* versus *ethers, ester* versus *esters,* and once this choice of number is resolved, the remaining choice between the lexical meanings has to be effected (or vice versa). Likewise, the entry for кислот— will have to contain a choice of *acid* versus *acids.*

On the basis of the above, the right partials of my example will have to be entered into the glossary with following translations:

пирокатехинов—   - *pyrocatechol*

—ые   - *zero*

эфир— - *ether*

- *ethers*

- *ester*

- *esters*

—ы various translations based on the genitival, nominatival, or accusatival meaning of this morph (the corresponding number meaning will affect the translation of the preceding base).

триарилметилфосфинов— - *triarylmethylphosphinic*

—ой various translations based on the nominatival, genitival, datival, accusatival or instrumental meaning of this morph.

кислот— - *acid*

- *acids*

Let me now discuss in series the routines developed to effect the appropriate choice from among several alternatives, where required, and the linguistic considerations that have entered into the routines.

эфир— - *ether*

- *ethers*

- *ester*

- *esters*

I shall first discuss the routine to cover the choice of number translation, since it has greater generality.

This routine is based on a classification of Russian noun bases in terms of the particular set of suffix morphs found to occur after each. The set of morphs particular to each type is the primary cue for the number translation of the base; if the morph unambiguously establishes the singularity or plurality of the translation, no further cueing is required.

эфир— belongs to base type 209, which may be followed by the suffix morphs ∅, -а, -у, -∅ ,-е, -ом; -ы, -ов, -ам, -ы, -ах, -ами. For type 209, any of these morphs constitutes an unambiguous and sufficient number-translation cue; if a type 209 base is followed by one of the morphs -∅ , -а, -у, -∅, -е, -ом it will have to be given a singular translation; if it is followed by one of the morphs —ы, —ов, —ам, —ы, —ах, —ами, it will have to be given a plural translation.

For this type, the following suffix morph will constitute an unambiguous cue to trigger off the selection of the appropriate number translation.

This translation routine is coded as follows:

The left partial эфир— is a decision point (i.e., a point in the text at which a translation decision has to be made and an appropriate routine initiated to effect it). The glossary entries for all decision points will have to be provided with a suitable diacritic which will have to indicate: (a) that this is a decision point, and (b) which routine is to be initiated in order to effect the decision appropriate to this decision point in this or any context. The diacritic chosen for this particular decision point was P209, where *P* stands for "this is a decision point, initiate a routine here", and 209 stands for "the routine required here is that indicated by rule 209".

This is one of a set of singular/plural base-translation rules, numbered 201 up.

Since the cue to this particular decision is always contained in the suffix immediately following the base in question, the rule will be relatively simple to construct, as compared to rules based on more extensive and/or more distant cues.

For purpose of rule 209, therefore, it is enough if the glossary entries for all suffixes that may follow a base of type 209 are provided with a diacritic indicating whether this suffix is a cue for the singular translation or for the plural translation of the base. Specifically this diacritic will have to convey the following information: (a) "this is a cue"—indicated by *C,* (b) "this is a cue for the routine covered by rule 209"—indicated by 209, (c) "this is a cue only if located in the suffix attached to that same base (or, in MT parlance, in the other partial of the same subdivided item)"—indicated by *s,*

(d) "this is a cue to effect a translation by a singular or plural form"—indicated by X or *Y* respectively. Thus, in order to insure the consistent operation of rule 209, suffixes -а, -у, -е, —ом will have to carry a diacritic C209sX (the zero suffixes are handled by a totally different routine not involving subdivision); suffixes —ы, —ов, —ам, —ах, —ами will have to carry a diacritic C209sY. Furthermore, the translations of эфир— *ether* and *ester* will have to be marked *X,* the translations *ethers* and *esters* will have to be marked *Y*.

