designates the performer of an action with causatives having a valence higher than two. Quadrivalence is practically never encountered, and quintivalence is apparently impossible due to the bulkiness of the structure that would result.

31. In the table above, each derivational step corresponded to an increase by one unit in syntactic valence. We have seen above that this possibility is far from always realized, and in some cases is not realizable.

We can make one refinement in the question of syntactic valence which is connected with those cases where there is reason to speak of a *preservation of the original syntactic valence*. These are cases encountered very often with verbs of emotion (classified in the table as univalent) and verbs relating to the receiving and taking of smth (classified in the table as bivalent). Cf., for example, in Swahili: *Mtoto*^{1} *a-li-mw-ogop-a*^{2} *mgeni*^{3} 'The baby^1 was frightened^2 by the stranger^3' (a- is a congruence morpheme for the subject, mw- a congruence morpheme for the object, li- the past tense marker) – *Mgeni*^{1} *a-li-mw-ogop-esh-a*^{2} *mtoto*^{3} 'The stranger^1 frightened^2 the baby^3' (-esh- is the causative suffix). In such cases, however, it seems more exact to speak not of the full preservation of valence (although there is a direct object in both examples), but of a transformation of valence from syntactically optional to obligatory. Since, however, the reason for fright, just like the reason for any other state, can be expressed by various syntactic forms (cf.: *Giza*^{1} *ilipoingia*^{2} *mtoto*^{3} *a-li-ogop-a*^{4} 'When^2 it became^2 dark^1 the baby^3 became frightened^4'; *Akiona*^{1} *mgeni*^{2} *mtoto*^{3} *a-li-ogop-a*^{4} 'When he saw^1 the stranger^2, the baby^3 became frightened^4', etc.) or can even remain unknown to the speaker, who observes only the outward manifestation of emotion (i.e., there is no mention at all of the cause in the utterance), there is also reason in most cases to look for an increase in valence for verbs of this type, with the reservation that in certain syntactic constructions there can occur an alternation between a construction with several disjunctive syntactic types of optional environments with a non-causative and a construction with one obligatory type of environment (a direct object) with a causative. Of course, far from all specific indications of cause with a V_i can be transformed into the subject in the V_j as was done in the examples above.

Soviet Academy of Sciences, Leningrad

TOWARDS A LINGUISTIC 'MEANING⇔TEXT' MODEL

SUMMARY. **1.** The proposed model is a translative (= transformative), rather than a generative, system; its purpose consists in establishing correspondences between any given meaning and (ideally) all synonymous texts having this meaning. We assume here that we can formally describe meaning as an invariant of a set of equisignificant texts, that analysis of meaning as such lays outside the model, and that the 'Meaning⇔Text' Model (MTM) is a fragment of the more general model 'Reality⇔(Meaning⇔Text)⇔⇔Speech'.

2. Five levels of utterance representation are used in the MTM: (I) The semantic level – a semantic graph (semantic atoms – *semes* – connected by semantic relations) plus indications of topicalization, etc.; (II) The syntactic level with two sub-levels: (IIa) The deep-syntax representation of a sentence – a deep dependency tree with universal deep-syntax relations plus indications of topicalization, anaphoric relations, and certain constituents (where necessary), (IIb) The surface-syntax representation of the sentence – a surface dependency tree with the specific syntactic relations of the language in question plus the same additional information as in (IIa); (III) The morphological level: (IIIa) The deep-morphology representation of word-forms (the name of the respective lexeme supplied with a full morphological description), (IIIb) The surface-morphology representation of word-forms (a string of morphs and/or morphological operations); (IV) The phonological level; (V) The phonetic/graphic-orthographic level.

3. The MTM consists of five basic components which establish correspondences between the representations of the above levels. Examples of the rules of each component are given.

4. Certain linguistic problems connected with the MTM are treated: a new type of dictionary (two dictionary entries are given); word-derivation and the possible formal-semantic relations between linguistic signs (signifié, signifiant, syntactics – information on the combinatorial properties of the sign); and a tentative definition of conversion in English.

In this paper we will proceed from the following hypothesis: a natural language may be viewed as a special kind of (logical) device which provides for the comprehension of a given utterance, i.e., the *perception of its meaning(s)*, or the construction of utterances *which express a given meaning*. Then it should be a device which provides correspondences between meanings and texts, or maps the set of all possible meanings onto the set of all possible texts – and vice versa.

Hypotheses about such devices can be formulated as functional models, or logical systems, describing the mapping 'Meaning⇔Text'. We propose to discuss below one such model.

Here, of course, we can give only a very general description of the model; our exposition will of necessity be sketchy and somewhat dogmatic. The relationship between 'Meaning⇔Text' model and the very closely related

F. Kiefer (ed.), Trends in Soviet Theoretical Linguistics, 33–57.

linguistic models proposed by the theories of generative grammars and generative semantics will not be touched upon at all. We will limit ourselves to emphasizing the fact that N. Chomsky's theories and their modern development by M. Halle, C. Fillmore, G. Lakoff, J. McCawley, and many others have significantly influenced the author, who is in addition particularly indebted to A. K. Žolkovskij, with whom he has worked in close contact for several years. Many of our propositions below have been taken from the joint works of A. K. Žolkovskij and the author (most notably, Žolkovskij and Mel'čuk, 1967).

N.B.: Reference will be made only to those publications which have a direct connection with the 'Meaning⇔Text' model described in this paper.

I. PROPERTIES OF THE 'MEANING⇔TEXT' MODELS

The main feature of the 'Meaning⇔Text' model (MTM) consists in the following: it is not a generative, but a translative (=transformative) system; it does not seek to generate grammatically correct (or meaningful, etc.) texts, but merely to match, ideally, *any* given meaning with all symonymous texts having this meaning, and conversely, to match *any* given text with all the meanings this text can have. Here we make the following three assumptions:

(a) We are able to describe the meaning of any utterance in a special semantic language. – Meaning is understood to be an invariant of a set of equisignificant texts (this equisignificance is considered to be intuitively obvious to a native speaker).

(b) The analysis of the meaning itself (the discovery of various semantic anomalies – contradictions, absurdities, trivialities, etc.) goes beyond the MTM as such; a different type of device is needed for this purpose.[1]

(c) The 'Meaning⇔Text' model should be a fragment of the more general and complete model of human (intellectual+linguistic) behaviour: 'Reality⇔Speech', i.e.,

$$\text{'}\underbrace{\text{Reality}\Leftrightarrow\text{Meaning}}_{\text{I}}\Leftrightarrow\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\!\overbrace{}\underbrace{\text{Meaning}\Leftrightarrow\text{Text}}_{\text{II}}\ \ \underbrace{\text{Text}\Leftrightarrow\text{Speech}}_{\text{III}}\text{'.}$$

In our opinion, only fragment II is object of linguistics proper; only this fragment is represented in the MTM.

The following limitations have been observed in our work on the MTM:

[1] Thus, we meet here with the essential *asymmetry* of texts and meanings: our model should catch all formal, i.e. linguistic, anomalies of a text, but it does not deal with the semantic ones.

(1) The MTM is a *purely functional* model. No attempts have been made to relate it experimentally with psychological or neurological reality; for the time being, therefore, the MTM is nothing more than a logical means for describing observable correspondences between meanings and texts.

(2) The transformation 'meanings⇔texts' is described in the MTM only *as a set of correspondences* between the former and the latter. The possible *procedures* for moving from meanings to texts and vice versa will not be treated here at all. In other words, the MTM in its present state models only *competence*, and not *performance*.

(3) Possible 'feedback' between texts and meanings in the process of speaking (changes in the original semantic message under the influence of an already constructed text, etc.) are not taken into account.

(4) Functions of natural language other than the communicative one are not considered at all, which amounts to viewing language as a communication system only, that is, as a 'Meaning⇔Text' transformer.

(5) The extremely important question of how language is acquired and perfected will be left completely untouched.

The MTM has been developed primarily on the basis of Russian linguistic data, which will also be used as illustrative materials in the present paper.

For a general description of the 'Meaning⇔Text' model see Mel'čuk (1970).

