Semantic Primitives from the Viewpoint of the Meaning-Text Linguistic Theory *

1989

by Igor A. Mel'čuk

To Alik [1], who a quarter of a century ago planted in my brain the seeds of semantic ideas, with warmth and gratitude.

The expression «Semantic Primitives» has come to be a buzz phrase in all domains of language- or thought-related research. People seem to find special fascination in semantic primitives, so that audiences start to warm up to a presentation in semantics, linguistic behavior, language understanding and the like as soon as the speaker indicates that his explorations might eventually lead to semantic primitives. «Oh, oh, Semantic Primitives! . . .» they say with mystical awe, turning their eyes towards Heaven.

But at the same time, semantic primitives are nowadays a hotly debated issue. There is no concensus even as to the most essential points concerning them. Do semantic primitives exist? If so, do we linguists need them? If so, what do we need them for? What are they? Where are they to be found? And so forth. (In this respect, the phrase «semantic primitives» reminds me of such expressions as «the meaning of life».) I think that in the ensuing discussions much effort was wasted and many claims missed the point because of *concep-*

* This paper has been written and prepared for publication with the help of a grant from the Canadian Federation for the Humanities, using funds provided by the Social Sciences and Humanities Research Council of Canada (Grant # 411-85-0001).

[1] Alik, *alias* Alexander K. Zholkovsky (Žolkovskij, in his previous life), is now Professor of Slavic Languages and Literatures, University of Southern California, Los Angeles. He turned 50 on Sept. 8, 1987, and on this occasion, I am joining all his friends and admirers who whish him a happy birthday, as well as many no less happy returns of the day. – It was Zholkovsky who actually launched semantic research in the URSS in 1960 (see Žolkovskij [1961, 1964a, b, c] Žolkovskij et al. [1961]); he probably was the first person to propose a formal semantic representation of utterances based on predicate calculus and to present a large series of semantic decompositions of Russian lexemes («vocabulary of purposeful activity») which seem to be quite valid 25 years later. I had the priviledge of having closely collaborated with Zholkovsky for more than 10 years; the groundwork for the Meaning-Text theory, the corresponding linguistic model and the Explanatory Combinatorial Dictionary – see below – was laid down by us together (Žolkovskij and Mel'čuk [1967], Mel'čuk and Žolkovskij [1970], Mel'čuk and Zholkovsky [1984]). I owe Alik Zholkovsky so much that my indebtedness cannot be properly acknowledged in details; therefore let this note be an over-all acknowledgement of Zholkovsky's profound influence on my theoretical approach in general and on my semantic explorations in particular.

tual vagueness. With the notable exception of A. Wierzbicka's impressive work (to quote only her books Wierzbicka [1972, 1980, 1985, 1987]), as far as I know, there have been in modern theoretical linguistcs no systematic attempts to outline the notion of semantic primitives or, a least, to stake out where to hunt for them. What was probably the first workable list of semantic primitives for linguistic use was suggested 25 years ago by A. Zholkovsky [Žolkovskij 1964c: 91-92]; these primitives were applied in the description of the meaning of about 150 Russian words [2]. Since then, a few other people working in the area (among others, Wierzbicka and Apresjan) have been using similar lists, trying to improve on Zholkovsky's list or reduce its size (cf., e.g., 15 semantic primitives of Wierzbicka's below: footnote 2). Nonetheless, we still lack a solid, logical discussion of the *concept* itself of these evasive entities. Yet what might be appropriately called «semantic primitives» in Artificial Intelligence (= minimal units of behavior) need not necessarily be identifical with what are semantic primitives in Linguistic Semantics (= minimal units of linguistic meaning). Therefore, I believe that a study of semantic primitives should be based upon the following thesis:

A clear concept of semantic primitive is needed, and this concept should be deli-

[2] Let me adduce here Zholkovsky's 23 semantic primitives, since I think that they are still interesting, and not only for historical reasons. I will quote them in literal English translation and a slightly modified form:

1) set [in the mathematical sense]
object
property
relation
include

2) no
and
or
if-then
identical
true

3) all
there is
more [than]

4) dimension
number
time
space

5) person
want
think

6) speaker [= I]
norm

For purposes of quick comparison, I will also quote Wierbicka 15 semantic primitives (as they appear in 1989; cf., e.g., Wierzbicka [1987: 31]:

1) I
you

2) this

3) someone
something

4) want
not want

5) think of
imagine
know
say

6) place
time

7) become
part

This list of primitives obviously embodies a different approach: emphasis of what might be called «human factor», attempt at universality, use of basic natural-language notions, etc. However, I cannot delve here into a substantial analysis and comparison of these two sets of semantic primitives.

Historical fairness requires to mention Zholkovsky's predecessors but there is no space for a serious bibliographical study. Let me limit myself to the indication of the interesting work done in this connection, in the late fifties and early sixties, by the members of the Cambridge Language Research Unit, headed by Margaret Masterman; see, e.g., Masterman [1961], where a list of 100 semantic primitives is proposed. – For a good coverage of the prehistory of semantic primitives in linguistics, see Wierzbicka [1972: 1-12].

neated and examined within a precise framework, i.e., within a well specified conceptual system, which must be carefully defined PRIOR to any theoretical discussion of semantic primitives proper.

Consequently, the present paper concentrates on one such possible framework: namely, the MEANING-TEXT THEORY (MTT) of natural languages. More specifically, I will need several concepts related to what is called MEANING-TEXT MODEL (MTM), in particular to formal representations of utterances and to the MTM's components; in order to introduce them, I have to talk first about the MTT in general. As a result, the paper has the following structure:

1. The Meaning-Text Theory in Brief Outline.
2. The Meaning-Text Model: Representations and Components.
3. Semantic Decomposition and Semantic Primitives.
4. The Explanatory Combinatorial Dictionary and Theoretical Semantics.
5. Semantic Primitives: Conclusions.

Let me emphasize that because of limitations of space, my characterization of the MTT and the MTM will be of necessity sketchy and without proper justification [3]. For the same reason, I will not consider, even briefly, the problem of semantic primitives and semantic decomposition as treated under componential analysis (cf., e.g., Nida [1975] and Lyons [1977: 317-335]) and within the generative transformational school, especially by generative semantics (cf., e.g., Fodor [1980: 144-155]). I will indicate, nevertheless, two earlier little known papers: Kay and Samuels [1975], because it proposes a list of 36 semantic primitives, which are then applied in the lexicographic definitions of a couple hundred English words, and Hofmann [1974], which makes a strong case for «semantic atoms».

I draw basically on the work done within the Meaning-Text theory, as well as on Anna Wierzbicka's work. She and Jurij Apresjan are my main sources of inspiration, and I may have many more unconscious quotations from them than I have managed to aknowledge.