Rule 209 will then read as follows: when the glossary entry matched against the current input item contains diacritic P209, check other partial of same subdivided item for the presence of either C209sX or C209sY. If C209sX is present, select translation X of current item (that is, *singular);* if C209sY is present, select translation Y of current item (that is, *plural).*

At this stage of the procedure, the glossary entries for эфир— and —ы will look as follows:

эфир-	- *ether* X	P209
	- *ethers* Y	
	- *ester* X	
	- *esters* Y	
—ы various translations as stated above		C209sY

If the routine by rule 209 is effected, the translation of the source item эфир—ы is plural: *ethers, esters,* rather than *ether, ester.*

Once the grammatical translation decision has been made, эфир— still remains an unresolved decision point for the lexical translation decision between *ethers* and *esters*. This decision will require a rule other than 209, since the cues are obviously not the same; a second *P* diacritic is therefore required to indicate that a second routine is to be initiated to effect the appropriate lexical choice. The diacritic chosen for this particular decision point was P1001, where *1001* stands for "the routine required here is that indicated by rule 1001". This is one of a set of lexical rules, numbered 1000 up.

In order to formulate rule 1001, all cards with эфир— were extracted from the lexical file of the Translation Analysis Group of the Georgetown MT Project, each of 93 cards showing one context in which эфир— has been found to occur in our corpus. The contexts were examined, to determine whether adjacent Russian words occurred with sufficient consistency to warrant serving as cues for translation by either *ether/ethers* or *ester/esters,* or neither. It was found that in 70 of the 93 cases the immediately adjacent word constituted a suitable cue, in 20 cases a cue was present within the sentence, but at a distance and in a relation to the current item such that they would require a very complex coding procedure which it was decided to defer to a later stage in the research, and in 3 cases no cue whatsoever could be found within the sentence.

The decision was made to incorporate the cues constituted by the adjacent words (73.5% of the cases) into rule 1001, and to leave the residual cases to be covered by an ambiguous translation. Translations *ester* and *esters* were designated by Б, translation *ether* and *ethers* were designated by Г, ambiguous translations for those cases where cues would not be handled or were not present were designated by Д, yielding the following glossary entry:

эфир—   - *ester* Х Б   Р209

                 - *esters* У Б   Р1001

                 - *ether*   Х Г

                 - *ethers* У Г

                 - *ester-ether* Х Д

                 - *esters-ethers* У Д

Examination of adjacent cues revealed that the translation *ester* or *esters* applies whenever эфир— is preceded by certain adjectives and/or followed by certain nouns or adjective-noun constructions in the genitive, the translation *ether* or *ethers* applies when it is preceded by certain other adjectives and/or followed by certain other nouns or adjective-noun constructions in the genitive. Code diacritics thus had to be devised to be attached to the glossary entries for all potential preceding and following cues.

They would be decision cue diacritics, and thus carry the indication *C*, followed by the rule number *1001*.

The location of the cue would be in either the preceding adjective or in the following noun or adjective (it was found that in adjective-noun constructions constituting cues for this lexical choice, the adjective—which is first in the sequence—was by itself sufficient to cue the decision).

Since, however, the sensing-and-matching routine subdivides all adjectives and all nouns in the genitive (with the reservation stated below) into left and right partials (base and suffix), the location of the cue had to be further specified.

For purpose of simplicity in stating this rule, I am here ignoring the feminine and neuter plural genitives, which have zero suffixes and therefore constitute complete items. The rule as actually formulated by our group does, however, take these forms into account.

In the case of preceding adjectives, the assumption could be made that (unless they were adjectives with nominal function, which would require a totally separate routine anyway) they would, in the overwhelming majority of cases, agree with the following эфир— in case and hence the right partial could be ignored as non-differential. Thus, the cue diacritic would have to be affixed to the glossary entry for the appropriate left partial. In the code of the Experimental Group, this location (left partial of immediately preceding subdivided item) is designated by ℓ-1. This designation would be followed by the code letter indicating that the choice to be made on the basis of this cue is either *ester* and *esters* or *ether* and *ethers*, thus yielding diacritics C1001ℓ-1Б and C1001ℓ-1Г.

Glossary entries for adjective bases like пирокатехинов— or алкилфосфорист— would thus be provided with diacritic C1001ℓ-1Б, and entries for adjectives like этилов— or изомерн— with C1001ℓ-1Г, to yield translations *ester* and *esters, ether* and *ethers,* respectively.

For the following nouns and adjectives as potential cues, both partials would have to be provided with the appropriate diacritics in the glossary, since they function as cues only if they are in the genitive—that is, if the appropriate suffix is present. Thus, two cue locations have to be checked by this routine: left partial of

immediately following subdivided item, and also right partial of immediately following subdivided item. These locations are designated by $\ell+1$ and $r+1$ respectively. This designation would, for the left partial, be followed by the code letter indicating the choice to be made; for the right partial, the location designation would be followed by the code letters indicating both choices (since for either choice, the right partial would have to be constituted by one of the same set of genitive suffixes). The resulting diacritics are C1001ℓ+1Б, C1001 ℓ+1Г, and C1001r+1БГ.