II. UTTERANCE REPRESENTATION LEVELS IN THE 'MEANING⇔TEXT' MODEL

In view of the fact that homonymy and synonymy so widely spread in natural languages, highly complicate the direct correspondence between meanings and texts, a number of *representation levels* have been established in the MTM.[2] Five levels for the representation of utterances are distinguished: the semantic, syntactic, morphological, phonological, and phonetic/graphic levels.

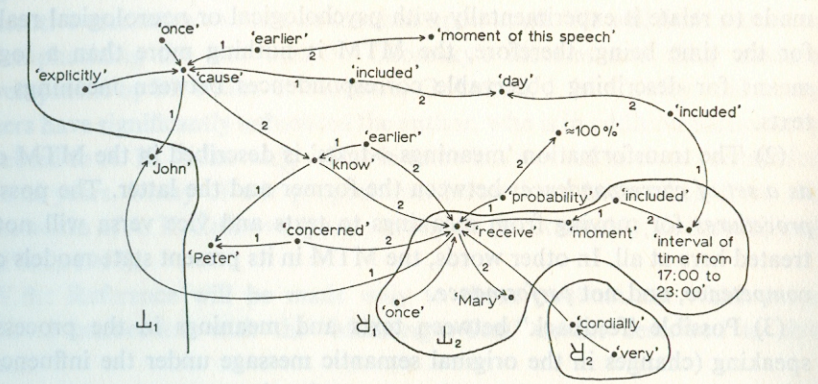
(I) *The semantic level: a semantic representation* is assigned to the utterance; see (1) – semantic representation of a set of synonymous Russian utterances exemplified by the sentence:

Ivan tverdo obeščal Petru, čto večerom on primet Mariju samym teplym obrazom,

'John firmly promised Peter that [this] evening he would receive Mary in a most cordial manner'.

[2] In complete accordance with the generally held view: cf. the stratificational grammar of S. Lamb, the Mehrstufiges Generatives System of P. Sgall, and others.

(1)



(*N.B.*: This and all subsequent examples do not claim to be semantically precise. They are purely illustrative in nature.)

A possible approximate reading: 'John explicitly caused once [past form of verb is represented by 'earlier than the moment of this speech'] Peter to know that after he, John, would receive Mary, with which reception Peter is concerned, in a moment included in the time interval from 17:00 to 23:00 in the day when it is said, and the reception, which has a probability of nearly 100%, would be very cordial'.

The semantic representation consists of the following two components:

(1) A *semantic graph* gives the meaning of an utterance without distributing it into sentences or words. (Such linguistic features, as the selection of specific words and syntactic constructions, are deliberately not shown at all in the semantic graph.) This is a connected directed graph with no additional restrictions whatsoever.[3]

The nodes of the graph are *semantic units* (SU) which generally represent 'semantic atoms' – meanings which within the given description are considered to be elementary; these meanings are called *semes*. By way of abbreviation, however, we allow 'intermediate' SU's – symbolizing complexes of semes, i.e., whole semantic (sub)graphs. Thus, in (1) the SU '$A \leftarrow$ earlier [than] $\rightarrow B$' is an abbreviation for 'the time of *A* precedes the time of *B*', and the SU 'know' is an abbreviation of something like 'have true information about'; it is obvious that 'concerned', 'receive', and 'cordially' are all far from

[3] The use of a graph (= network) to represent a semantic content of utterances is by no means a novelty in linguistic analysis. Suffice it to mention here the pioneer paper by K. I. Babickij (*Naučno-texničeskaja Informacija*, 1965, No. 6), the book by W. Hutchins (1971) with further references and many well-known publications in and on the inference and question-answering systems (by Quillian, Bobrow, Simmons, and others).

elementary. A complete analysis into semes would make the graph unreadable.

The arcs of the graph are *semantic relations*, i.e., relations between SU's.

SU's divide into two classes. One type of SU consists of names (of classes) of objects or single objects, in particular, proper names; semantic relation arcs can only enter them. The other type of SU, conventionally called *predicates*, consists of predicates, quantifiers, and logical constants; arcs can also leave them. (*N.B.*: the arrows on the arcs point from predicates to their arguments.)

Predicate semes are never more than two-place; intermediate SU's can have up to five (or even more?) places. The various arcs leaving a single node are numbered: $A \xleftarrow{1}$'cause'$\xrightarrow{2} B$ means that *A* is the first, and *B* the second argument of the predicate 'cause' (*A* is the causer, or the cause; *B* is the caused, or the result, etc.). How the nodes of a semantic graph are physically distributed on a sheet is of no significance.

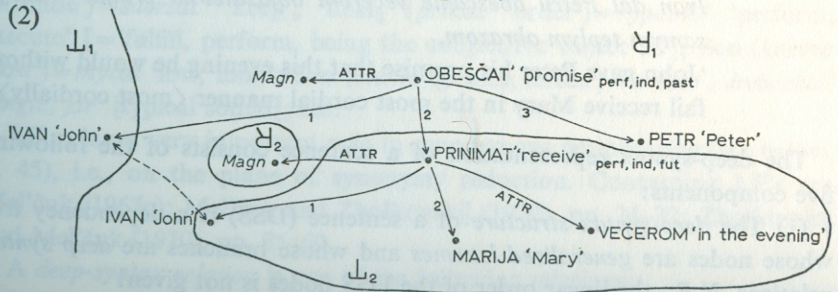
(2) *Information about the communicative organization* of an utterance: about the topic (ꓕ, Fr. thème) – comment (ꓤ, Fr. rhème), about the old – the new, about the psychological value of a particular meaning fragment for the speaker, about emotional emphasis. This information stands in approximately the same relationship to a semantic graph as prosodic phenomena to the string of phonemes which make up a sentence. (In our simplified examples only information about the topic-comment is indicated.)

The division into topic and comment can have successive 'strata'; thus, within the comment of the first stratum in (1), there is a division into $ꓕ_2$ – $ꓤ_2$.

(II) *The syntatic level* can be differentiated into two sublevels.

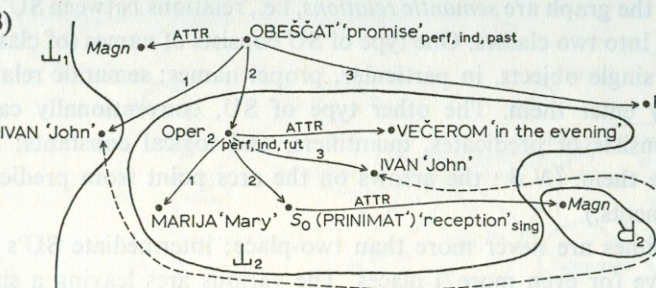
(IIa) *Deep syntax*: the utterance is given as a sequence of sentences; each sentence is assigned a so-called *deep-syntax representation* (DSR); see (2)–(4) – the deep-syntax representations of some synonymous Russian sentences corresponding to semantic graph (1):

(2)



Ivan tverdo obeščal Petru, čto večerom on primet Mariju samym teplym obrazom, 'John firmly promised Peter that this evening he would receive Mary in the most cordial manner'.

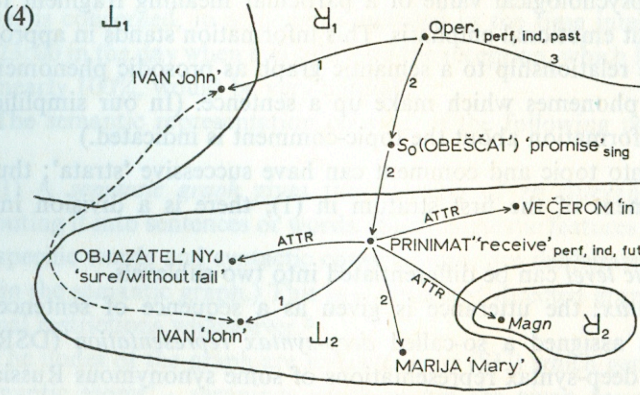
(3)



Ivan tverdo obeščal Petru, čto večerom Marija najdet u nego samyj teplyj ⟨radušnyj⟩ priem,

'John firmly promised Peter that this evening Mary would receive [find] from him a most cordial ⟨hearty⟩ welcome/that he would welcome Mary most cordially ⟨heartily⟩'.

