

Session 2: CURRENT RESEARCH

REPORT ON SOME PRINCIPLES OF THE UNIFIED
TRANSFER SYSTEM

Ariadne Lukjanow

C-E-I-R, Inc.

I. Introduction

Several approaches have been employed in machine translation in the course of the past few years. These approaches were either determined by specific objectives or influenced by the background of the research workers. The objectives range from automatic dictionaries to translations with varying degrees of accuracy, readability, and perfection. The background of a researcher can influence his approach to machine translation in three basic ways. One approach may be influenced by machines in such a way that only the development of a new language computer would lead to acceptable results. Another approach may consist of an attempt to simulate human reasoning on a standard computer.

A third approach would be to make machine translation as mechanical and utilitarian as possible, by adapting this attempt to the capabilities of the machine and by clearly defining the relationship between man and machine. Since present-day computers are best suited to repetitive mathematical operations and man is still the better thinker, this last approach will make it possible to utilize both of these capabilities to their fullest extent. All thinking will be expressed in the form of codes in the dictionary in the manner provided for by the system.

In order to translate at all, any system must provide solutions to the problem of transferring structure, function, form, and meaning from the source language into the target language. Thus, we can call translation a fourfold transfer process consisting of:

- (1) Transfer of the function of words (parts of speech)
- (2) Transfer of the form of words (morphology)
- (3) Transfer of the meaning of words (semantics)
- (4) Transfer of the location of words (syntax)

Every word has a meaning, even if there occurs a so-called "zero-translation", or non-translation. In this system, we shall

Session 2: CURRENT RESEARCH

accept a 1:1 translation as equivalent to the non-existence of a meaning problem.

Every word in a language has its function; i.e., it is a part of speech and, unless it is a non-translation item, it also has a location or position (syntax) qualification. The transfer process can be visualized as a combination of the following six concepts:

- (1) Function (some particles, some adverbs)
- (2) Function + location (some punctuation marks, some adverbs, some gerunds)
- (3) Function + form + location (groups from all parts of speech)
- (4) Function + form (some prepositions, some adverbs, some gerunds, negations, etc.)
- (5) Function + form + meaning + location (groups from every part of speech)
- (6) Function + meaning + location (some adverbs, some conjunctions, etc.)

Example:

Combination of function and location:

posle - "later"; adverb with a 1:1 translation

equivalent and location after verb,

colon - punctuation mark; 1:1 equivalent, position is
at the end of a clause.

It is obvious that the elements of the transfer form sets with variants in each of the elements. We can visualize them as shown on the following page.

Session 2: CURRENT RESEARCH

Function	Form	Meaning	Location
x	0	0	0
0	x	0	0
0	0	x	0
0	0	0	x
x	x	0	0
x	0	x	0
x	0	0	x
0	x	x	0
0	x	0	x
0	0	x	x
x	x	x	0
x	x	0	x
x	0	x	x
0	x	x	x

0 - non-variant or absent

x - variant

It would seem that these variations could be expressed in mathematical formulae, but this is not true because the relationship between the variants does not follow the rules of permutation or random combinations. In contrast, these variations follow definite linguistic rules which permit only certain variants within certain combinations. In order to determine these linguistic combinations for the elements of the transfer, it is necessary to define and classify each variant for every element of the transfer, as well as the relationship between the variants of each element of the transfer and the variants of the other three.

This can best be illustrated on prepositions:

Session 2: CURRENT RESEARCH

Element of Transfer	Definition
function	preposition
form	case government; i. e. .prepositions demanding the genitive, dative, accusative, instrumental, or locative
meaning	prepositions of time (static, earlier, later), location or space (where, to where, from where), cause, goal, substitution, division, etc.
location	first item in prepositional phrase, or position 1 in prepositional phrase

Theoretically, we could produce a transfer combination of preposition dative location (from where?) plus first position of prepositional phrase, but the grammatical rules and semantic connections do not permit this type of combination. The prepositions of location are subject to the following division only:

Location	Genitive	Dative	Accusative	Instrumental	Locative
a) where?	bliz vne mezhdu sredi u	po		za mezhdu nad pered pod	v na pri
b) to where?	do	k	v za na pod skvoz6 cherez		o
c) from where?	iz iz-za iz-pod ot s				

Session 2: CURRENT RESEARCH

The above table shows that the "from where?" definition is used only with the genitive case. Thus, the only usable and meaningful combination is:

preposition + genitive + location (from where?) +
first position of prepositional phrase.

In the Unified Transfer System we accept any meaningful and valid combination of elements of the transfer expressed in the form of numerical digits as a single unified transfer code.

Since many words of the source language can be associated with several function, form, meaning, and location qualifications, it is necessary to combine single transfer-code units into sets of codes which can express these variations.

Examples:

<u>dannye</u>	nominal modifier
<u>vdol6</u>	preposition of genitive adverb
<u>s</u>	preposition of - genitive - accusative - instrumental
<u>sredi</u>	preposition of - location (where?) - time (static)

If we consider that we have four elements in the transfer, each of which has a definite and limited number of variants, it is safe to assume that the number of transfer codes is limited and that we may likewise assume that the same applies to sets of transfer codes. This leads us to the concept that numerous words in the dictionary are associated with identical transfer codes or identical sets of transfer codes. This fact makes possible the concept of code patterns. The number of single transfer-code units in the pattern can vary from one to several. After examining some 50, 000 canonical entries (stems) in the dictionary of Smirnitskij, we have decided to set the limit at a maximum of 25 single code units in the pattern.