1. *The Meaning-Text Theory in Brief Outline*

The basic idea underlying the Meaning-Text linguistic theory is as follows. Any act of linguistic communication is believed to involve three major entities:

(a) A CONTENT to be communicated by linguistic signals; we will refer to this as *meaning*.

(b) A complex SIGNAL (acoustic or graphical), to be used in order to communicate the content; we will refer to this as *text*.

[3] As indicated in note 1, the MTT and the MTM were put forward by A. Zholkovsky and the present writer 23 years ago. Later we were joined by Jurij Apresjan and several other people, who collaborated actively in the development of the theory and the proposed model. – Basic readings on MTT and MTM include, besides the titles mentioned in note 1, Mel'čuk [1974, 1981, 1988a], Apresjan [1974, 1980], Mel'čuk and Pertsov [1987].

(c) A MAPPING, or a set of correspondences, between meanings and texts; this is nothing else but a (natural) *language*. (This is so because basic correspondences are established between ELEMENTARY meanings and ELEMENTARY texts, both of which are finite in number.)

Note that:

- We presuppose the DISCRETE character of both linguistic meanings and texts. They are taken to be formal objects describable by means of formal languages and specifiable by a formal device.
- The set of meanings and the set of texts are INFINITE (but denumerable); the set of correspondences between them is, on the contrary, FINITE.
- The correspondence between a specific set of meanings and a specific set of texts is, as a rule, MANY-TO-MANY. One meaning can be expressed by quite a few texts (synonymy), and a text can express several meanings (ambiguity) [4].

Given this basic idea, the three postulates of the Meaning-Text theory can be formulated.

POSTULATE I

A natural language L is a finite set of many-to-many correspondences between a denumerable set of meanings and a denumerable set of texts.

Symbolically:

$$(1)\ \{\text{MEANING}_i\} \overset{\text{language}}{\Longleftrightarrow} \{\text{TEXT}_j\} \mid 0 < i, j < \infty,$$

where i and j are positive integers.

[4] *Synonymy vs. Ambiguity*. It should be most strongly emphasized that synonymy and ambiguity in natural language are by no means symmetrical:

- Synonymy is extremely important and useful for natural language; in a sense, a language is a system for production of synonymous utterances. Ambiguity, on the contrary, is not only useless (its only purposeful use is in puns) but harmful – and has to be constantly fought against. It is *un mal nécessaire*, while synonymy is the essence itself of any language.
- Synonymy is rarely typical of short linguistic items: thus, exact synonymy of lexemes or phrases is infrequent, whereas it is extremely widespread on the level of sentences or paragraphs. Ambiguity, on the contrary, is mostly related to short items: morphs, lexemes and phrases tend to be ambiguous; ambiguity of sentences is less frequent, and ambiguity of whole paragraphs is hardly possible to all.
- Constructing synonymous expressions for a given meaning is a purely linguistic task. Resolving ambiguity of an expression is, on the contrary, mostly a logical or encyclopedic task: to choose the intended meaning among a few possible ones requires real understanding of the extralinguistic situation.

I could quote other differences as well, but I think that the above is sufficient to make my point. Because of this asymmetry, the Meaning-Text theory is oriented toward synonymy and (more or less) ignores ambiguity. The SYNONYMY RELATION is taken as basic, or intuitively obvious, and the Meaning-Text model of natural language is being developed proceeding from synonymy of utterances. Synonymy/quasi-synonymy/non-synonymy of linguistic expressions is the primary tool of our research. The over-all importance of synonymy for the Meaning-Text approach will emerge many times in the present paper. – Interestingly, logicians have long since understood the importance of synonymy for semantics; cf. e.g., Quine [1953: 22]: «. . . the primary business of the theory of meaning [is] . . . the synonymy of linguistic forms and the analyticity of statements». In linguistics, the primary role of synonymy for semantics was strongly emphasized by Zholkovsky [1964a: 12-13].

As far as *meaning* (in our technical sense) is concerned, the following holds:

- We deal with strictly linguist meaning – the shallowest, literal meaning of utterances which can be arrived at solely on the basis of linguistic knowledge, without any reference to the extralinguistic context, encylopedic information or common sense. In the MTT, meaning is no more than the INVARIANT OF SYNONYMOUS PARAPHRASES. Thus the concept of meaning is derived from the concept of «the same meaning», that is, from the concept of synonymy (this latter being considered indefinable; cfr. footnote 4).
- Meaning is taken to be DIRECTLY ACCESSIBLE to speakers, i.e. to be a part of their intuitive knowledge of their language, much like its sounds; therefore, meaning belongs to the linguist's data.
- In actual fact, the MTT deals with FORMAL REPRESENTATIONS of meanings, called Sem(antic) R(epresentation)s, rather than with real meanings. Developing a formal language for the description of meaning – a SEMANTIC TRANSCRIPTION, or METALANGUAGE – is, from the standpoint of the MTT, one of the primary tasks of linguistic research.

The same is true, *mutatis mutandis*, for *texts* (once again, in the proposed technical sense): a text is a linguistically valid segment of speech, be it a morph, a wordform, a phrase, a sentence, or a paragraph).

(1) can be now rewritten in a more precisely as (2):

$$(2)\quad \{\mathrm{SemR}_i\} \overset{\text{language}}{\Longleftrightarrow} \{\mathrm{PhonR}_j\} \mid 0 < i, j < \infty,$$

where i and j are positive integers.

Phon(etic) R(epresentation) stands for text.

From Postulate 1 it follows that linguistcs should aim at describing correspondences of type (2). However, while meanings and texts are immediately accessible to a linguist, the correspondence between them is not. Linguists face here the classical «black box» situation: we perceive and control the inputs and outputs of language **L**, i.e. its meanings (SemRs) and its texts (PhonRs) but not its inner circuits linking inputs with outputs; in other words, we do not have access to LINGUISTIC RULES, which are encoded in the speaker's brain and implement the mapping diagrammed in (2). The most obvious option we have in such a situation can be expressed by the following postulate:

POSTULATE 2

A natural language **L**, viewed as mapping (2), can be described only by a cybernetic, or functional, model.

A *functional model of a language* **L** is a system of formal rules which simulates the linguistic behavior of **L**'s speakers, i.e., which associates, on the one hand, with a given SemR of **L** all the PhonRs that (according to the speakers) can carry the corresponding meaning and, on the other hand, with a given

PhonR of **L** all the SemRs the corresponding text can express. Such a model is called a MEANING-TEXT LINGUISTIC MODEL of **L**. The Meaning-Text linguistic theory is a theory for building MTMs of actual languages – that is, for writing specific mappings of type (2).