(4)



Ivan dal Petru obeščanie večerom objazatel'no prinjat' Mariju samym teplym obrazom,

'John gave Peter his promise that this evening he would without fail receive Mary in the most cordial manner ⟨most cordially⟩'.

The deep-syntax representation of a sentence consists of the following five components:

(1) The *deep-syntax structure* of a sentence (DSS) is a dependency tree whose nodes are *generalized lexemes* and whose branches are *deep syntax relations*. *N.B.*: the linear order of the DSS nodes is not given!

A generalized lexeme is one of the following objects:

Either it is a full lexeme of the language in question (empty words, strongly

governed prepositions and conjunctions, auxiliary verbs in complex forms, etc., are not represented in the DSS);

Or it is a fictive lexeme: for example, the symbol for an indefinite personal subject (≈ Fr. *on*, Ger. *man*), which has no expression in an actual Russian text;

Or it is a whole idiom: e.g., *s'est' sobaku* 'know something backwards and forwards, know one's stuff', *sinij čulok* 'bluestocking', etc.;

Or it is a symbol for a lexical function (see below).

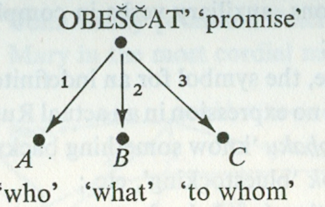
The symbol for a generalized lexeme can have subscripts for the morphological features which have full meaning and are not determined syntactically: number of the noun; aspect, tense, and mood of the verb.

A *lexical function* (LF) **f** is a relation which connects a keyword (or word-group) W – the *argument* of LF – with a set of other words or word-groups $\mathbf{f}(W)$ – the *value* of LF – in such a way that for any W^1 and W^2, if only $\mathbf{f}(W^1)$ and $\mathbf{f}(W^2)$ exist, both $\mathbf{f}(W^1)$ and $\mathbf{f}(W^2)$ hold an identical relationship – with respect to meaning and syntactic role – to W^1 and W^2, respectively; in the majority of cases $\mathbf{f}(W)$ is also different for different W's which means that $\mathbf{f}(W)$ is 'phraseologically bound' by W. We have arrived at about 50 *standard elementary* LF's, that is, those whose number of possible arguments and number of possible values is sufficiently large. *Complex* LF's, which are composed of standard LF's, are also possible. Some examples of LF's: **Syn** (*priglašat'* 'invite') = *zvat'* 'ask, call'; **Syn** (*xudoj* 'thin') = *toščij* 'skinny' [synonym]; $\mathbf{Conv}_{21}$ (*pered* 'in front of') = *szadi* 'in back of', $\mathbf{Conv}_{21}$ (*sledovat'* 'follow') = *predšestvovat'* 'precede' [conversive]; S_0 (*polagat'* 'believe') = *mnenie* 'opinion' [nomen actionis]; **Magn** (*priglašenie* 'invitation') = *nastojčivoe* 'urgent, persistent', **Magn** (*bereč'* 'keep, cherish') = *kak zenicu oka* 'like the apple of one's eye', **Magn** (*xudoj* 'thin') = *kak skelet/ščepka* 'as a skeleton/a lath' [= 'very']; $\mathbf{Oper}_1$ (*prikaz* 'order') = *davat'* 'give', $\mathbf{Oper}_1$ (*mnenie* 'opinion') = *imet'*, *predërživat'sja* 'have, hold' [= 'be the subject of']; $\mathbf{Oper}_2$ (*priglašenie* 'invitation') = *polučat'* 'receive'; $\mathbf{Oper}_2$ (*kontrol'* 'control') = *byt' pod* 'be under' [= 'be the object of']; $\mathbf{Real}_1$ (*obeščanie* 'promise') = *sderžat'* 'keep'; $\mathbf{Real}_2$ (*prikaz* 'order') = *vypolnit'* 'perform, execute' [= 'fulfill, perform, being the subject/the object of']; **Son** (*korova* 'cow') = *myčat* 'low, moo'; **Son** (*stekla* 'glasses, lenses') = *zvenet'*, *drebezžat'* '*jingle, jar*' [typical sound]; etc.

LF's play a very important role in synonymous periphrasing (see below, p. 45), i.e., on the plane of synonymy reduction. Concerning LF's see Mel'čuk (1967c); Mel'čuk and Zholkovskij (1970), pp. 24–32; Zholkovskij and Mel'čuk (1970), pp. 35–60.

A *deep-syntax relation* is one of the following relations:

1, 2, 3, 4, 5 (maybe, 6)–relations which connect a lexeme-predicate with its first, second, third, fourth, fifth (or sixth) arguments, respectively:



ATTR – a relation which connects any entity with its attribute, or modifier (in the broadest sense of the term):

TOČKA 'point', ÈTOT 'this',
• ——ATTR——→ •

PRINIMAT' 'greet', 'receive' VEČEROM 'in the evening', etc.
• ——ATTR——→ •

COORD – a relation which connects any two coordinated entities:

IVAN 'John' PETR 'Peter' I 'and' MARIJA 'Mary',
• ——COORD——→ • ——COORD——→ • ——→ •

i.e. 'John, Peter, and Mary'.

APPEND – a relation which connects the root of the sentence deep-syntax tree with any so-called 'independent' element (such as parenthetical expression, interjection, address, etc.):

PRIXODIT' 'come'$_{\text{perf, ind, past}}$
O 'oh' • ←—APPEND—— • ——1——→ •
ON 'he',
i.e. 'Oh, he came'.

We presume that these relations are sufficient to be able to describe any syntactic constructions of any language on the deep level.

(2) Information about the communicative organization of the sentence, specifically, indications of the topic and comment, see ⊥ and Ʀ in (2)–(4).

(3) Information about coreferentiality of particular phrases; in (2)–(4) the coreferential nodes ('the same John') are connected by a dotted arrow.

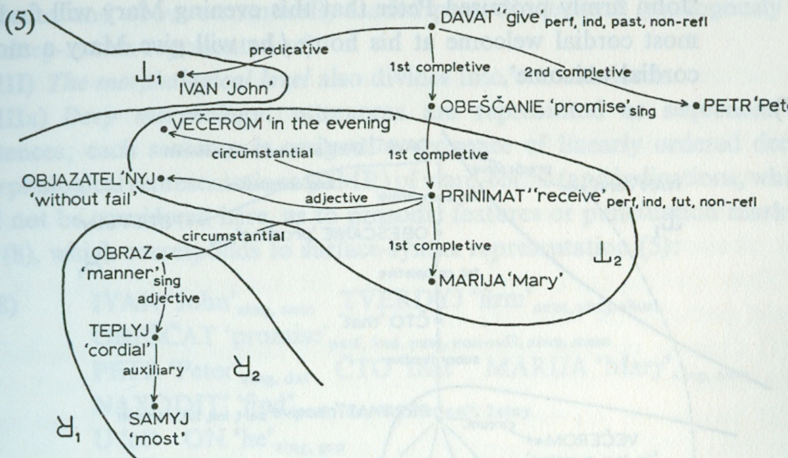
(4) Information about constituents – those which cannot be represented in a natural way by a dependency tree (like *old man and women* etc).

(5) Information about meaningful (i.e. not syntactically conditioned) prosodic phenomena, like intonation contours, pauses, junctures, emphatic stresses, and the like.