Now let us examine the actual elements of each transfer. Since in translation we are dealing with at least two languages simultaneously, we have to develop a criterion for parts of speech, morphology, semantics, and syntax which would accommodate both languages under consideration, or we must establish a classification system which in the form of transfer codes would permit us to place an equal sign

Session 2: CURRENT RESEARCH

between the two languages. This necessitates a certain type of analysis and of synthesis of the grammars of both languages.

II. The Function of Words or the Categorization of Word Behavior

When examining conventional parts of speech in Russian and English grammars separately, we note that they contain identical categories such as prepositions, adverbs, nominals, modifiers, etc. But when we compare these categories of both languages, we discover that they differ considerably in usage, behavior, and function. In terms of a translation system, this means that either we have to introduce new synthetic categories or we have to divide and redistribute words differently within these categories. Categorizing is, of course, a somewhat subjective process. That can best be illustrated by examining the English preposition "to" in the following manner:

Qualification	English	Russian Equivalents	Bilingual Data	Transfer data (Classification)
Function	1. preposition	1. preposition	1. preposition-like item	1. preposition code
Behavior	2. introducer of infinitive	2. non-existent	2. particle- like item	2. particle code

Obviously, the second category in the above table might as well be classified as a special auxiliary verb (instead of particle), but to the author of the system the definition as particle appears more reasonable, perhaps because of the occurrence of the Russian particle *by* in the verbal phrase.

In the process of comparative analysis-synthesis, we have established the following basic categories as transfer parts of speech (listed alphabetically):

- (1) adjectival modifier
- (2) adjective/noun
- (3) adverb (including some gerunds and the particle *li*)
- (4) adverbial modifier (type: *bolee*, *menee*, etc.)
- (5) auxiliary verb (*byl*, *byli*, etc.)
- (6) auxiliary verb (*moch6*, *khotet6*, etc.)

Session 2: CURRENT RESEARCH

- (7) conjunction
- (8) negation (including some negative adverbs)
- (9) nominal (animate), including some pronouns
- (10) nominal (inanimate), including some pronouns and numerals
- (11) nominal (formulae, cardinal numbers, missing words)
- (12) numerical modifier
- (13) particle
- (14) participial modifier
- (15) preposition
- (16) pronominal modifier
- (17) pronoun (type: nami, vami, imi, etc.)
- (18) pronoun (soboi)
- (19) punctuation marks (each treated as a separate category, a total of six)
- (20) verb (including participles such as izucheny, otkryty, etc.)

The assignment of these basic categories to individual words is a discrete and subjective process. It can give valid results only if all other factors and constituent parts of the transfer are being taken into consideration. We proceed from the parts of speech as categories to their classification. That can be expressed in the form of a numeric code.

We know that sentences and phrases are combinations of these categories and that these combinations cannot be produced by a random distribution of words. Words have to occupy certain positions in order to form a meaningful combination or phrase.

If we take the three-word phrase "in this room", we cannot convey the same idea by a redistribution of the participating words:

"this in room"
"this room in"
"room in this"
"room this in"
"in room this"

We will either get a meaningless jumble of words or convey a different idea. We say "our new building", but not "new our building". We place some adverbs before verbs, some after them. Some of these phenomena can be explained, some are ascribed to usage, but others escape any logical explanation.

Session 2: CURRENT RESEARCH

Dealing with 26 categories and considering each of them in relation to the other 25, we can establish a hierarchy within the meaningful combinations of parts of speech, i.e., logical sequences.

This point can be illustrated by the position of words within the sequence of a prepositional phrase consisting of a preposition (P), a nominal (N), two adjectival modifiers (AM), and a pronominal modifier (PM):

P before N
AM before N
PM before N
PM before AN
P before PM
P before AM
AM = AM

Thus, we arrive at P-PM-AM-AM-N; or if we assign numerical values to these categories and would like them to form a progression of i_1, i_2, i_3, \dots , we will emerge with the following correlations:

$P < N$; $AM < N$; $PM < N$; $PM < AM$; $P < PM$; $P < AM$; $AM = AM$:

e.g., $P < PM < AM = AM < N$.

Approaching our categories of parts of speech with these criteria, we can assign numerical values or codes to parts of speech (all codes are in octal notation):

01	comma
02	conjunction
16	preposition
17	adverb
20	negation
21	participial modifier
22	pronoun (<u>nami</u> , <u>vami</u> , etc.)
23	auxiliary verb (<u>byl</u> , <u>bylo</u> , etc.)
24	auxiliary verb (<u>moch6</u> , <u>khotet6</u> , etc.)
25	particle
26	verb
27	pronoun (<u>soboj</u>)
37	adverbial modifier
45	pronominal modifier
46	numerical modifier

Session 2: CURRENT RESEARCH

47	adjectival modifier
55	adjective/noun
65	nominal (animate)
66	nominal (inanimate)
67	nominal (formulae, numbers, missing words)
70-77	punctuation marks (colon, semicolon, dash, period, etc.)