Glossary entries for noun and adjective bases like кислот—, фосфорист— would then be provided with diacritic C1001ℓ+1Б, and entries for noun and adjective bases like хлоргидрин— and иодоватист— with C1001ℓ+1Г, to yield translations *ester* and *esters, ether* and *ethers* respectively, whenever the appropriate diacritic is present in the entry for the accompanying suffix. Thus, entries for all morphs assignable to the genitive would have to carry diacritic C1001r+1БГ.

Rule 1001 will then read as follows: when the glossary entry matched against the current input item contains diacritic P1001, proceed to step 1.

Step 1. Check the entry for the left partial of the immediately preceding subdivided item for the presence of C100lℓ-1Б and check the entries for both right and left partials of the immediately following subdivided item for the presence of both C1001ℓ+1Б in the left partial, and of C1001r+1БГ in the right partial; if either the preceding condition or both following conditions or all three conditions apply, select translation Б of current item (that is, *ester* and *esters);* if not, proceed to step 2.

Step 2. Check the entry for the left partial of the immediately preceding subdivided item for the presence of C1001ℓ-1Г, and check the entries for both left and right partials of the immediately following subdivided item for the presence of both C1001ℓ+1Г in the left partial, and of C1001r+lБГ in the right partial; if either the preceding condition or both following conditions or all three conditions apply, select translation Г of current item (that is, *ether* and *ethers);* if not, proceed to step 3.

Step 3. Select translation Д of current item (that is *ester/ether* and *esters/ethers).*

Step 3 of rule 1001 then covers those cases (26.5%) for which no appropriate cueing was worked out for the present, by putting out a multiple translation.

In our example, the entry for the left partial of the preceding subdivided item, пирокатехинов—, will carry С1001ℓ-1Б, and the entries for the left partial of the following subdivided item, триарилметилфосфинов—, and for the right partial —ой, will carry С1001ℓ+1Б and С1001𝑟+1БГ respectively, and thus, since the previous application of rule 209 has already narrowed down our choice to plurals only *(esters, ethers,* or *esters/ethers),* rule 1001 will now resolve our lexical ambiguity and will unequivocally translate эфир— as *esters.*

— ы *various translations based on the genitival, nominatival, or accusatival meaning of this morph.*

The assignment of the morph —ы to the genitive, nominative, or accusative will affect its translation as follows: if assigned to nominative or accusative, its English translation has been found to be zero in all cases covered so far; if assigned to genitive, its translation will usually be *of* when preceded by a noun, it will be *than* when preceded by a one-word comparative form such as сильнее, and it will be zero in the remaining cases. A sampling of our chemical corpus and of some other Russian text has revealed that, at the present stage of our research, the above information is sufficient to determine the translation of this morph, and that an inspection of the base to which the morph is attached is unnecessary to effect the translation of the morph. Thus, as cues for the translation of this morph will function an immediately preceding noun, an immediately preceding one-word comparative form, or the absence of either.

The rule applicable to the morph —ы is 105, which is one of a set of suffix translation rules numbered 101 up, and the translation routine is coded as follows:

The entry for —ы, the decision point, gets a diacritic P105.

Any noun in the language will, if it precedes a subdivided item the right partial of which is —ы, constitute a cue for its translation by *of.* Therefore, all nouns entered into the glossary will have to

carry an appropriate diacritic. This diacritic will be a cue diacritic for rule 105, hence *C105* will be the initial code symbols.

Other information incorporated in the diacritic will be as follows:

Since some nouns will be entered in the glossary without requiring subdivision through the sensing-and-matching routine, that is, will be entered as complete items, and some nouns will be entered as subdivided items consisting of left-partial base and right-partial suffix, provisions for the location designation of the *C105* diacritic will have to apply to both possibilities. Nouns entered as complete items will have to carry this diacritic, as well as the left-partial (base) entries for nouns entered as subdivided items; in either case, the location is to apply only if the complete or subdivided item immediately precedes the current subdivided item containing —ы. We mark this location by the code designation *cℓ1,* where *c* stands for complete item, *ℓ* for left partial of other subdivided item and *-1* for immediately preceding.