(IIb) *Surface syntax*: utterances are given as sequences of sentences; each sentence is assigned a surface-syntax representation (SSR); see (5), which is

the SSR corresponding to DSR (3), and (6)–(7), which are the SSR's corresponding to DSR (4):

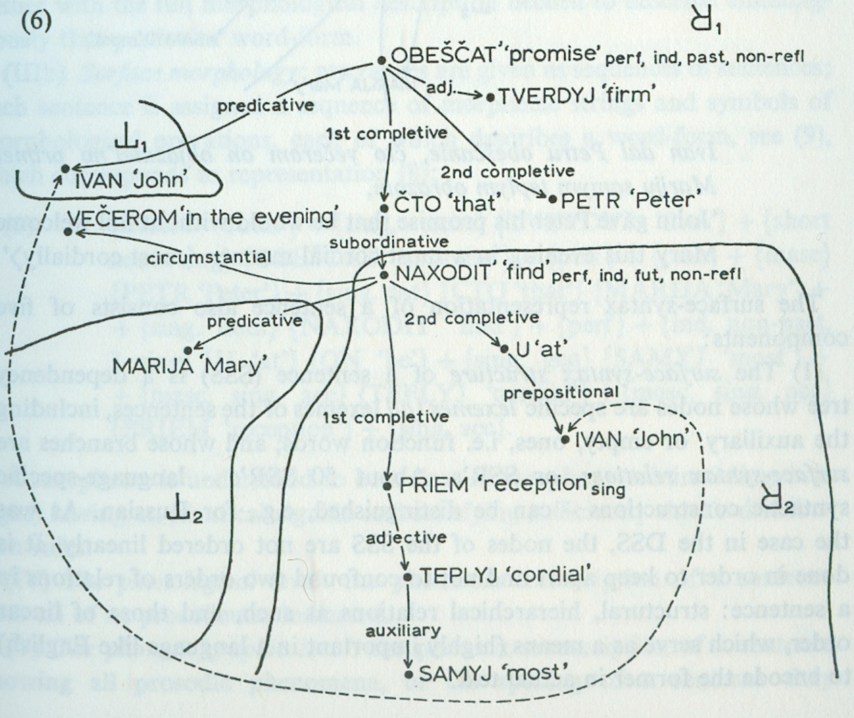
(5)



Ivan dal Petru obeščanie večerom objazatel'no prinjat' Mariju samym teplym obrazom,

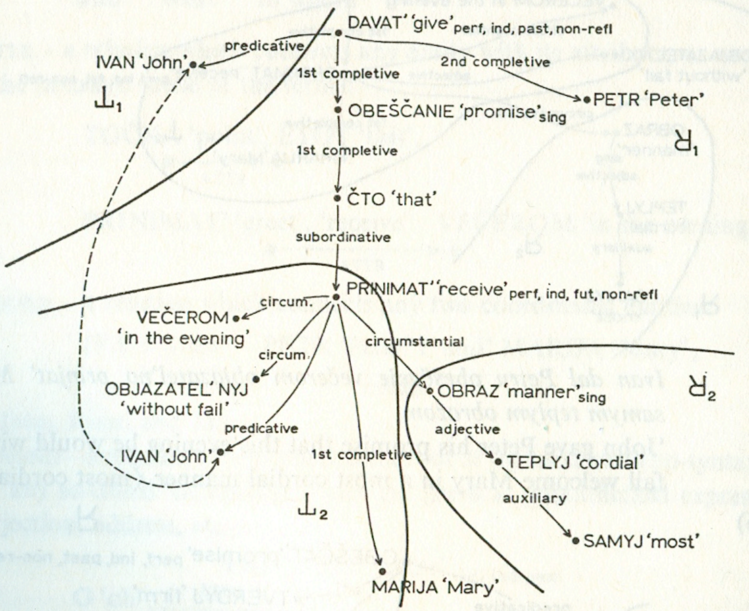
'John gave Peter his promise that this evening he would without fail welcome Mary in a most cordial manner ⟨most cordially⟩',

(6)



Ivan tverdo obeščal Petru, čto večerom Marija najdet u nego samyj teplyj priem,
'John firmly promised Peter that this evening Mary will find a most cordial welcome at his home / he will give Mary a most cordial welcome',

(7)



Ivan dal Petru obeščanie, čto večerom on objazatel'no primet Mariju samym teplym obrazom,
'John gave Peter his promise that he would without fail welcome Mary this evening in a most cordial manner ⟨most cordially⟩'.

The surface-syntax representation of a sentence also consists of five components:

(1) The *surface-syntax structure* of a sentence (SSS) is a dependency tree whose nodes are specific *lexemes* (*all* lexemes of the sentences, including the auxiliary, or empty, ones, i.e. function words, and whose branches are *surface-syntax relations*, or SSR's. About 50 SSR's – language-specific syntactic constructions – can be distinguished, e.g., for Russian. As was the case in the DSS, the nodes of the SSS are not ordered linearly. It is done in order to keep apart and not to confound two orders of relations in a sentence: structural, hierarchical relations as such, and those of linear order, which serve as a means (highly important in a language like English) to encode the former in actual text.

(2)–(5) Information about the communicative organization of the sentence, about coreferentiality of nominal and other phrases, about the constituents, and about semantically loaded prosodic features (analogously to the deep-syntax representation).

(III) *The morphological level* also divides into two sub-levels.

(IIIa) *Deep morphology*: utterances are represented as sequences of sentences; each sentence is assigned a sequence of linearly ordered deep-morphological representations (DMR) of word-forms (and indications, which will not be considered here, as to prosodic features or punctuation marks); see (8), which corresponds to surface-syntax representation (5):

(8) IVAN 'John'$_{\text{sing, nom}}$ TVERDYJ 'firm'$_{\text{neut, sing, short}}$
OBEŠČAT 'promise'$_{\text{perf, ind, past, non-refl, sing, masc}}$
PETR 'Peter'$_{\text{sing, dat}}$ ČTO 'that' MARIJA 'Mary'$_{\text{sing, nom}}$
NAXODIT' 'find'$_{\text{perf, ind, fut, non-refl, 3sing}}$
U 'at' ON 'he'$_{\text{sing, gen}}$
VEČEROM 'this evening' SAMYJ 'most'$_{\text{masc, sing, acc}}$
TEPLYJ 'cordial'$_{\text{masc, sing, acc}}$
PRIEM 'reception'$_{\text{sing, acc}}$

A word-form DMR consists of the name of the respective lexeme together with the full morphological description needed to describe unambiguously that particular word-form.

(IIIb) *Surface morphology*: utterances are given as sequences of sentences; each sentence is assigned a sequence of morpheme strings and symbols of morphological operations, each of which describes a word-form, see (9), which corresponds to representation (8):

(9) {IVAN 'John'} + {sing, nom} {TVERDYJ 'firm'} + {short neut. sing} {OBEŠČAT' 'promise'} + {perf} + {past} + {masc} {PETR 'Peter'} + {sing. dat} {ČTO 'that'} {MARIJA 'Mary'} + + {sing. nom} {NAXODIT' 'find'} + {perf} + {ind. non-past. 3 sing} {U 'at'} {ON 'he'} + {sing. gen} {SAMYJ 'most'} + + {masc. sing. acc} {TEPLYJ 'cordial'} + {masc. sing. acc} {PRIEM 'reception'} + {sing. acc}.

A morpheme is understood to be a class of morphs (= minimal linguistic signs) having an identical signifié and satisfying sufficiently simple distribution rules.

(IV) *The phonological level*: the phonemic transcription of a sentence with all of its prosodemes indicated.

(V) *The phonetic/graphic level*: the phonetic transcription of a sentence showing all prosodic phenomena, or the spelling of a sentence duly punctuated (*N.B.*: levels (IV) and (V) are not treated in the present paper).