We are fully aware that this progression method for the identification of a phrase or logical sequence is reliable only in so-called normal sequences. Interrupted sequences or inverted word order require additional re-examination and even actual recognition of constituent parts of sequences. In such cases specific instructions are necessary.

We have, however, established the fact that more than 80%¹ of sequences are so-called normal sequences. That frees us of the necessity to recognize at all times every constituent part of all sequences, as well as of every possible combination of the constituent parts.

The sequences established through progression codes are by no means permanent or final divisions within the sentence. They can become smaller or disintegrate into single items either through the demands of other components of the transfer process, or through so-called verification instructions.

Example:

A sequence ending with code 47 (adjectival modifier) will call for verification instructions of:

- (1) a sequence within a sequence;
- (2) a sequence with homogeneous parts of speech plus conjunction and/or comma; etc.

We can therefore state that progression codes divide sentences into working units which may or may not become final sequences or phrases. This once more confirms the idea of a total or unified transfer versus a single transfer concept on a different level within the limitations of each phrase (structural, morphological, semantic).

The division into sequence is made in accordance with:

¹ Between 20, 000 and 25, 000 words in various fields of knowledge have been examined for this purpose.

Session 2: CURRENT RESEARCH

A_1, \dots, A_n = part-of-speech code
 B_1, \dots, B_n = part-of-speech code with 1st digit being 6
 C_1, \dots, C_n = part-of-speech code with value of 10

STOPS

A_1 A_n
 B_1 B_n even if $B_1 = B_n$
 A_1 $C_1 \hat{C}_2 A_n A_1 = A_n$

From: JOURNAL OF CHEMICAL INDUSTRY, vol. 22, no. 9
 (1952):

Izucheny reaktsii mezhdou ehfilovym ehfirom pirokatekhinfosforistoj kisloty i triarilbrommetanami.

Pri vzaimodejstvii ukazannykh soedinenij obrazuiutsia pirokatekhinovye ehfiry triarilmetilfosfinovykh kislot.

Pri omylenii poslednikh slaboj solianoj kisloty polucheny pirokatekhin i triarilmetilfosfinovykh kisloty.

V nastoiashem issledovanii nami izuchalis reaktsii mezhdou smeshannymi ehfirami fosforistoj kisloty, tipa $A_2 \dots$ i triarilbrommetanami.

Reaktsiia mezhdou ehfilpirokatekhinovym ehfirom fosforistoj kisloty i triarilbrommetanami po analogii s alkilfosforistymi ehfirami dolzhna idti po reaktsii: $A_2 \dots$

Ehksperimentalnye dannye pokazali, chto reaktsiia dejstvitelno protekaet po ukazannomu uravneniiu.

Tak, naprimer, pri nagrevanii smesi triarilbrommetana i ehfilpirokatekhinovogo ehfira fosforistoj kisloty proiskhodit vydelenie bromistogo ehfila i obrazovanie kristallicheskogo veschestva predstavliaiushego soboj pirokatekhinovyy ehfir triarilmetilfosfinovoj kisloty.

Dlia ustanovleniia stroeniia poluchennogo soedineniia byla provedena reaktsiia omyleniia razbavlennoj solianoj kisloty pri nagrevanii ot 180 do 200° v zapaiannykh trubkakh.

Produktom omyleniia iavliaiutsia pirokatekhin i triarilmetilfosfinovaia kislota.

Poluchennye nami ehfiry tipa $A_2 \dots$ vesoma ustojchivy k vlage vozdukha.

Session 2: CURRENT RESEARCH

(As Translated for the Revised English Text)

Concerning the action of triarylbromomethanes on alkylpyrocatechol esters of phosphorous acid.

Reactions between the ethyl ester of pyrocatechol-phosphorous acid and triarylbromomethanes were studied.

(Up)on the interaction of the above-mentioned compounds, pyrocatechol esters of triarylmethylphosphinic acids are formed.

(Up)on hydrolysis of the latter with dilute hydrochloric acid, pyrocatechol and triarylmethylphosphinic acids were obtained.

In the present investigation, the reactions between mixed esters of phosphorous acid of the type . . . and triarylbromomethanes were studied [by us] .

The reaction between the ethylpyrocatechol ester of phosphorous acid and triarylbromomethanes should proceed, by analogy with alkylphosphorous esters, according to the reaction: . . .

Experimental data showed that the reaction actually proceeds according to the above-mentioned equation.

Thus, for example, upon heating of a mixture of triarylbromomethane and the ethylpyrocatechol ester of phosphorous acid, evolution of ethyl bromide occurs and (there occurs) the formation of a crystalline substance which is the pyrocatechol ester of triarylmethylphosphinic acid.

In order to establish the structure of the compound obtained, a (reaction of) hydrolysis with dilute hydrochloric acid was carried out on heating (at) from 180° to 200° in sealed tubes.

The product(s) of hydrolysis are pyrocatechol and triarylmethylphosphinic acid.

The esters obtained by us of the type . . . are extremely resistant to the moisture of the air.