Finally, the diacritic attached to all nouns entered as complete items, and all noun bases entered as left partial, will carry the choice letter *A,* indicating that the translation to be chosen is *of,* The full diacritic will thus be C105cℓ-1A.

Any one-word comparative in the language will, under the same circumstances, constitute a cue for the translation of —ы by *than.* One-word comparatives will be entered in the glossary as complete items, hence the code diacritic will contain *c-1* as its location designation, and it will contain *B* as its choice designation, resulting in the diacritic sequence C105c-1B.

The glossary entry for —ы will now be as follows:

-ы	of	A	P105	C209sY
	than	B		
	-	C		

Rule 105 will then read as follows: when glossary item matched against the current input item contains diacritic P105, proceed to step 1.

Step 1. Check the entry for the immediately preceding complete item, or the entry for the left partial of the immediately preceding

subdivided item, for the presence of C105cℓ-1A; if it is present, select translation A of current right partial item (that is, *of)*, and reverse the order of the translations of the right and left partials of the subdivided item of which the current item is part (this transposition is required, since while a Russian suffix follows its base, the preposition which may translate it will precede the English noun which translates the base); if C105cℓ-1A is not present, proceed to step 2.

Step 2. Check the entry for the immediately preceding complete item for the presence of C105c-1B; if it is present, select translation B (that is, *than)* and reverse the order of the translations of the right and left partials of the subdivided item of which the current item is part (the reasons for the transposition being similar to the above); if C105c-1B is not present, proceed to step 3.

Step 3. Select translation C of current item (that is, *zero).*

In our example the diacritics required for the first and second step of rule 105 are absent, therefore step 3 will be carried out which results in translation *zero.*

   —ой *various translations based on the nominatival, genitival, datival, accusatival, or instrumental meaning of this morph.*

The assignment of this morph to the various case categories listed above depends on the class of the preceding base: if a noun base precedes, the morph will be assignable to the instrumental only; if an adjective base precedes, the morph will be assignable to all of the listed cases. The first step in the translation routine will thus be to determine the class of the preceding base, in order to proceed with a simpler routine if a noun base precedes, or with a more elaborate routine if an adjective base precedes. In our example, the latter is the case.

In this elaborate routine, it was found that the translation of the morph, if assigned to the nominative or accusative, is zero, as for the previously treated morph —ы. In other instances, irrespective of its assignment to a particular case, its translation will usually be *of* when preceded by one group of nouns, thus both in дирекция такой *direction of such* where —ой is genitive and управление такой *management of such* where it is instrumental; it will be *by* when preceded by a second group of nouns and a group of verbs, *with*

when preceded by a third group of nouns and a second group of verbs, *at* when preceded by a fourth group of nouns and a third group of verbs; it will be *than* when preceded by a one-word comparative, *as* when preceded by a fourth group of verbs, *to* when preceded by a fifth group of nouns, a fifth group of verbs, or a noun in the accusative, and finally, when preceded by a sixth small group of verbs, its translation will be *as* when certain nouns follow and *with* when certain other nouns follow. Thus, as cues for the translation of this morph will function various immediately preceding verbs and nouns, an immediately preceding one-word comparative, an immediately preceding noun suffix morph with accusative function (we have not yet coded for accusatives with zero suffix in this context), and in a limited number of instances, an immediately following noun as well.

The rule applicable to the morph —ой is 102, one of the suffix-translation sets numbered 101 up, and the translation routine is coded as follows:

The entry for —ой, the decision point, gets a diacritic P102.

The entry for the preceding base will have to be the cue for the choice between the simple noun-suffix routine, called routine 1, and the more elaborate adjective-suffix routine, called routine 2. A *C102* diacritic will thus have to be attached to all feminine hard noun bases, and another *C102* diacritic will have to be attached to all hard adjective bases, the two sets of bases after which —ой occurs. The location designation for both will be *s,* since we are here dealing with the other partial of the same subdivided item. The full diacritic attached to appropriate noun base entries will be C102s1, to indicate choice of routine 1, the full diacritic attached to adjective base entries will be Cl02s2, to indicate choice of routine 2.