III. THE DESIGN OF THE 'MEANING⇔TEXT' MODEL

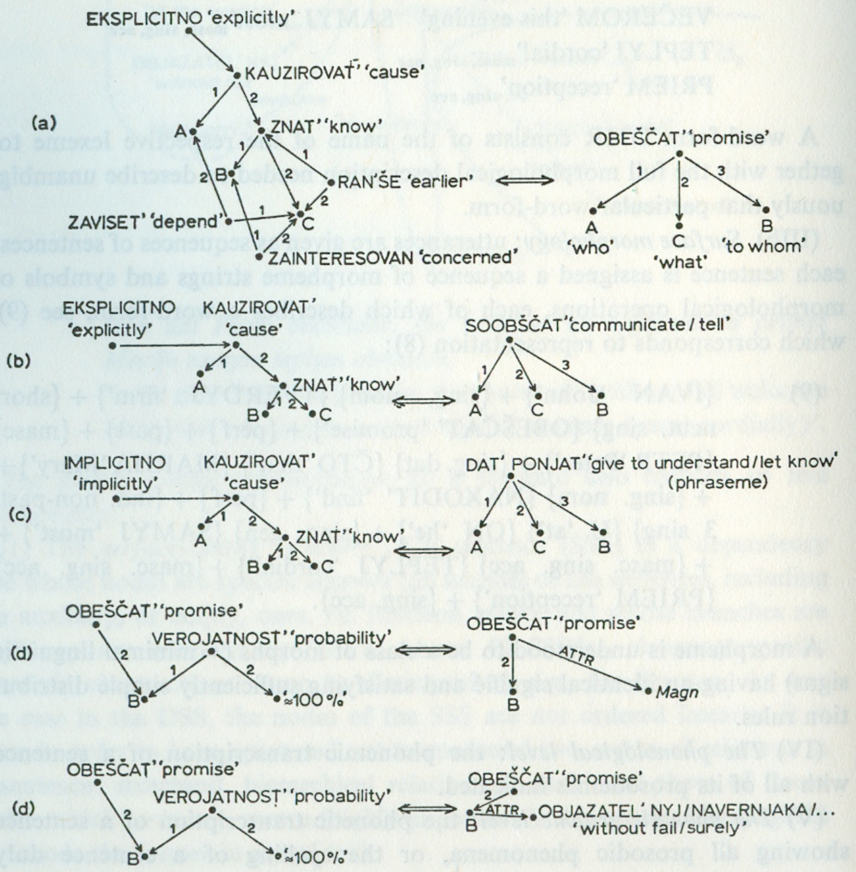
Transitions from one level of utterance representation to another are accomplished by the following basic components of the model.

(I) *The semantic component* establishes correspondences between the semantic representation of an utterance and all alternative (=synonymous) sequences of deep-syntax representations of the sentences which make up this utterance.

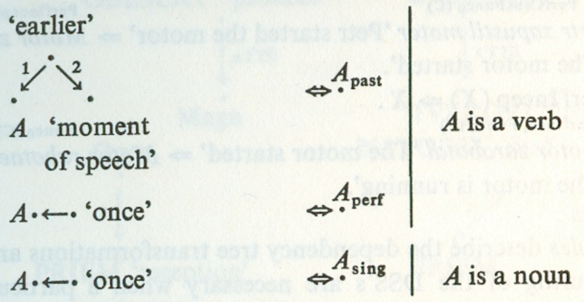
When moving from meaning to text, the semantic component of the model performs the following operations:

(1) Cuts the semantic graph into subgraphs, each of which corresponds to a sentence.

(2) Selects the corresponding words by means of rules of the type:



(3) Selects the semantic (i.e., syntactically unconditioned) morphological characteristics of the lexical items by means of rules of the type:



(4) Forms the deep-syntax structure of the sentenses.

(5) Processes the remaining components of the deep-syntax representations.

(6) For each DSS are constructed all its synonymous DSS's such that this synonymy can be described in terms of lexical functions. In other words, an 'algebra' of the transformations on the DSS which contain LF symbols is given.

These transformations can be described by rules of two classes:

Lexical rules (at present there are about 60 of them) are either semantic equivalencies or semantic implications. Examples:

Equivalencies:

(1) $\overset{X}{C} \Leftrightarrow \overset{Y}{\mathrm{Conv}_{21}(C)}$.

$\overset{C}{\text{The set } A \text{ also contains the point } x} \Leftrightarrow$

$\overset{\mathrm{Conv}_{21}(C)}{\text{The point } x \text{ also belongs to the set } A}$.

(2) $\overset{X}{C} \Leftrightarrow \overset{Y}{\mathrm{Adv}_{1B}(C)}$.

On $\overset{C}{\textit{pospešil}}$ *vyjti* 'He hurried to leave' ⇔ *On* $\overset{\mathrm{Adv}_{1B}(C)}{\textit{pospešno}}$ *vyšel* 'He hurriedly left'.

(3) $\overset{X}{C} \Leftrightarrow \overset{Y}{S_0(C) + \mathrm{Oper}_1(S_0(C))}$.

On $\overset{C}{\textit{boretsja}}$ 'He is struggling' ⇔ *On* $\overset{\mathrm{Oper}_1}{\textit{vedet}}$ $\overset{S_0(C)}{\textit{bor'bu}}$ 'He is waging a struggle'.

(4) $\overset{X}{\mathrm{Real}_2(C)} \Leftrightarrow \overset{Y}{\mathrm{Adv}_{1B}(\mathrm{Real}_2(C))}$.

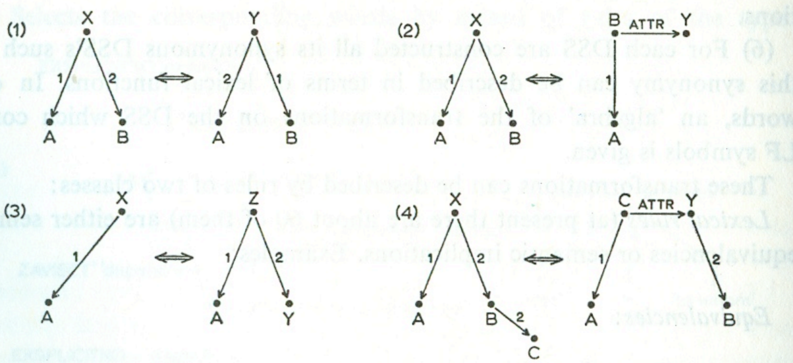
On posledoval ee sovetu uexat' 'He followed her advice to leave' ⇔ *On uexal po ee sovetu* 'He left on her advice'.

Implications:

(1) PerfCaus $(X) \Rightarrow$ PerfIncep (X).
PerfCausFunc₀ (C) PerfIncepFunc₀ (C)
Petr zapustil motor 'Petr started the motor' $\Rightarrow$ *Motor zarabotal* 'The motor started'.

(2) PerfIncep $(X) \Rightarrow X$.
PerfIncepFunc₀ (C) C Func₀ (C)
Motor zarabotal 'The motor started' $\Rightarrow$ *Motor rabotaet* 'The motor is running'.

Syntactic rules describe the dependency tree transformations and indicate what restructuring of the DSS's are necessary when a particular lexical rule is applied. Thus, in order to operate the lexical rules given above, the following types of syntactic rules are necessary:



For DSS transformation rules, see Zholkovskij and Mel'čuk (1970), pp. 60–81.

A special formalism has been devised for describing unordered dependency tree transformations – so called Δ-grammars (Gladkij and Mel'čuk (1969), Gladkij and Mel'čuk (1971)).

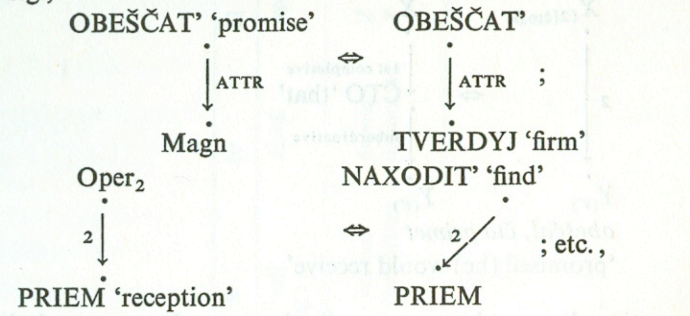
(II) *The syntactic component* establishes correspondences between the deep-syntax representation of a sentence and all the deep-morphology representations which correspond to it. These correspondences are established by two stages.

(IIa) The transition from the deep-syntax representation of a sentence to all its alternative surface-syntax representations can be conceived of as two kinds of operations:

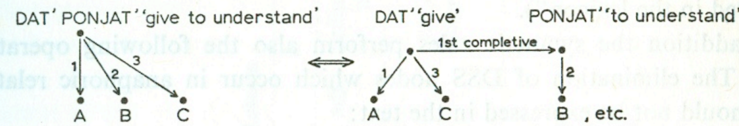
(1) DSS⇔SSS transformations. DSS to SSS transformation rules are used when moving from meaning to text, and these rules again divide into two classes:

– Lexical rules transform the nodes of the trees.