Examining the preceding Russian text sample in terms of the progression code, we find the following:

Session 2: CURRENT RESEARCH

Sentence Number	
1	26-66 16-47-66 47-66 02-66-77
2	16-66 21-66 26-47-66 47-66-77
3	16-66 55 47-47-66 27-66 02-47-66-77
4	16-55-66 22-26-66 16-21-66 47-66 01-67 67 01-66-77
5	66 16-47-66 47-66 02-66 16-66 16-47-66 23-26 16-66-70 66 77
6	47-55 26 01-02-66 17-26 16-21-66-77
7	02-01-17 01-16-66 66 66 02-47-66 47-66 26-66 47-66 02-66 47-66-77 01-21-27-47-66-47-66 77
8	16-66 66 21-66 23-26-66 66 21-47-66 16-66 16-67 16-67 16-21-66-77
9	66 66 26-66 02-47-66-77
10	21-22-66 67 67 17-26 16-66 66-77

It has been found convenient to make the part-of-speech code part of the pattern number, so that we can determine the possible logical sequence or wording area immediately after the dictionary lookup.

III. Form of Words (Morphology)

With the part-of-speech codes, we have devised the means to divide the sentence into possible structural (constituents), sequential (progressive), meaningful combinations, i.e., phrases or fractions of the sentence.

The next step would be to establish in which way the constituents of the sequence depend on each other, and what demands they place on each other, if any (i.e., either to confirm the sequence or to

Session 2: CURRENT RESEARCH

divide the original sequence into smaller sequences or even single items).

The morphological criteria we are using for this purpose are case, gender, number, and absence of these. For the sake of convenience, we define the demands of government, agreement, and influence as agreement in case, gender, and number.

Numerical values used are:

- (1) Agreement in case #1-7
- (2) Agreement in gender #1-3
- (3) Agreement in number #1-2
- (4) No agreement necessary 0

Note: Case #7 represents the usage as per example in V riadu, sadu, na lugu etc.

v - preposition, accusative, locative

riadu - nominal in dative

Despite the disagreement in case, it is a meaningful combination in which words belong together or form a valid sequence.

Positions of the morphological three-digit codes are as follows:

case	gender	number
------	--------	--------

Digits representing case agreement:

- 0 - no case
- 1 - nominative
- 2 - genitive
- 3 - dative
- 4 - accusative
- 5 - instrumental
- 6 - locative
- 7 - auxiliary

Session 2: CURRENT RESEARCH

Digits representing gender agreement:

- 0 - no gender
- 1 - masculine
- 2 - feminine
- 3 - neuter

Digits representing number agreement:

- 0 - singular or no number
- 1 - plural
- 2 - number disagreement (used in impersonal verbs, etc.)

All these morphological qualifications can occur singly or in combinations.

Session 2: CURRENT RESEARCH

Table of Possible Morphological Codes

000	010	020	030
100	110	120	130
200	210	220	230
300	310	320	330
400	410	420	430
500	510	520	530
600	610	620	630
700	710	720	730
001	011	021	031
101	111	121	131
201	211	221	231
301	311	321	331
401	411	421	431
501	511	521	531
601	611	621	631
701	711	721	731
002	012	022	032
102	112	122	132
202	212	222	232
302	312	322	332
402	412	422	432
502	512	522	532
602	612	622	632
702	712	722	732

Session 2: CURRENT RESEARCH

If we consider now the relationships possible between the concepts expressed in the part-of-speech code and the units of the morphological codes, we can establish combination sets of codes, i. e. , the partial code patterns.

In this report, we shall do so for one part of speech--the prepositions.

Prepositions, as we know, do not demand agreement in number or gender. Therefore, we are dealing with only a case agreement.

Table of Prepositions

Cases	1-Case Prepositions	2-Case Prepositions	3-Case Prepositions	2-Case & Auxiliary Prepositions
2 (genitive)	bez, bliz, vdol', vmesto, vne, vnutri, vozle, vokrug, dlia, do, iz, iz-za, iz-pod, krome, mimo, nakanune, okolo, ot, posle, posredi, protiv, radi, sredi, u	mezhdu (mezh)	s	
3 (dative)	k, blagodaria, vopreki, podobno, soglasno, naperekor, navstrechu		po	
4 (accusative)	pro, skvoz', cherez	v, na, za, pod, o(ob)	s, po	v, na
5 (instrumental)	nad, pered	za, pod, mezhdu (mezh)	s	
6 (locative)	pri	v, na, o(ob)	po	v, na
7 (auxiliary)				v, na

Session 2: CURRENT RESEARCH

On the basis of this table, we can say that some prepositions (code 16 or 17) can be associated with one, two, or three morphological units. The total code patterns will be then as follows:

1-unit Patterns	2-unit Patterns	3-unit Patterns
16 - 200	16 - 200 - 500	16 - 200 - 400 - 500
	16 - 400 - 600	
16 - 300		16 - 300 - 400 - 600
	16 - 400 - 500	
16 - 400	17 - 200 - 000	16 - 400 - 600 - 700
	17 - 300 - 000	
16 - 500		
16 - 600	17 - 400 - 000	

In this fashion, the 49 prepositions of the table of prepositions are associated with 14 code patterns which would accomplish function and form transfer.