The said mappings, however, are, as was already stated, many-to-many, and *many* here means really 'many'. Thus, sentence (3) has more than 200000 paraphrases [Mel'čuk 1981: 31-32]:

(3) *The Food and Drug Administration has seriously cautioned expectant mothers to avoid one of life's simple pleasures: a cup of coffee.*

To put it differently, the SemR of (3) corresponds to about 200000 different PhonRs; for instance:

(3') a. *The FDA has strongly warned pregnant women against one of the small joys of life: drinking coffee.*

b. *The FDA has issued an earnest warning to all women expecting a baby: they should not indulge in consuming coffee, which is among life's simple joys* [5].

Similarly, a PhonR (= representation of a text) may give rise to quite a few SemRs, i.e., a text may have several meanings; since this phenomenon – ambiguity – is too well known, I will not illustrate it here. (Concerning the essential asymmetry of synonymy *vs* ambiguity, see footnote 4.)

The extremely involved character of the meaning-text correspondence makes it practically impossible to write linguistic rules for a DIRECT meaning-to-text

[5] *Paraphrases in Natural Language.* The following important remark seems necessary at this point. I take sentences (3) and (3') to be exact paraphrases of each other, i.e. synonymous. There are, however, many linguists who believe that there are no real paraphrases, no really synonymous linguistic expressions: to them, in natural language, every difference in form entails a difference in meaning. (An excellent exposition of this viewpoint can be found in Bolinger [1977].) In other words, given two different expressions, we always can find a semantic difference between them, provided we look hard enough.

Since the whole Meaning-Text approach is built upon the notion of paraphrase, which amounts to utterance synonymy, I have to address the above view and explain why I still stand my ground.

The answer is simple: It is probably true that by looking hard enough we can find a semantic difference between two given expressions in most cases, even if not always. But why should we look hard enough? In their actual linguistic behavior, speakers are normally not overly precise. If we look hard enough, not only will there be no two identical meanings, but no two identical lengths, weights, or speeds either! Two pairs of shoes of the same size are in fact never of the same size; and two 5-pound bags of potatoes can be easily shown to have different weights. This is, however, without interest: within the limits of a specific problem, it is harmful to be too precise. (You should not weigh potatoes using a pharmacist's scale.) In the same vein, in practical instances of linguistic communication, people do not exploit all semantic possibilities of their language. A SEMANTIC NEUTRALIZATION takes place rather often (cf. Apresjan [1974: 156-163, 239-242, 281-283]). Capitalizing on this fact, we accept the existence of real paraphrases within the limits of a given degree of precision. To put it slightly differently: we agree to disregard potential distinctions in contexts where they are not, so we believe, actually exploited. Some semantic differences can perhaps be unearthed between the sentences in (3) and (3'); I think, however, that for the purposes of a journalist writing for *Newsweek* on the given topic they are fully equivalent.

transition. To make the task more tractable, we break down the above correspondence into a number of simpler correspondences, which entails the introduction of INTERMEDIATE LEVELS of representation. Capitalizing on the fact that natural language objectively displays sets of special organizational properties centered around two basic linguistic units – the sentence and the word, we posit the following postulate:

POSTULATE 3

To describe the correspondence $\{SemR_i\} \Longleftrightarrow \{PhonR_j\}$ in **L**, two intermediate levels of representation are introduced: Synt(actic) R(epresentation) and Morph(ological) R(epresentation).

Therefore, (2) can be rewritten, in a more developed form, as (4):

(4) $\{SemR_i\} \Longleftrightarrow \{SyntR_k\} \Longleftrightarrow \{MorphR_l\} \Longleftrightarrow \{PhonR_j\}$

semantics **syntax** **morphology + phonology**

In (4), I have also indicated in boldface the divisions, or COMPONENTS, of natural language that ensure specific subcorrespondences. Thus, the semantics of **L** is that component of **L** that maps its meanings (= **L**'s SemRs) onto its sentence structures (= **L**'s SyntRs), and so forth. Each of these components corresponds to one of the components of **L**'s Meaning-Text Model: to the Sem- the Synt- and the Morph-Phon-component, respectively.

However, it turns out that in order to describe the meaning-to-text correspondence in a natural and elegant way, we need subtler distinctions in the utterance representations used, as well as in the components of the model. Namely, the syntactic, the morphological and the phonetic levels have to be split into two sublevels each: a deep (D-) one, geared to meaning, and a surface (S-) one, geared to text, i.e. to actual linear expressions. As a result, (4) appears as (5):

(5) $\{SemR_i\} \Longleftrightarrow \{DSyntR_{k1}\} \Longleftrightarrow \{SSyntR_{k2}\} \Longleftrightarrow \{DMorphR_{l1}\} \Longleftrightarrow \ldots$

semantics **deep syntax** **surface syntax**

(For a more detailed discussion see Mel'čuk [1988a: 50 ff]).

In the present paper, only SemR and DSyntR are relevant, so that all the other representations will be ignored. Accordingly, only the Sem-component of the MTM will be considered.

These three postulates express the most general properties of the Meaning-Text theory. More specific points, such as directionality of the description (we work exclusively FROM the meaning TO text), concrete properties of various utterance representations, the structure of rules and the like, regardless of their utmost importance, cannot be dealt with here. The interested reader can consult the titles indicated in footnote 3.

A Meaning-Text Model is specified, first, by the set of utterance representations it uses, and second, by the set of rules that correlate said representations. A group of rules dealing with a given pair of representations constitutes a component of the MTM.