To develop routine 2, which is the one of interest here, appropriate cue diacritics were selected and assigned to entries for the various groups of nouns (left partials and complete items) as well as verbs (complete items), one-word comparatives (complete items), and noun suffix morphs with accusatival function (right partials), in a manner similar to the above, and to the procedure used in developing the translation for morph —ы.

I shall not present all detailed considerations leading to the formulation of rule 102, but confine myself to citing it.

The glossary entry for —ой will now be as follows:

—ой   (routine 2 only)   P 102

        as   A

        with   B

        of   C

        than   D

        by   E

        to   F

        at   G

        --   H

Rule 102 will read as follows:

When the glossary entry matched against the current input item contains diacritic P102, check other partial of same subdivided item for the presence of either C102s1 or C102s2. If C102s1 is present, go into routine 1; if C102s2 is present, go into routine 2.

Routine 2:

Step 1. Check the entry for the preceding complete item for presence of C102c-1AB; if this is present, then check the entry for the left partial of the immediately following subdivided item for the presence of either C102ℓ+1A or C102ℓ+1B; if C102ℓ+1A is present, select translation A *(as)*; if C102ℓ+1B is present, select translation B *(with)*; if these conditions do not apply, proceed to step 2.

Step 2. Check the entry for the immediately preceding complete item, or the entry for the left partial of the immediately preceding subdivided item, for the presence of C102cℓ-1B or C102cℓ-1C or C102cℓ-1E or C102cℓ-1G; if C102cℓ-13 is present, select translation B *(with)*; if C102cℓ-1C is present, select translation C *(of)*; if C102cℓ-1E is present, select translation E *(by)*; if C102cℓ-1G is

present,   select translation G *(at);* if none of the above conditions apply, proceed to step 3.

Step 3. Check the entry for the immediately preceding complete item for the presence of either C102c-1A or C102c-1D; if C102c-1A is present, select translation A *(as);* if C102c-1D is present, select translation D *(than);* if neither is present, proceed to step 4.

Step 4. Check the entry for the immediately preceding complete item, or the entry for either the left or the right partial of the immediately preceding subdivided item, for the presence of C102cfr-1F; if it is present, select translation F *(to);* if it is not present, proceed to step 5.

Steps 1—4. If a translation is selected from among A through G, reverse the order of the translations of the right and left partials of the subdivided item of which the current item is part.

Step 5.   Select translation H (zero).

In our example, step 2 applies, yielding the translation *of.*

кислот—   - *acid*

acids

This noun base entry, unlike the previously discussed эфир—, requires only a number translation routine.

In line with our base classification procedure, кислот— is assigned to type 202, which may be followed by the suffix morphs -а, -ы, -е, -у, -е, -ой; -ы, -Ø, -ам, -ы, -ах, -ами. If a type 202 base is followed by one of the morphs —а, —е, —у, —е, —ой, it will have to be given a singular translation; if it is followed by one of the morphs —Ø, —а, —ах, —ами, it will have to be given a plural translation. Any of these morphs, therefore, constitutes a sufficient number-translation cue. The remaining suffix morph —ы, however, allows both a singular and a plural translation of a type 202 base, and additional cues must therefore be found whenever a type 202 base is followed by —ы.

When a type 202 base is preceded by an adjective, the adjective suffix serves as a cue to resolve the number ambiguity: —ой before

type 202 base followed by —ы will give singular translation for the base, —ые before type 202 base followed by —ы will give plural translation for the base.

When no adjective precedes, the cueing for the number-translation decision can be based on the relation between the number and the case meanings of morph —ы: when the morph has genitive case meaning, the number translation of the type 202 base will be singular; when the morph has nominative or accusative case meaning, the number translation of the type 202 base will be plural. Preceding nouns, prepositions, one-word comparatives, and various particles serving as cues for genitive identification are then utilized to translate the base as singular, absence of these serves to translate the base as plural.

The routine is then coded as follows:

The entry for кислот—, the decision point, gets a diacritic P202.

All nominal suffix morphs of the set found after type 202 bases will be given *C202* diacritics, with location designation *s*, since they are other partials of the same subdivided items. Suffix morphs —а, —е, —у, —ой will be entered with diacritic C202sX, to effect singular translation; suffix morphs —ам, —ах, —ами with C202sY to effect plural translation, and morph —ы with C202sXY to indicate that the number ambiguity remains unresolved (the zero suffix is again handled by a totally different routine not involving subdivision).