IV. Meaning of Words and Meaning Classes

The analysis of languages in establishing meaning categories is of subjective character and is based on a mental process which requires not only an intimate knowledge of the languages to be analyzed but also a very careful manipulation of the numbering system in order to prevent an unintentional conflict of meaning classes in the code patterns within sequences.

In this report we shall attempt to establish some of the criteria of analysis and the nature of the classification of semantic or meaning definitions.

Session 2: CURRENT RESEARCH

In this area we have to make a distinction between situations which can be described as "a word by itself" and "a word in different environments". There are distinctly two levels of meaning ambiguity: (1) a so-called subject-matter ambiguity which fits into the category of "a word by itself", and (2) an environmental ambiguity, i. e. , "a word in different environments".

Example of the first type of ambiguity:

- | | | |
|-------------|---|---|
| akt | = | 1. "act" |
| | | 2. "legal deed" (law) |
| | | 3. "convocation" (education) |
| obrazovanie | = | 1. "education" (education) |
| | | 2. "formation" (technical subject matter) |
| "board" | = | 1. "piece of wood" |
| | | 2. "food" (household arrangement) |
| | | 3. "stage" (theater) |
| | | 4. "council" (political science) |
| | | 5. an action, as in "to board a train" |

The subject-matter ambiguity can be solved sometimes through the environment; for instance, if we encounter the word obrazovanie with modifiers like kristallicheskoe (crystal), kislotnoe (acid), etc. , there is no doubt that the meaning of this word is "formation". Outside of environmental influences, we have to depend on the subject matter of the article or book to be translated, i. e. , microglossary, and use that as a cue for selection. On the level of subject-matter meaning ambiguity, we have to expect a certain amount of inaccuracy or errors, which will have to be corrected by hand.

The second level of meaning ambiguity, the environmental ambiguity, is subject to meaning categorization or classes.

We shall describe the method of arriving at these classes, as well as some class definitions, through the examination of environment relationships in preposition-nominal sequences.

By their meaning connotations, prepositions can be divided into a variety of groups. We shall list some of them here.

Session 2: CURRENT RESEARCH

(1) Prepositions of time

Simultaneous/ Static (when)	Earlier (before when)	Later (after when)
v, za, na, po, pri, s, sredi	do, k, za, pered	ot, po, s, cherez

(2) Prepositions of space and location

where	where to	where from
bliz, v, vne, za, na, nad, mezhd (mezh), pered, po, pod, pri, sredi, u	do, v, k, za, na, o (ob), pod, skvoz', cherez	iz, iz-za, iz-pod, ot, s

(3) Prepositions of cause

for whom, for what, why, etc.
za, iz, iz-za, ot, po, s

(4) Prepositions of goal

dlia, do, v, za, k, na, po, radi

(5) Part of the whole

iz, po

(6) Exchange or replacement

za, vmesto

(7) What is it made from

iz

Now let us consider the category of space or location connotation. We have already divided this category into three sub-categories: (1) The first implies the specific position of something, generally recognized by yielding an answer to the question, "Where?" It implies a point of location. (2) The second sub-category implies the concept of something proceeding towards a certain location, generally recognized by yielding the answer to the question "To where?" It implies a point of destination. (3) The third sub-category expresses the idea of something coming from a certain location, generally recognized by yielding the answer to the question, "From where?" It implies a point of origin.

These sub-categories in turn can be divided further by analyzing specific prepositions.

The prepositions s, iz, and iz-pod, all belong to the "From where?" or point-of-origin class. They differ in their individual semantic content. When s is used, it designates either the place from which the object is removed by some agent, or the place from which an object capable of locomotion is leaving. This last instance usually involves geographic locations in connection with persons or modes of transportation of persons. When iz is used, it designates an object leaving by any means any location that has an egress, or the emergence of an object from another object. When iz-pod is used, it designates an emergence in any manner from under something on the part of an object.

Examples: s Kavkaza, s gory, s sobraniia, etc.
iz goroda, iz derevni, iz avtomobiia, etc.
iz-pod kamnia, iz-pod stola, iz-pod knigi, etc.

Session 2; CURRENT RESEARCH

The prepositions do, k, cherez, all belonging to the "To where?" or point-of-destination sub-category, again differ in their individual meaning content. When do is used, it designates the direction of movement with the definite connotation of limitation or boundary. When k is used, it again designates the direction of the movement, but its definite connotation is to achieve only proximity to the destination. When cherez is used, it designates a penetrating movement through some medium, usually with some difficulty attached to it, and it also designates a movement of directly surmounting a difficult medium.

Examples: do goroda, do Washingtona, etc.

k beregu, k gorodu, k reke, etc.

cherez les, cherez bar'er, etc.

The prepositions u, bliz', and pri, all belonging to the "Where?" or specific location sub-category, differ in their individual semantic content. When pri is used, it designates that one object is adjoining another one. On the other hand, u indicates immediate closeness of objects; bliz', in turn, indicates only closeness of objects.

Examples: u reki, u berega,

bliz' goroda, pri stantsii, etc.

In most of these instances, the translation of the prepositions is at variance with their literal meaning (1:1 equivalent).