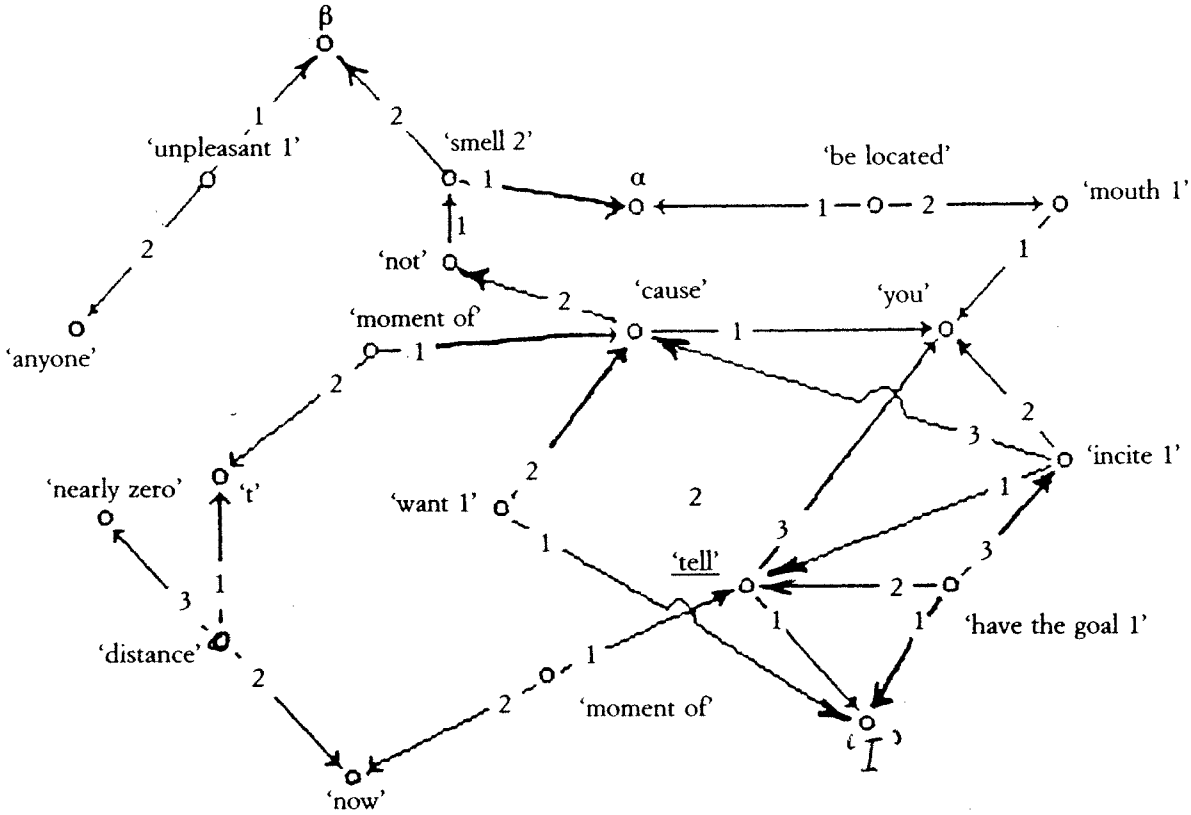
UTTERANCE REPRESENTATIONS. In the MTM framework, a representation is a set of formal objects called *structures*, each of which depicts separately and autonomously one aspect of the linguistic phenomenon represented. Thus, the SemR of an utterance consists of three structures:

- the *semantic structure*, specifying the meaning of the utterance as such (so to speak, the situational, or propositional, meaning);
- the *semantic-communicative structure*, specifying the organization of the message by the speaker (topic *vs.* comment, old *vs.* new, assertion *vs.* presupposition, and the like);
- and the *rhetorical structure*, specifying the «artistic» intentions of the speakers (the style and the character of the text to be produced: neutral, poetic, bureaucratic, journalese, etc.).

However, in this paper, I will cite only one structure per representation – the MAIN structure, – calling it, *par abus de langage*, «representation» (so that what is called the SemR below is in fact the semantic structure; and what is called the DSyntR is but the deep-syntactic structure).

Semantic Representation

(6) Let me start from an example



The SemR in (6) expresses the meaning of an advertisement for a new brand of toothpaste I have gleaned from a newsmagazine. It can be read in English LITERALLY, i.e. word for word, as follows:

(6') 'I tell you – having the goal that my telling incite you to do what I want you to – that I want you to cause that there be no unpleasant (for anyone) smell of something located in your mouth, the distance between the moment of my telling, which is now, and the moment *t* of your causing what I want you to cause being nearly zero (i.e., causing should follow 'now' immediately)'.

This meaning can be expressed in idiomatic English in many different ways, so that (6) is in fact the SemR of A WHOLE FAMILY OF (more or less) SYNONYMOUS SENTENCES (see examples in (7), p. 74).

Formally speaking, a SemR is an oriented acyclic connected graph – a NETWORK with labeled nodes and arcs.

A node of a SemR is labeled with a *semantic unit* of **L**, a semantic unit being a separate sense of a lexical unit, i.e. of a word or a set phrase (of **L**). A word taken in one well-specified sense is called a *lexeme* and its meaning, a *semanteme*. Thus the semantic units in (6) are semantemes of English, denoted with corresponding English lexemes. (The numbers that identify senses in our examples are borrowed from the *Longman Dictionary of Contemporary English*, 1978). The semantemes, as well as the meanings in general, are notated with single quotes. Thus, if **L** is a lexical unit, then '**L**' is its meaning, i.e. a semanteme. The underlining of an item – 'tell' in the lower right quarter of the diagram in (6) – identifies the «central», or «starting», point of the SemR. Thus, SemR (6) is about 'I tell you that . . .', and not about 'something located in your mouth').

Generally speaking, semantemes fall into two major types: *functors* (including predicates, quantifiers and logical connectives), and *names* (of (classes of) objects). A functor is characterized by the number and nature of arguments it can take; arcs are directed from a functor to its arguments.

An *arc* of a Sem is labeled with a number identifying the corresponding argument; for instance,

I
'cause' o → o 'you'

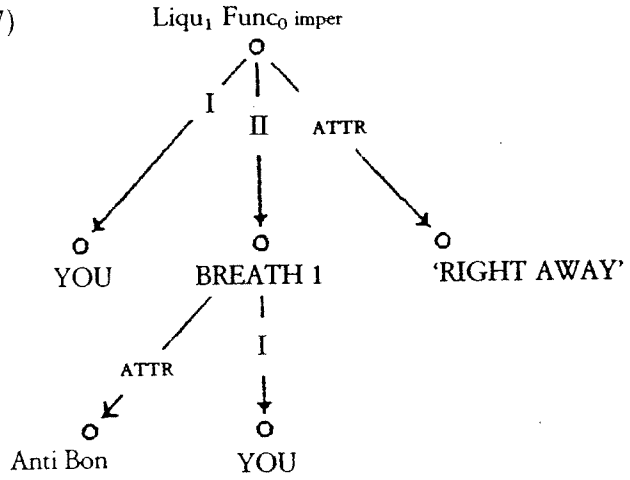
means that 'you' is the 1st argument of 'cause', i.e. the causer (= 'you cause . . .'). Thus, as one can easily see, the SemR of the MTM is based in predicate calculus.

Deep-Syntactic Representation

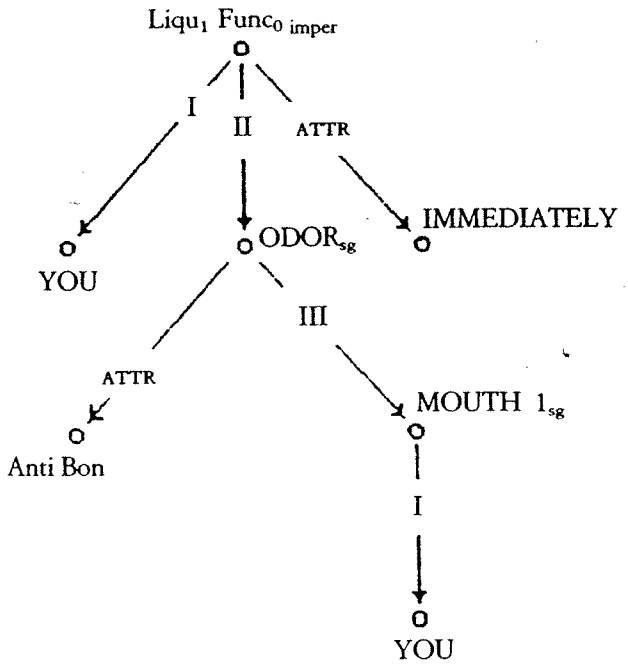
As has been indicated, SemR (6) can be expressed by several English sentences; to illustrate this, I will give here three synonymous DSyntRs (= the DSyntRs of the three synonymous sentences (7')) which correspond to SemR (6).