With the help of the lexicon, they 'compute' the values of the lexical functions, e.g.,

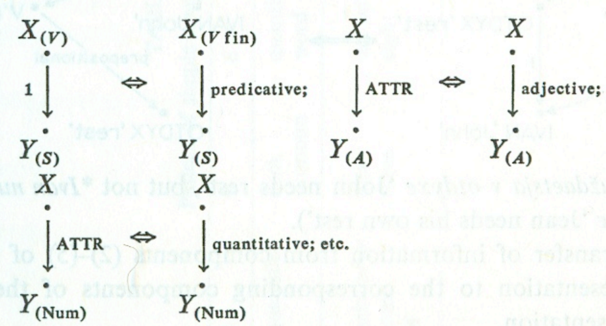


or expand the symbols of the phrasemes into surface-syntax sub-trees, e.g.,

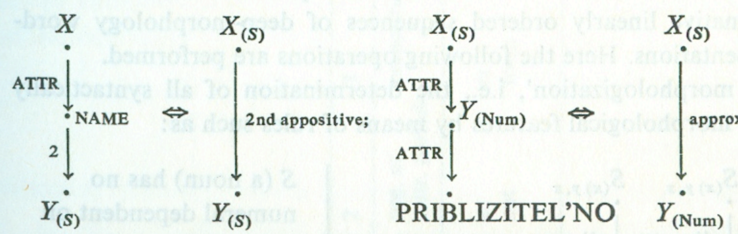


– Syntactic rules transform the tree itself, performing the following operations:

(a) The replacement of a deep relation by a surface one, e.g.,



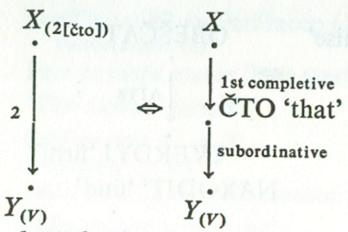
(b) The replacement of a deep node by a surface relation:



Množestvo Q, *gorod Nansi* — 'approximately' *čelovek vosem*

'The set *Q*', 'The city of Nancy' 'about eight people'

(c) The replacement of a deep relation by a surface node:

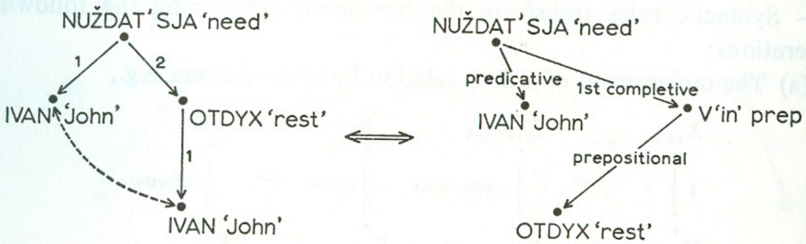


obeščal, čto primet
'promised (he) would receive'

(the notation $X_{(2[\check{c}to])}$ here means "a lexeme whose second deep-syntax valency is filled in the SSS by the conjunction ČTO 'that'; this information is stored in the lexicon").

In addition the syntactic rules perform also the following operation:

(d) The elimination of DSS nodes which occur in anaphoric relations and should not be expressed in the text:

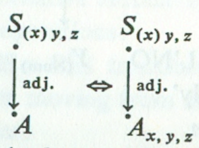


(i.e., *Ivan nuždaetsja v otdyxe* 'John needs rest', but not **Ivan nuždaetsja v svoem otdyxe* 'Jean needs his own rest').

(2) The transfer of information from components (2)–(5) of the deep-syntax representation to the corresponding components of the surface-syntax representation.

(IIb) The transition from the surface-syntax representation of a sentence to all alternative linearly ordered sequences of deep-morphology word-form representations. Here the following operations are performed.

(1) 'SSS morphologization', i.e., the determination of all syntactically contitioned morphological features by means of rules such as:



S (a noun) has no numeral dependent on it (x, y, z – gender, number, case; A is an adjective)

(*polnaja tablica* 'complete table')

TABLE I

Pattern of simple noun phrase in Russian

1	2	3	4	5	6	7	8	9	10	11
Coord. conj.	neg. or restr. particle	preposition	quantifier	dem. pronoun	quantity word	poss. adj.	ordinal num. (adj.)	adj.	head	apposition
ili	*ne*	*iz*	*nekotoryx*	*ètix*	*vos'mi*	*našix*	*vtoryx*	*važnyx*	*formul*	*Z*
or	not	from	some of	these	eight	our	second	important	formulae	Z

$V \quad\quad V$ (*V* is a verb)

$\downarrow_{\text{adj.}} \Leftrightarrow \downarrow_{\text{adj.}}$

$A \quad\quad A_{\text{neut, sing}}$

(*lovko povernulsja* 'turned nimbly')

$Pr_{(z)} \quad\quad Pr_{(z)}$ ($Pr_{(z)}$ is a preposition governing a case *z*)

$\downarrow_{\text{prepositional}} \Leftrightarrow \downarrow_{\text{prepositional}}$

$S \quad\quad S_z$

As a result we obtain deep-morphology representations (DMR's) in the SSS nodes.

(2) 'SSS linearization', i.e., the determination of the linear order of the word-form DMR's (word order) for subtrees of the SSS corresponding to clauses (within a complex sentence) by means of rules of three following types (Mel'čuk, 1967a):

– The determination of word order in the simple phrases in accordance with patterns of the type shown in Table I.

– The determination of the order of simple phrases within derived (full, or compound) phrases.

– The determination of the order of the derived phrases within clauses taking into account topicalization and a number of other intricate mutually interacting factors.

(3) The combination of the DMR strings which correspond to clauses, into a single DMR string for the whole sentence.

(4) The introduction of pronouns into the DMR string (=pronominalization).

(5) The obligatory and optional ellipses carried out on the DMR string.

(III) *The morphological component* establishes – also in two steps – the correspondence between the deep-morphology representations of a word-form and the word-form itself in phonemic transcription.

(IIIa) The transition from a DMR of a word-form to the SMR of this word-form, i.e., to a string of morphemes and morphological operations; this is done by means of rules such as:

$$T_{(S,\,p)\,y,\,z} \Leftrightarrow \{T\}_{(p)} + \{YZ\}.$$

(*T* is a stem; *y*, *z* represent number and case; *p* are combinatorial properties of the stem supplied by the lexicon and called 'syntactics'.)

$$\text{PRIEM 'reception, welcome'}_{(p)\,\text{sing, acc}} \Leftrightarrow \{\text{PRIEM}\}_{(p)} + \{\text{SING. ACC}\}$$

$$T_{(A,\,p)\,x,\,\text{sing, short}} \Leftrightarrow \{T\}_{(p)} + \{\text{SHORT. X. SING}\}$$

$$\text{TVERDYJ}_{(p)\,\text{neut, sing, short}} \Leftrightarrow \{\text{TVERDYJ}\} + \{\text{SHORT. NEUT. SING}\}$$

$$T_{(V,\,p)\,\text{perf. fut}} \Leftrightarrow \{T\}_{(p)} + \{\text{PERF}\} + \{\text{PRES}\}$$

(since the perfective future in Russian is formally constructed like the present).

(IIIb) The transition from the SMR of a word-form to its phonemic transcription; this is performed by means of four groups of rules (Mel'čuk, 1967b, d, 1968; Es'kova *et al.*, 1971):

(1) Morphemo-morphic rules of the type

{PRIEM 'welcome'}$_{(p_1)}$ ⇔ /pr,ijom/$_{(p_1)}$,
{TVERDYJ 'firm'}$_{(p_2)}$ ⇔ /tv,ord/$_{(p_2)}$,
{SING. ACC} ⇔ /∅/ | either decl. II, masc, inanimate (*dom* 'house') or decl. III, masc/fem (*put'* 'way, road', *noč'* 'night')
⇔ /a/ | either decl. II, animate (*kota* 'cat') or decl. III, neut. (*vremja 'time'*)
{SHORT. NEUT. SING} ⇔ /o/
{NAXODIT' 'find'} + {PERF} ⇔ /najd/ | *not* past
⇔ /našed/ | past and part
⇔ /našod/ | past and *not* part
{IND. NON-PAST 3 SING} ⇔ /of/ | conj. I
⇔ /it/ | conj. II
{ON 'he'} + {SING. NOM} ⇔ /on/
{ON 'he'} ⇔ /j/ | *not* nom | *not* after a preposition
⇔ /n/ | *not* nom | after a preposition.