Sub-Categories	Prepositions	Literal Meanings	Special Meanings
From where	s iz iz-pod	with from from under	from from from under
To where	do k cherez	to to through	to toward through/over
Where	u bliz' pri	at near at	by near at

Session 2: CURRENT RESEARCH

All these sub-categories, in turn, have to be divided again into smaller groups. For example: iz-pod in relation to location-objects does definitely mean "from under", but with location-cities the meaning of iz-pod becomes "from the vicinity of".

Example: iz-pod stola = "from under the table"
iz-pod Washingtona = "from the vicinity of Washington"

The prepositions so far have been analyzed for their special or locational relationships. The same prepositions can also be analyzed with respect to other semantic criteria.

For example, the preposition iz with the connotation of selection will in some instances keep the translation "from", but in the following environments:

- (a) before plural pronouns: nikh, vsekh, nas, tekh, etc.
- (b) before numerals: dvukh, trekh, etc.
- (c) before collective nominals like: chlenov, predstavitelej, iuristov, etc.

iz will become the preposition of selection, that is, one of many or part of the whole with the translation "of".

Example: luchshij iz vsekh = "best of all"
komitet iz predstavitelej = "committee of representatives"

Let us follow through the analysis of the same prepositions with respect to a time-relation concept. In this case we will find that three sub-categories become apparent: (1) The first implies that an action or state of being occurs after a fixed time span. These prepositions are: ot, s, cherez, po. (2) The second sub-category connotes that an action or state of being occurs before a fixed time span. These prepositions are: do, k, iz, pered. (3) The third sub-category implies that the action or state of being occurs during a fixed time span. These prepositions are: sredi, po, v, s, za, na, pri. Therefore, we can now draw an analogy with the three previously examined sub-classes and can call these time sub-categories "after when", "before when", and "when".

It becomes apparent at this point that some prepositions occurring in these new sub-categories have participated in the previous ones.

Of special interest in analyzing these prepositions are some that coincide with the sub-classes which were previously established:

Session 2: CURRENT RESEARCH

Sub-Categories	Prepositions	Literal Meanings	Where Class	When Class
From-after	s	with	from	since
To-before	k	to	toward	toward
Where-when	pri	at	at	during

For large-scale translation it is necessary not only to apply these larger categories and their sub- and sub-sub-categories, but also to analyze them in terms of each other in order to establish the similarities, as well as conflicts, and then establish final definite categories. This task has been accomplished in the United Transfer System and the precise description of each category will be included in the projected United Transfer System Manual. In this report, owing to the limitations of time and the size of the report, we shall limit ourselves to the method of arriving at the categories, rather than categorization itself.

Let us continue the analysis of the three prepositions s, k, and pri, limiting ourselves to the "where" and "when" categories.

From initial inspection it would appear necessary to assign at least 10 meaning classes, 1 for each concept of the 3 prepositions. It becomes apparent that the number of meaning classes then would increase geometrically with the increasing number of preposition-participants and/or the introduction of new categories. We therefore begin to search for means of reducing the progression. The cues for this reduction come from two basic sources; we can estimate that morphology provides about 70% of the cues by imposing the case restrictions, and that the usage of language should provide the remaining 30%.

In establishing meaning classes, we can combine several concepts into a single class, as long as there can be no conflict at the morphological or usage levels. The idea is to bring together the three elements of transfer (function, form, and meaning) and to reduce

Session 2: CURRENT RESEARCH

the number of code patterns to a minimum. Thus, the next step would be to establish what part morphology plays in reducing the number of meaning classes in the sample prepositions.

This can be illustrated in the following table.

Preposition	Class	Cases				
		genitive	dative	accusative	instrumental	locative
s	1	from		for	with	
	2	from				
	3	since				
k	1	to				
	2	toward				
	3	toward				
pri	1					at
	2					at
	3					during

Note: 1 = literal; 2 = "where" case; 3 = "when" case.

From the table above we can see that for the preposition s only one morphological category is affected; therefore, we need not assign these particular meaning classes for the accusative and instrumental cases and thus achieve a reduction from the original possible 10 down to 6 meaning classes.

Next we note that the meaning classes for each preposition of the sample belong into separate cases, and this would permit us to assign only two meaning classes, a "where" and a "when" class, coded up as entries in the proper case for the respective preposition.

If we now examine these sample prepositions further, we will find that apparently in the case of the preposition s we would not be able to reduce the number of meaning classes any further by virtue of the fact that the "where" translation differs from the "when" translation. On the other hand, k would not present any problem since both translations are identical, while pri again presents us with separate translations for the "where" and "when" cases; but the identical translation of the literal meaning and the "where" case would permit us to eliminate the "where" class in this instance. Sometimes, when the translation of prepositions with the same meaning-definition

Session 2: CURRENT RESEARCH

changes several times, we find it practical to give the preposition a so-called "zero translation" and to attach the translation of the preposition to the nominals. Thus, we again achieve a reduction in the number of meaning classes. The same type of classification is applied to the remaining members of the prepositional phrase (nominals, modifiers) and the classes are assigned within the boundaries of function-form criteria. The definitions of meaning have to be used carefully and discretely, since the criterion of time element is not necessarily derived from strict time terms but can be arrived at in combinations of prepositions with nominals of action.