(7)

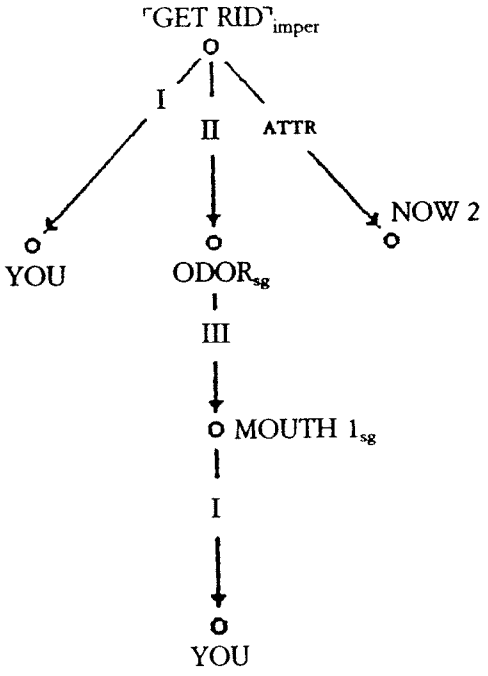
a.



b.



c.



The corresponding sentences are as follows:

(7') a. *Stop bad breath right away!*
b. *Eliminate foul mouth odor immediately!*
c. *Get rid of mouth odor now!*

Sentences (7' a-c) by no means exhaust the possibilities an English speaker has to express the meaning specified by the SemR in (6); he can, for instance, say (7'd) as well:

d. *Down with ⟨Beat, fight⟩ halitosis!*

Enough said on this topic for the moment.

Formally speaking a DSyntR is a particular case of a network – a DEPENDENCY TREE (a node of a tree receives only one arrow, and there is exactly one node – the *top node* – that receives no arrow at all).

A *node* of a deep-syntactic tree is labeled with a *deep lexical unit* of **L**, which can be: a full lexeme, a phraseme (like *right away, get rid, if worst comes to worst, kick the bucket*, etc.), or a lexical function.

A *lexical function* is a dependency **f** that associates with a given lexical unit L – its *argument*, of *key word* – a set $\{L_i\}$ of (more or less) synonymous lexical units – its *value* – that express, contingent on L, a specific meaning associated with **f**:

$$\mathbf{f}(L) = \{L_i\}$$

In natural languages there have been discovered about 60 standard elementary lexical functions (LF). Let us give a few examples.

FUNCTION ARGUMENT	VALUE
Magn (*shave*)	= *close, clean*
Magn (*easy*)	= *as pie, as a piece of cake*
Magn (*condemn*)	= *strongly*
Magn (*contrast*)	= *sharp; vivid*
$Oper_1$ (*cry*)	= *let out* [ART ~]
$Oper_1$ (*figure*)	= *cut* [ART ~] [*He cut a miserable figure*]
$Oper_1$ (*strike*)	= *be* [on ~]
$Oper_1$ (*attention*)	= *pay* [~]
$Real_2$ (*mine*)	= *strike* [ART ~] [*The car struck a land mine*]
$Real_2$ (*test*)	= *withstand* [ART ~]
$Real_2$ (*joke*)	= *get* [ART ~]
$Real_2$ (*attack*)	= *fall* [to ART ~]

[The symbol ART indicates the necessity of an article, which is chosen according to general rules of English].

In the DSyntR's of (7) we find two LFs: $LiquFunc_0$ means 'to cause not to be' (like in to *disturb* the peace of mind, to *remove* the doubts, to *scotch* rumors, . . .), and $Liqu_1Func_0$ specifies that the causer is necessarily the 1st actant of the key word: he causes something to himself; AntiBon means 'bad' (like in *foul* weather) [6].

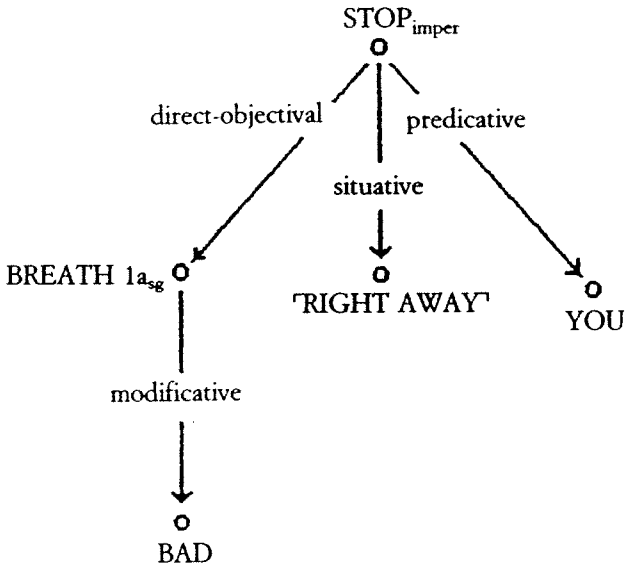
[6] For more on lexical functions, introduced in Žolkovskij and Mel'čuk [1967], see Mel'čuk [1982] and the previously mentioned titles.

A *branch* of a deep-syntactic tree is labeled with the name of a *deep-syntactic relation*; DSyntRels are universal and very few in number: COORD(inative), ATTR(ibutive) and six actantial DSyntRels I-VI.

Let it be emphasized that there is NO LINEAR ORDER in a deep-syntactic tree. Word order is considered to be a means for encoding the syntactic structure of a sentence, and, therefore, it cannot be part of the said structure. Linear order is also absent from the SSyntR: it appears only in the DMorphR of the sentence in question, computed by the rules of the SSynt-component (of the MTM) – on the basis of DSyntRels linking the sentence elements, lexicographic properties of these elements, data about the communicative organization of the sentence, etc.

The subsequent representations (SSyntR, DMorphR, etc.) will not be discussed here, but for the sake of better understanding I will quote the SSyntR and the DMorphR for the DSyntR (7a) [= sentence (7'a)]:

(8) a. The surface-syntactic representation of sentence (7'a)



b. The deep-morphological representation of sentence (7'a).

STOP$_{imper}$ BAD BREATH RIGHT AWAY!