For those cases where a type 202 base is followed by morph —ы, the additional cueing is provided as follows:

Suffix morph —ой, which will function as a cue for singular translation whenever it occurs as a right partial of the subdivided item immediately preceding the type 202 base, is assigned a diacritic C202r-1X; suffix morph —ые, which will function as a cue for plural translation whenever it occurs in the same location (unless it is attached to an adjective with nominal function, in which case a special routine is required), is assigned a diacritic C202r-1Y.

Finally, appropriate *C202* diacritics are assigned to the nouns, prepositions, one-word comparatives and particles serving as cues when no utilizable adjective suffix is present.

Rule 202 will then read as follows:

When the glossary entry matched against the current input item contains diacritic P202, proceed to step 1.

Step 1. Check other partial of same subdivided item for the presence of either C202sX or C202sY or C202sXY; if C202sX is present, select translation X *(singular);* if C202sY is present, select translation Y *(plural);* if C202sXY is present, proceed to step 2.

Step 2. Check right partial of preceding subdivided item for the presence of either C202r-1X or C202r-1Y. If C202r-1X is present, select translation X *(singular);* if C202r-1Y is present, select translation Y *(plural);* if neither is present, proceed to step 3.

Steps 3-5 contain the routine for number translations based on the noun, preposition and other cues serving when the preceding steps are inapplicable. They are formulated in a manner similar to the comparable steps of rules 102 and 105.

In our example, step 2 applies, yielding the singular translation *acid.*

Providing the glossary entries for my example with the appropriate diacritics, and applying the rules as formulated above, the sequence пирокатехиновые эфиры триарилметилфосфиновой кислоты will be translated properly by *pyrocatechol esters of triarylmethyl-phosphinic acid.*

In conclusion, I would like to discuss in some detail in what way morphological features and syntactic relations enter into the four grammatical translation routines discussed above: the routines covered by rule 102 and 105 for the translation of suffix morphs —ой and —ы, and the routines covered by rules 202 and 209 for the number translation of noun bases.

It is clear from the above discussion that in translating the suffix morphs, their grammatical identification as to case was obviously taken into account, but the various steps of the routines were not designed to detour around case identification—they lead directly to translation. Cues were classified and provided with diacritics directly in terms of the English translation required, as for instance in

the translation *of* for morph —ой in terms of a preceding noun, irrespective of whether the morph itself represents a genitive or instrumental.

On the other hand, the case function of the suffix morph was specifically taken into account in the final steps of the number-translation routine for bases of type 202, by using cues grammatically applicable to the case relation, and not the number category.

The syntactic relation of agreement between adjective and noun in an adjective-noun construction is utilized by letting the first case suffix morph of the construction serve as the initiator of the proper translation of the case meaning of the entire construction. The reason is that any non-zero translation of a case morph will have to undergo rearrangement in order to precede the translation of the base, and to result in normal English, it will have to precede the translation of the first of the several bases contained in the construction. Subsequent case morphs in the construction will then be programmed to receive zero translation. Thus, rearrangement is held down to the necessary minimum. While the agreement relation itself is here utilized translationally, the agreement-dependence of the adjective and other modifiers on the noun within the construction is ignored.

The relation of government between two adjacent nouns or adjective-noun constructions is utilized extensively for the translation of case-suffix morphs, but without an intervening formally expressed grammatical identification routine, as can be seen in several steps of rules 102 and 105; this same relation, together with the government of noun or noun-adjective construction by a preceding preposition or particle is also utilized in the final steps of rule 202.

Note, however, that these morphological features and syntactic relations are not utilized whenever present, but only when required.

Thus, the nominal suffix property of morph —ы is ignored in rule 105, while the nominal and adjectival suffix properties respectively of morph —ой are utilized to initiate two separate routines in rule 102.

Likewise, the agreement between adjectival suffix morph —ые and nominal suffix morph —ы are ignored in arriving at a plural translation of эфир—, while the agreement between adjectival suffix morph —ой and nominal suffix morph —ы is the basis for step 2 of rule 202 for putting out a singular translation of кислот—

Thus, structural linguistic information is utilized selectively in terms of its relevance to the translational result, which remains the final criterion.