Examples: pri okislenii = "during the oxidation"
pri rabote = "during the work"
pri issledovanii = "during the investigation"

In assigning these classes to nouns it becomes apparent that these concepts are not rigid rules but are the result of subjective judgment. It is impossible, for instance, to say that all nouns of action in the locative case will yield the same translation "during" for the preposition pri. There are vagaries of usage which defy any definition. For instance, if zhelanii, a noun of action in the locative (desire, wish) occurs with the preposition pri, the translation of the preposition changes to "if" and the meaning connotation from time to condition.

Example: pri zhelanii = "if desired"

A final consideration in assigning meaning classes must be an expression of the physical location of the participating members (words), i.e., whether they must be immediately adjacent, or whether they can be separated by non-participating words; in other words, whether the preposition-nominal relationship is dependent on their immediately adjacent physical location.

Example: pri tshchatel'nom issledovanii = "during careful investigation"
pri issledovanii = "during investigation"
versus: na drugoi den6 = "next day"
na den6 = "for a day"

Thus the indication of the boundary, known in the system as boundary indicator or item count, is introduced together with some of the meaning classes. If no boundary is necessary, this indicator is

Session 2: CURRENT RESEARCH

coded as zero; otherwise it corresponds to the number of participating items, e.g. , four participant-members (words) require digit 4 as a boundary indicator.

Thus, after complete analysis, the code pattern for preposition with:

(a) one literal translation

(b) P + N = adverb

would look as follows:

Pattern #1 = 16-0112-200 = 0 translation (adverb class)

-0111-200 = literal meaning (1:1 equivalent)

This brings together the function-meaning-form transfer categories.

Example:

Preposition <u>bez</u> :	Nominals: <u>interesa</u>	<u>pol'zy</u>
literal meaning =	<u>somneniia</u> ,	<u>nuzhdy</u>
"without"	etc.	etc.

bez interesa = "without interest"

bez somneniia = "without doubt"

bez pol'zy = "useless"

bez nuzhdy = "needless"

V. Location, Arrangement, or Syntax

Once the function-form-meaning transfer has been achieved in the code, the next step in translation is the transfer of structure, Which we can visualize as consisting of two sub-transfers:

- (1) actual structure transfer;
- (2) pure relocation of items within the structural boundaries.

The structure sub-transfer is the division of a total sentence into "sentences" and/or clauses; clauses into blocks; and blocks into phrases.

The first phase of this sub-transfer is the identification of all punctuation marks within the sentence, in regard to their meaning and function within the sentence.

For example, let us examine the possible meaning and functions of the semicolon ";".

Positions of the semicolon:

- (1) between independent "sentences" such that combinations of them form the total sentence without use of conjunctions. These "sentences" can have commas inside themselves.

Session 2: CURRENT RESEARCH

(2) between independent sentences which are combined into a total sentence by means of

- (a) conjunctions no, odnako, vse zhe, tem ne menee, etc.
- (b) conjunctions i, da

(3) between phrase-type homogeneous members of the sentence, specifically where these "phrases" have modifiers or modifier-groups separated by commas.

(4) between several subordinate clauses with one main clause present in the sentence; in this case, however, the semicolon would not be followed by conjunctions.

(5) between "sentences" which consist of main clause and subordinate clauses, i.e., independent "sentences".

(6) between enumeration or recapitulation.

If we examine the above we can see that with the exception of (3) and (6), the semicolon ";" is always a division mark between "sentences" and, upon recognition on the basis of its function code (71), can be considered as a stop signal between "sentences" in structural analysis.

For identification of the case in point (6), we have to locate the colon. Then the situation would be:

xxxxxx: xxx; xxxx; xxx; xxxxx. Note: x = word
"sentence" "sentences or phrases"

For identification of the case in point (3), the situation is either similar to point (6) (i.e., we will locate the colon), or if the colon is absent, we are dealing with homogeneous phrases which can be treated in the same fashion as point (1).

Therefore, in the Unified Transfer System the semicolon is accepted as a stop signal for the division of a sentence into clauses or "sentences."

The same process is applied to other punctuation marks until we divide the sentence into "sentences", which then in turn have to be divided into introduction, subject, predicate, and final blocks, in the order of their occurrence in a sentence. Then whichever of these blocks are found to be present are rearranged into a model structure of introduction-subject-predicate-final blocks.

For the sake of the discussion, let us consider the identification of the subject.

Session 2: CURRENT RESEARCH

A subject can be:

- (1) any part of speech in the nominative
- (2) a combination of words (cluster) with the connotation of "jointness", like: my s toboj, sestra s bratom, etc.
- (3) numerical combinations with a precise or approximate definition of objects, like: dva priiatelia, neskolko chisel; minogo liudej, etc.