(2) Accentuation rules; perform transformations of the type:

/pr,ijom + ∅/ ⇒ /pr,ijóm + ∅/
/tv,ord + o/ ⇒ /tv,órd + o/.

(3) Morphonemic rules; perform different kinds of morphologically conditioned phoneme alternations:

/d/ ⇒ /ž/ (/tv, órd/ – /tv, órže/)
/c/ ⇒ /č/ (/pt, íca/ – /pt, íčka/)
{/t/, /d/} ⇒ Λ (/pl, ot + u/ – /pl, ó + 1/; /v, od + ú/ – /v, ó + 1/)
| in certain morphs only, according to syntactics of these.

(4) Phonological rules; perform morphologically unconditioned phonemic transformations like the following:

$$/C_{[\text{voiced}]}\{C_{[-\text{voiced}]}\}/ \Rightarrow /C_{[-\text{voiced}]}\{C_{[-\text{voiced}]}\}/$$

IV. *The phonological component* establishes the correspondence between the phonemic and the phonetic transcription of a word-form.

V. *The graphico-orthographic component* establishes the correspondence between the phonemic transcription of a word-form and its spelling. Its rules have the form:

/Xja/⇔/X/ A		(/struja/ *струя*)
/Xju/⇔/X/ Ю	/X/ is not a consonant	(/strujú/ *струю*)
/Xji/⇔/X/ И		(/strují/ *струи*)
/jX/⇔Й /X/	/X/ is not a vowel	(/kraj/ *край*)
/C,j/⇒/C/ Ъ /j/	/C/ is a consonant	(/sv,inja/ *свинья*)
/C,a/⇒/C/ Я		(/m,áso/ *мясо*)

IV. SOME LINGUISTIC IMPLICATIONS OF THE 'MEANING⇔TEXT' MODEL

Our work on the 'Meaning⇔Text' model has resulted in a number of interesting linguistic problems. We will only mention three of them here (cf. Mel'čuk and Žolkovskij, 1970; p. 40–46).

(1) One of the most important components of the MTM is the *lexicon* which stores all the information (semantic, syntactic, morphological, phonological) about each word necessary for all the components of the model to be able to handle this word correctly. Such a lexicon, or dictionary, known as an *explanatory combinatorial dictionary*, is at present being developed for Russian; see in particular Apresyan *et al.* (1969). Two entries from this type of dictionary are given below by way of illustration (due to lack of space, without any preliminary explanations. We trust the examples will be sufficiently obvious).

1. *OBEŠČAT' 'To promise'*

OBEŠČAT', *obéščan, ju, ješ'*, imperf. 1. *X obeščaet Y Z-u* 'X promises Y to Z' = X explicitly causes Z to know that Y, with which Z is concerned and which depends on X, will occur. Cf. GARANTIROVAT' 'guarantee'; UGROŽAT' 'threaten'.

1 = X who	2 = Y what	3 = Z to whom	
S_{nom}	(1) S_{acc} (2) V_{inf} (3) *čto* + SENT	S_{dat}	(1) X and Z are persons, (2) If Y = 2.2, then M_1 (Y) = X [M_1 (Y) is the first place, or first actant, of Y].

Ivan obeščal Petru knigu 'Ivan promised Peter a book'; *Ivan obeščal (Petru) dostat' (emu) knigu* 'Ivan promised (Peter) to get (him) ⟨he would get (him)⟩ a book'; *Ivan obeščal, čto kniga budet u Petra zavtra že* 'Ivan promised that Peter would have the book (no later than) tomorrow'.

Syn: arch. bookish *sulit'* 'promise'; colloq. *obeščat'sja*	Promise something in the near future ['tomorrow'] without fulfilling these promises:	*kormit'* [$S_{acc}=Z$] *zavtrakami* 'feed somebody with (false) hopes'
$Syn_\cap$: *kljast'sja* 'vow'; *zaverjat'* 'assure'; *zarekat'sja* 'promise not to, renounce'; *objazyvat'sja* 'pledge oneself'	Promise something instead of doing it:	*kormit' obeščanijami* 'feed with hopes'

[$Syn_\cap$ stands for an inexact ('semantically intersecting') synonym]

$S_0=S_2$: *obeščanie* 'a promise'	One cannot rely on the fulfillment of that which has been promised:	*obeščannogo (govorjat) tri goda ždut* (proverb) lit. 'you have to wait 3 years for what you've been promised' = 'words are wind'
Sing = Perf: colloq. *poobeščat'* 'have promised' $Magn_2^{intens}$: *tverdo* 'firmly'	$Magn_2^{quant}$ + $AntiVer_2$: *zolotye gory*//colloq. *naobeščat'* [S_{acc}], *naobeščat' s tri koroba* 'promise the sun and the stars, make a lot of (empty) promises' [X promises a lot of Y, but the speaker does not believe that the probability of Y is great]	

(2) *X obeščaet Y Z-u* 'X promises Y to Z' = X causes Z to conclude that Y, which is connected with X by cause-and-effect relations, will occur.

1 = X	2 = Y	3 = Z	
S_{nom}	(1) S_{acc} (2) V_{inf} (3) čto + SENT obligatory	S_{dat} rare	(1) X is not a living being, Z is a person, (2) If Y = 2.2, then $M_1(Y) = X$, (3) 2.1: S is a predicate.

Begstvo obeščalo nam spasenie 'Flight promised us salvation', *Den' obeščal byt' teplym* 'The day promised to be warm'.

$Syn_{\cap}$: *predveščat'* 'portend, foreshadow'; *predskazyvat'* 'foretell, predict'.

A_1 (promise a lot of something good): *mnogoobeščajuščij* 'promising, hopeful'.

2. *OBEŠČANIE 'A promise'*

OBEŠČANIE, *ja*, neut. $S_{0,2}$ *(obeščat' 1)* [the fact that something is being promised and also that which is promised]. Cf. GARANTIJA 'guarantee'; UGROZA 'threat'.

1 = X whose	2 = Y what	3 = Z to whom	
(1) S_{gen} (2) A_{poss}	(1) S_{gen} (2) V_{inf} (3) čto + SENT	S_{dat} rare	(1) X and Z are persons, (2) If Y = 2.2, then $M_1(Y) = X$, (3) Y = 2.1, if O. depends on a verb-LF, [*dal obeščanie prixoda* 'gave promise of arrival', etc.], (4) Impossible: 1.1, 2 + 2.1.

Obeščanie Petra pridti 'Petr's promise to come'; *moe obeščanie, čto kniga budet zavtra* 'my promise that the book will come tomorrow'; *obeščanie vstreči* 'promise of a meeting'; but **ego obeščanie vstreči* 'his promise of a meeting' is impossible.

$Syn_{\cap}$:
bookish *posul* '(lavish) promise'

$Syn_{\cap}$:
kljatva 'vow'; *zaverenie* 'assurance'; *slovo* 'word';
objazatel'stvo 'obligation'; *obet* 'vow'; *zarok* 'pledge'

V_0:
obeščat' 1 'to promise'

$Magn_1$:
toržestvennoe 'solemn'

$Magn_2$:
tverdoe 'firm'

$AntiVer_2$:
pustoe 'empty'; slang *lipovoe* 'fake, phony'

$AntiMagn + AntiVer_2$:
legkomyslennoe 'flippant, frivolous'

$Magn_2^{quant} + AntiVer_2$:
širokoveščatel'nye 'wide, alluring' | O. in plural

$Oper_1$:
davat' [$S_{dat} - e$] 'give'

$Liqu_1Oper_1$:
brat' obratno ⟨*nazad*⟩ [(*svoe*) − *e*] 'take back'

$Perm_3Liqu_1Oper_1$:
osvoboždat [$S_{acc} = X$ *ot* − *ja*] 'release from'

$Oper_3$:
polučat [*ot* $S_{gen} - e$] 'receive from'

$Oper_1$(Magnquant + AntiVer):
davat' kuču [−*j*] 'give a bunch of, a lot of'