As can be seen, the nodes of a surface-syntactic tree are labeled with actual lexemes of the sentence, and its branches, with the names of actual SSynt-relations obtaining among the lexeme occurrences (in **L**, in this case – in English). The DMorphR of a sentence is the linearly orderd sequence (= string) of its lexeme occurrences.

Let me briefly summarize three theoretically important differences between

semantic representation and deep-syntactic representation. I will do this by formulating their properties in parallel columns.

1. What is represented?

A SemR represents the meaning of a sentence of **L**, not the sentence itself. Neither the lexical inventory, nor the syntactic organization of the sentence is taken into consideration.	A DSyntR represents a given sentence of L, not its meaning. Both the lexical inventory and the syntactic organization of the sentence are reflected in its DSyntR as such.

2. The vocabulary used.

A SemR uses semantic units, i.e. senses of lexical units of L, which need not, however, be the lexical units of the sentence represented. The vocabulary of a SemR is language-specific – but not necessarily identical to that of the actual sentence in question.	A DSyntR uses lexical units of the sentence represented, although they may appear in a generalized form (e.g., as lexical functions). The vocabulary of a DSyntR is that of the sentence represented – except for the generalizations allowed.

3. The syntax used.

A SemR uses the syntax of labeled networks, the numbers identifying the arguments of the corresponding functors. This is thus a universal *a*linguistic syntax, based on a logical formal system.	A DSyntR uses the dependency syntax of universal syntactic relations, which are generalizations of actually observed syntactic relations in natural languages. This is thus a universal *inter*linguistic syntax.

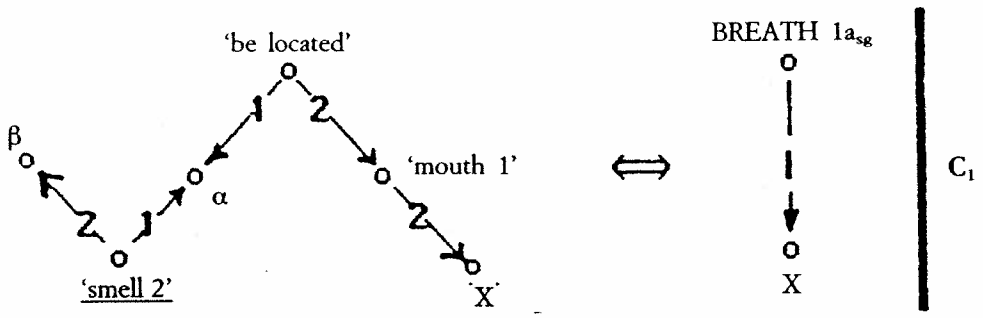
Components of the MT-Model. The MTM carries out both the transition from (6) to (7') and to all sentences synonymous with (7') and vice versa. This is done in a series of autonomous steps, only one of which is of interest to us here: the transition between the SemR (6) and the corresponding DSyntRs in (7), an operation which is the responsability of the semantic component of the model. What happens afterwards, that is, between a DSyntR and the actual sentence, belongs to the realm of syntax, morphology and phonology and is, therefore, irrelevant to our topic.

The Semantic Component

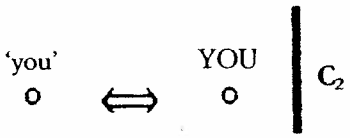
This consists of SEMANTIC RULES belonging to several different types. To be more specific I will consider a few semantic rules necessary for the (6) $\Longleftrightarrow$ (7a) correspondence.

Lexical Sem-Rules

(9)

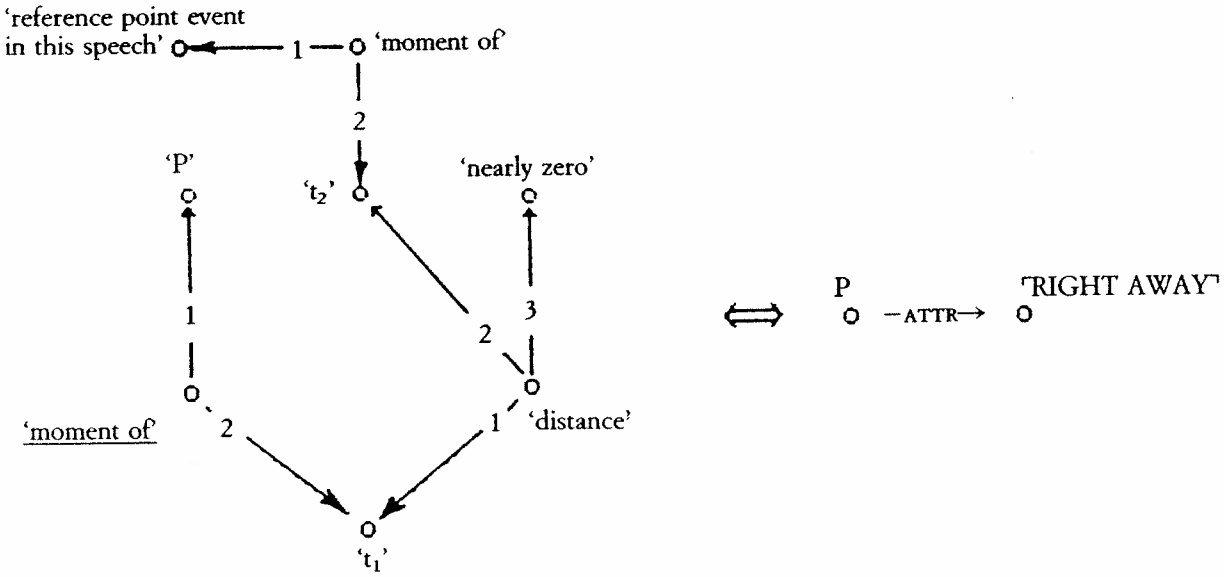


(10)



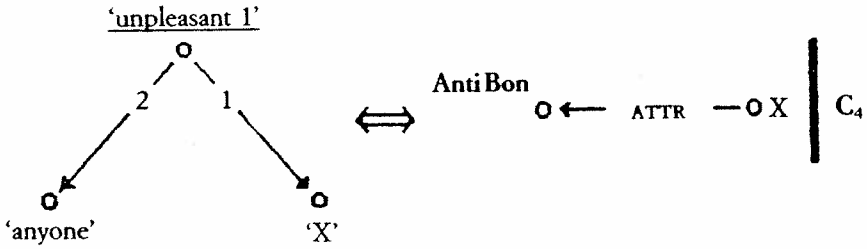
A Phraseological Sem-Rule

(11)