Therefore, we could say that the subject can be:

<u>SUBJECT</u>	<u>EXAMPLES</u>
1. Noun	1. <u>Kolba</u> stoiala na stole
2. Adjective	2. <u>Serye</u> ne podlezhata analizu
3. Participle	3. <u>Spavshie</u> prosnulis'
4. Numeral	4. V vode rastvorilis' tolko <u>dva</u>
5. Pronoun	5. Ne poshel <u>on</u> domoj
6. Verb in infinitive	6. <u>Pisat'</u> trudno
7. Non-inflected word	7. Gromkoe " <u>ura</u> " narushilo tishinu
8. Word cluster	8. <u>Brat s sestroj</u> uchat'sia v universitete
9. Numerical combination	9. <u>Dvesti studentov</u> izuchaiut inostrannye iazyki

We can identify the subject (or subjects) and its (or their) modifiers and rearrange the total into a block sequence by examining the "sentence" for the presence: (1) of a noun in the nominative; or (2) of any other part of speech in the nominative, which can be then either the subject itself or a modifier to the subject in another case (i.e. , cluster subject); or (3) of words like mnogo, malo, etc. , identifying them on the basis of their function-form codes.

Example:

Predicate(2)	Subject (1)	Final (3)	Original block order (inverted word order)
Ne poshel	on	domoj	
Subject(1)	Predicate(2)	Final(3)	Rearranged block order
on	ne poshel	domoi	

Session 2: CURRENT RESEARCH

Whenever the blocks are larger than one word (which happens in most of the cases), we check the rearrangement numeric codes, which are attached to the translation. These numbers are assigned on the basis of the target-language translations as well as the function of its source equivalent.

Example: Dlia naibolee nagliadnogo predstavleniia = "to get a clearer concept"

The source-language preposition *dlia* received a verbal translation, but retained a prepositional rearrangement code because of the source-language word function.

A list of rearrangement codes and their definitions or equivalents follows:

Syntax codes:

- 1 Preposition
- 2 Introductory words (if, that, which, what, how, why, as, since, etc.)
- 3 Conjunctions (and, but, either, or, neither, nor, comma, colon, semicolon, etc.)
- 4 Words like: only, just, then, perhaps, maybe, therefore, however, almost, likewise, etc.
- 5 Not
- 6 It is possible, it is not possible, it is known, etc.
- 7 Some, all, any, none, something, anything, any kind, nothing, etc.
- 11 My, yours, his, etc. (possessive form)
- 12 Numerals (one, two, etc., first, second, etc., few, many, much, more, most, last, etc.)
- 13 Other
- 14 Adjectives (including some pronouns)
- 15 Nouns and pronouns (nominals)
- 16 Myself, yourself, itself, etc.
- 17 Participial modifiers
- 21 Will, may, must, can, do, etc.
- 22 Have
- 23 To be
- 24 Seldom, often, really, verbally, continuous, ever, never (adverbs of time)

Session 2: CURRENT RESEARCH

- 25 Verbs and short forms of participles, adjectives
- 26 Here, there, away, beyond, upstairs (place)
- 27 Equally, rapidly, strangely, unequally, vastly, greatly, considerably, quite, etc. (manner)
- 31 Early, late, later, soon, etc. (time)
- 32 Period

On the basis of the above table, the constituent parts of the system are as follows:

1. Source-language dictionary (in our case the Russian Dictionary). Its format is: Russian word = Dictionary Line # + Code Pattern # .

Example:

predstavleniia = 14155* 66-212

* arbitrary Dictionary Line #

2. Target-language dictionary (in our case the English dictionary). Its format is: Dictionary Line # - English equivalent + rearrangement code.

Example:

14155	A	concept	015
	B	concept	015
	C	of the performance	015
	D	the performance	015
	E	performance	015
	F	of the concept	015
	H	the performances	015
	J	performances	015
	K	the performances	015
	L	performances	015

Session 2: CURRENT RESEARCH

3. Code patterns arranged by code pattern # .

Example:

66-212	4115 230 000
	3116 230 000
	0111 230 421
	0111 230 401
	0111 230 400
	0111 230 021
	0111 230 001
	0111 230 000
	0111 131 401
	0111 130 400
	0111 431 401
	0111 431 400

The code distribution in the pattern is:

66 is the functional part of speech

212 is the actual number of the pattern

The line of 10-digit code is the unified transfer code:

1	2	3	4	5	6
1 digit	3 digits	3 digits	1 digit	1 digit	1 digit

1. Semantic (meaning) boundary indicator
2. Semantic class
3. Form or morphology group
4. Subject-matter indicator (microglossary control)
5. Preposition control
6. Article control

The system instructions outside of strict data preparation (text preparation, dictionary lookup, sorts, etc.) are divided as follows:

1. Progression instructions (determination of working area)
2. Selection:
 - (a) comparison of codes in patterns for the selection of form and meaning
 - (b) article and preposition selection

Session 2: CURRENT RESEARCH

3. Verification or correction instructions for:
 - (a) changes in progression (larger or smaller string, single word selection, etc.)
 - (b) pronoun and conjunction selection
4. Sentence recognition and division instructions
5. Syntactic block recognition
6. Rearrangement instructions

Session 2: CURRENT RESEARCH

TRANSLITERATION SCHEME

LIBRARY OF CONGRESS

UNIFIED TRANSFER SYSTEM

A	A
B	B
V	V
G	G
D	D
E	E
ZH	ZH
Z	Z
I	I
Ī	J
K	K
L	L
M	M
N	N
O	O
P	P
R	R
S	S
T	T
U	U
F	F
KH	KH
TS	TS
CH	CH
SH	SH
SHCH	SC
"	'
y	y
'	6
Ě	EH
IU	IU
IA	IA