$Func_2$:
sostojat' [*v* S_{prep}] 'consist in'

$Real_1$:
vypolnjat', deržat', sderživat' [(*svoe*) − *e*] 'keep one's promise' | D_2 *(O.)* is an action of *X*

not $Real_1$:
ne vypolnjat', ne deržat', ne sderživat' [(*svoe*) − *e*]
'not fulfill, keep one's promise' [not do what one promised],
narušat [(*svoe*) − *e*] 'break one's promise' [do what one had promised not to do]
| D_2(*O.*) is an action of *X*

$Fact_0$:
sbyvat'sja 'come true, be realized' | D_2(*O.*) is not an action of *X*

(3) Our aim of investigating, within the framework of our work on the MTM, the relations between meaning and text, i.e. between signifiés and signifiants of linguistic units, has made it possible to take a new approach towards problems of word-derivation and treat it against the background of the universal picture of all the relations possible between words with respect to their meaning and form. Since both the signifiés ('*A*', '*B*'), and the signifiants (*A*, *B*) of two lexemes can (1) coincide (=), (2) be contained one within the other (⊃), (3) intersect (∩) and (4) have no part in common, we have 17 possible formal-semantic relations (Mel'čuk, 1968).

Both 'classical' facts, i.e., specifically, homonymy ($A = B$, $'A' \cap 'B' = \emptyset$), absolute synonymy ($A \cap B = \emptyset$, $'A' = 'B'$, 'normal' word-derivation ($A \supset B$,

'$A'\supset$'B') etc., and certain little studied phenomena such as 'contrary word-derivation': $A \supset B$, but '$A' \subset$ 'B', are provided for in 'cells' of the deductively derivable system. Examples of contrary derivation:

(1) *radovat'sja* (be glad of) 'experience emotion *X*' – *radovat'* (make glad of) 'cause to experience emotion *X*' (and many other similar pairs);

(2) *geolog-ij-a* (geology) 'the science of the Earth...', – *geolog* (geologist) 'a specialist in the science of the Earth...' (and many other pairs);

(3) *A tverž-e B* (*A* is harder than *B*) 'the hardness of *A* exceeds the hardness of *B*' – *A tverd* (*A* is hard) 'the hardness of *A* exceeds the norm established for the objects of the class *A*'. (*N.B.*: *tverže* $\neq$ *bolee tverdyj*, i.e. in Russian there exists a difference between 'harder' and 'more hard'): *A bolee tverdyj, čem B* = '*A* is hard, and *A* is harder than *B*'; but from *A tverže B* it does not follow that *A* is hard: *A* can be soft, but still harder than *B*.)

At the same time, such relations as *moskvič:moskvička* = *laborant: laborantka* can of course also be easily described, since *moskvič* $\subset$ *moskvička*, but 'moskvič' $\not\subset$ 'moskvička' ('moskvič' = 'a male resident of Moscow', 'moskvička' = 'a female resident of Moscow'), while *laborant* $\subset$ *laborant-k-a* and 'laborant' $\subset$ 'laborantka' (*laborant* = 'a technical laboratory worker', *laborantka* = 'a woman, who is a technical laboratory worker'), etc., as is, in particular, English conversion (see below).

(4) Work on the MTM has led the author to think of a linguistic sign as being *three-dimensional* entity, or an ordered triple $A = \langle A, \text{'}A\text{'}, \sum_A \rangle$, where A is the signifiant, 'A' the signifié, and $\sum_A$ is all information about the combinatorial properties of the sign, which in their totality may be spoken of as *syntactics*, cf. above, page 50. (Information about what part of speech a lexeme is; the gender of nouns, lexical functions, etc. belong to syntactics.) If we consider syntactics as an individual (together with the signifié and signifiant) component of the linguistic sign, we can provide a natural enough formal descriptions of phenomena such as the English *the cook – to cook*, known as conversion. We might define conversion as a linguistic sign whose signifiant is an *operation on the syntactics* of other signs (cf. meaningful alternation: a sign whose signifiant is an operation on the signifiants of other signs).

Examples of conversion:
$K_1 = \langle V \Rightarrow N$, 'he who...', $\sum_{K_1} \rangle$ *(to cook – the cook; to bore – the bore)*;
$K_2 = \langle N \Rightarrow V$, 'cause to act upon...', $\sum_{K_2} \rangle$ *(the bomb – to bomb; the machine gun – to machine-gun)*.

It seems possible to construct a calculus of all conceivable morphological means, or processes, in natural languages.

ACKNOWLEDGEMENTS

The first draft of this paper has been read by and discussed with S. J. Fitialov, A. J. Dikovskij, O. S. Kulagina, N. G. Mixajlova, L. S. Modina, and M. K. Valiev; for helpful comments on the final version I am indebted to Ju. D. Apresjan, L. N. Iordanskaja, N. V. Pertsov, V. J. Rozencvejg, and A. K. Žolkovskij. It is with great pleasure that I acknowledge here their valuable help and sympathetic understanding. Last, but not least, I would like to express my gratitude to Ferenc Kiefer for everything he has done in order to have this paper published in English.

Soviet Academy of Sciences, Moscow

BIBLIOGRAPHY

Apresyan, Ju. D., Mel'čuk, I. A., and Žolkovskij, A. K.: 1969, in F. Kiefer (ed.), 'Semantics and Lexicography: Towards a New Type of Unilingual Dictionary', *Studies in Syntax and Semantics*, Dordrecht, pp. 1–33. (Rev.: 1972, *Language* **48**, No. 3.)

Gladkij, A. V. and Mel'čuk, I. A.: 1969, *Tree Grammars*, International Conference on Computational Linguistics, Stockholm. Preprint No. 1.

Gladkij, A. V. and Mel'čuk, I. A.: 1971, in 'Informacionnye voprosy semiotiki, lingvistiki i avtomaticeskogo perevoda', *Grammatiki derev'ev, I. Opyt formalizacii preobrazovanii sintaksičeskix struktur estestvennogo jazyka*, M., pp. 16–41.

Es'kova, N. A., Mel'čuk, I. A., and Sannikov, V. Z.: 1971, *Formal'naja model' russkoj morfologij. I. Formoobrazovanie suščestivtel'nyx i prilagatel'nyx* (Russian Language Institute, AN S.S.S.R.), M.

Mel'čuk, I. A.: 1967a, *Ordre des mots en synthèse automatique des textes russes*, TA informations, No. 2, pp. 65–84.

Mel'čuk, I. A.: 1967b, *Model' sprjaženija v ispanskom jazyke. Mašinnyj perevod i prikladnaja lingvistika* **10**, 21–53;
also in: *Soviet Papers in Formal Linguistics* (V. Y. Rosencveig, ed.), Vol. I., Atheneum, Frankfurt, 1973.

Mel'čuk, I. A.: 1967c, in 'To Honor Roman Jakobson...', *K voprosu o 'vnešnix različitel'nyx élementax: semantičeskie parametry i opisanie leksičeskoj sočetaemosti*, pp. 1340–1361.

Mel'čuk, I. A.: 1967d, in 'Problemy strukturnoj lingvistiki', *Model' sklonenija vengerskix suščestvitel'nyx*, M., pp. 344–373.

Mel'čuk, I. A.: 1968, *Stroenie jazykovyx znakov i vozmožnye formal'no-smyslovye otnošenija meždu nimi.* – Izvestija AN S.S.S.R, *Literature and Language Series* **27**, 426–438.

Mel'čuk, I. A.: 1970, 'Towards a Functioning Model of Language', in *Progress in Linguistics*, The Hague-Paris, pp. 198–207.

Mel'čuk, I. A. and Žolkovskij, A. K.: 1970, 'Towards a Functioning 'Meaning–Text' Model of Language', *Linguistics*, No. 57, pp. 10–47.

Žolkovskij, A. K. and Mel'čuk, I. A.: 1967, 'O semantičeskom sinteze', *Problemy Kibernetiki* **19**, 177–238.

Žolkovskij, A. K. and Mel'čuk, I. A.: 1970, *Sur la synthèse sémantique*, TA Informations, 1–85 [translation of the preceding article].