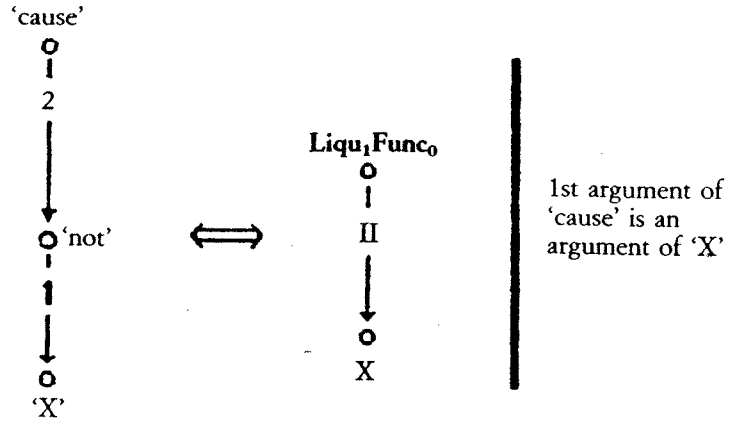


Lexical-Functional Sem-Rules

(12)

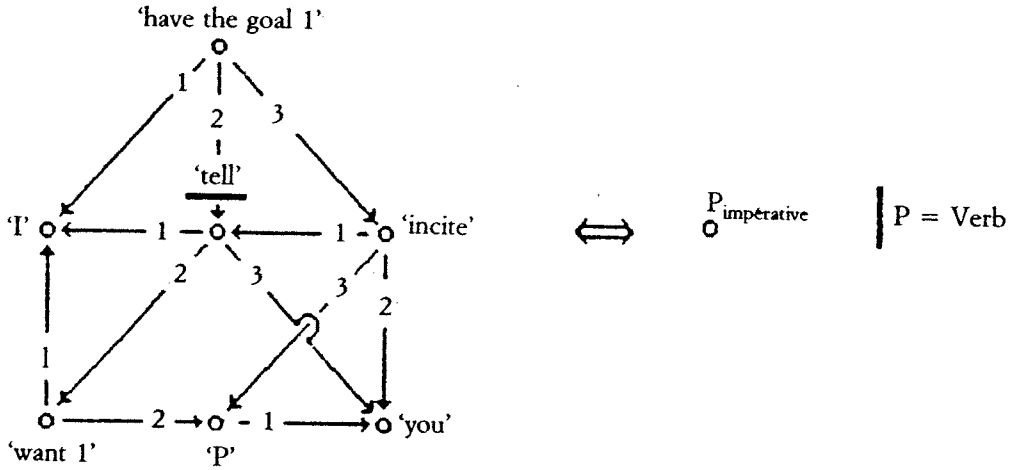


(13)



A Morphological Sem-Rule

(14)



In rules (9)-(14), the expression to the right of a vertical bar (i.e., | C_i) represents cooccurrence conditions that specify the syntactic behavior of the lexical item in question and its combinability with other lexical items.

The above six Sem-rules are not sufficient to specify fully the (6) $\Longleftrightarrow$ (7a) correspondence: we need a number of other rules as well, and first of all – general META-rules that can manipulate the Sem-rules in order to implement the actual construction of a DSyntR for a given SemR or of a SemR for a given DSyntR; etc. Also the rules (9)-(14) are probably underdefined, lacking precision in terms of semantic characterization and/or important cooccurrence conditions. Nevertheless, I think they are sufficient for illustrative purposes: now the reader can clearly see for himself what the Sem-rules of the MTM are like.

Let me emphasize specifically the following important fact: Sem-rules (9)-(11) are nothing else but LEXICAL or DICTIONARY, ENTRIES of a special dictionary, designed as an inalienable part of the Meaning-Text Model and within the framework of the Meaning-Text Theory: the Explanatory Combinatorial Dic-

tionary (ECD), which will be discussed later, in Section IV. For the time being, suffice it to say that lexical and phraseological Sem-rules as they are shown here are the same as entries of this dictionary (allowing for obvious differences in presentational layout and incompleteness of the Sem-rules in question). I will return to this problem when quoting six sample lexical entries of an ECD for English.

As for the other, i.e. closer-to-the-surface, components of the MTM, they are (as already mentioned) immaterial for this paper, and they will not be touched upon. Instead, I will say a few words about the Meaning-Text Model as such.

A General Characterization of the MTM

There are three important constraints on MTMs of the type discussed in this paper.

First, the MTM is designed to carry out synonymous and quasisynonymous PARAPHRASING. In other words, it is supposed to be able to produce, for any given sentence S of **L**, (ideally) all sentences S_i, that are judged (by the speakers of **L**) to be semantically identical or near-identical to S. (Thus the MTM is allowed to alter the meaning of the starting sentence – but it must be fully aware of this alteration and have necessary means to specify it.)

Second, the MTM must implement the paraphrasing in question VIA A SEMANTIC REPRESENTATION, i.e., by reducing sentence S to SemR(S) and then producing all possible sentences S_i synonymous to S on the basis of this SemR: S_i(SemR(S)). The SemR must be of the kind specified above: a semantic network whose syntax is that of predicate calculus but whose vocabulary is roughly that of **L**.

Third, the MTM must use, between a SemR and the corresponding sentences, their DEEP-SYNTACTIC REPRESENTATIONS, also of the kind specified above. In this manner, all linguistic phenomena which are not directly related to meaning and therefore only tangentially relevant to our topic (like government, word order, morphology, etc.) can safely be relegated to closer-to-the-surface components of the model and consequently excluded from our consideration, so that we can concentrate on the SEMANTIC description of linguistic items. (In Sem-rules (9)-(10) a semantic description of two English lexemes is proposed, in (11) – a semantic description of an English phraseme in (12)-(13), a semantic description of two lexical functions, and in (14), of the English imperative grammeme) [7].

This, then, will be our frame of reference: the Meaning-Text Model, as specified by the above three constraints. We will be looking for and at semantic primitives exclusively from the viewpoint of this MTM; we will evaluate them strictly within the limits of the latter.

[7] The proposed semantic description of the imperative follows that of Wierzbicka [1980: 342].

From what has been said so far it follows that a Meaning-Text model must indicate RELATIONS between lexical meanings of **L** (equality, inclusion, intersection). To be more specific, we want our MTM to paraphrase text – using, for instance, equations and entailments illustrated in (15):

(15) a. *John got Mary pregnant.* ⟨ = *John knocked Mary up* ⟩ ⇒
Mary is now with child. ⟨ = *Mary is in the family way.*⟩
b. *Mary is an expectant mother.* ⟨ = *Mary is a mother-to-be*⟩
⟺ *Mary will have a baby* ⟺ *Mary is expecting.*
c. *Mary is giving birth.* ⇐ *Dr. Blostein is delivering Mary's child.*

One natural way to enable the model to do so is by stating the following equalities:

(16) a. *X is pregnant with Y (by Z)* = 'As a result of the functioning of X's reproductive system, X has inside X's body a Y ⟨Ys⟩ which is developing until Y is able to live outside X's body and then will come out of X's body, (Y being an offspring of Z)';
b. *Y gets X pregnant* ⟨*Y knocks up X*⟩ = 'As a result of sexual intercourse with X, male Y causes that female X becomes pregnant';
c. *X is having a baby* = 'woman X pregnant with Y is in a state such that Y, which has developed to the point that Y is able to live outside X's body, comes out of X's body';
d. *expectant mother* = 'pregnant woman'; etc.

These statements are SEMANTIC DECOMPOSITIONS of the corresponding lexical expressions presented in the form of lexicographic definitions; they are logically equivalent to lexical Sem-rules, mentioned in (9) and (10), and they must be stocked in, and available for, any formal system having the same goals as the MTM, i.e., carrying out synonymic paraphrasing.

We require from such decomposition ABSOLUTE MUTUAL SUBSTITUTABILITY with the lexical units decomposed: a lexical unit must be replaceable by its decomposition and a multilexical expression which is the decomposition of a lexical unit must be replaceable by this unit in any imaginable context – *salva significatione* (i.e., the stylistic elegance or even lexical cooccurrence can be violated) [8]. This includes, of course, substitutability within the definitions. Thus, if

(17) a. *X reveals Y to Z* = 'X immediately causes that Y becomes known to Z',

[8] Mutual substitutability of the definition and the item defined, as well as the greater semantic simplicity of the defining elements with respect to the item defined (mentioned below), are major requirements in Wierzbicka's semantic approach, repeatedly stated and defended in her publications from the late sixties.

b. *Y becomes W* = 'Y begins to be^{2}3 W',
and
c. *Y is* [= BE^{2}3] *known to Z* = 'information "Y" is^{2}5 in Z's psyche',

then obviously, by substituting for 'become' and 'known' in (17a) their decompositions, we obtain the following:

d. *X reveals Y to Z* = 'X immediately causes that the information "Y" begins to be^{2}5 in Z's psyche.

Indeed, *John revealed to all his colleagues that he had knocked Mary up* DOES mean 'John immediately caused [by saying or writing] that the information "John, as a result of sexual intercourse with Mary, had caused that Mary became pregnant" began to be^{2}5 in the psyche of all his colleagues'. [9]

To put it differently: Given the above task of automatic paraphrasing, we require that a lexical unit L of **L** be semantically described (= defined) in terms of two or more lexical units L_1, L_2, . . ., L_n of **L** such that 'L' = 'L_1' ⊕ 'L_2' ⊕ . . . . ⊕ 'L_n'. (The symbol ⊕ stands for a highly specific operation of «semantic addition», or «amalgamation», see footnote 9). Thus, generally speaking, a lexical unit L is always defined in terms of lexical units L_i that are SEMANTICALLY SIMPLER than L. [10] As a consequence, vicious circles, the current plague of existing dictionaries, are banned from our system of lexical definitions. [11]

[9] *Syntactico-Semantic Accommodations in Definitorial Substitutions.* It should be constantly borne in mind that substituting a defining expression for the defined, be it in a text or in another definition, is by no means a simple mechanical concatenation of word strings. Such a substitution is carried out in semantic networks, not in actual sentences, and presupposes special syntactic and semantic operations, involving restructuring (deleting repeated or contradictory pieces of meaning, instantiating variables, removing or adding semantic constraints on variables, and so forth). This should be done by meta-rules of the type mentioned above (p. 79). One can see the necessity for such adjustments even in example (17). The complex operation of «putting together» the meanings of lexical units which form a sentence (or a phrase) can be notated ⊕: this symbol implies summation, but of a particular type; we mean taking into account all syntactic and semantic rules of **L**. – On the interaction of lexical meanings in the process of their summation, see Apresjan [1977] (first published in Russian in 1972) [1974: 79-94], and [1980: 73-94].

[10] Let me formulate, at this point, the following two important provisos.

First, not all lexical units of **L** can be semantically decomposed: are excepted those whose meanings are semantic primitives or quasi-primitives (see below, p. 98-99); some of the latter, however, can be decomposed partially.

Second, «*A* is semantically simpler than *B*» means that 'B' can be defined (= decomposed) in terms of 'A' but not vice versa:

'B' = 'A' ⊕ ... ⊕ 'x_n', while 'A' ≠ 'B' ⊕ ... ⊕ 'y_m'.

The possibility of defining 'B' in terms of 'A' but not inversely can be determined in the process of sunsequent decompositions and substitutions.

[11] *Circularity in Lexicographic Definitions.* There is no need to flog this dead horse. In principle, I am as against circularity as anyone, and, together with Apresjan, Wierzbicka and many others, I try to eliminate vicious circles in lexicographic definitions. (Here I need not advance explanations of why circularity is so vicious.) Yet I think, and here I must disagree with Wierzbicka, that in certain cases we need not avoid circularity. Take, for instance, the case of EYES, SEE and LIGHT. Wierzbicka proposes to define EYES solely by describing in detail the physical appearance of human eyes (*eyes* = 'parts of the body in the upper part of the face which can open and close', Wierzbicka [1980: 80]) and then to reduce SEE to EYES (*see* = 'perceive by eyes'); then LIGHT can be reduced to SEE (*light* = 'what can be only seen'). I, however, think that we should not omit

This, in turn, entails that, because the number of lexical units of any **L** is finite, we will inevitably arrive at such lexical units $\{L_i'\}$ that cannot be decomposed any further in terms of other lexical units of **L**. THE LEXICAL UNITS $\{L_i'\}$ ARE THE SEMANTIC PRIMITIVES OF **L**. (Note that I am not saying that semantic primitives are not definable: they are not definable only in terms of other lexical meanings of **L**; however, they are definable in terms of extralinguistic – logical, psychological, mathematical, physical, – notions.)

Let us stop for a second to retrace our steps:

(i) We require that the problem of semantic primitives be considered within the framework of the problem of automatic paraphrasing.

(ii) Automatic paraphrasing requires that all possible semantic links among lexical units of **L** be explicitly indicated; that is, that the meaning of a lexical unit be described in terms of other simpler lexical units. This entails semantic decomposition.

(iii) We require that semantic decomposition be carried out with utmost rigor; that is, the expression.

$$\text{'A'} = \text{'B'} \oplus \text{'C'}$$

means, first, that 'B' and 'C' are semantically simpler that 'A' and, second, that 'A' and 'B' $\oplus$ 'C' are mutually substitutable in all context *salva significatione*.

From this, it automatically follows that we will be led to a set of undecomposable lexical meanings. Thus we are now in a position to answer at least three of the questions about semantic primitives asked at the beginning of the paper:

1) Within the framework of a MTM, we can be sure that semantic primitives exist.

2) They are the simplest lexical meanings of **L** – lexical meanings that cannot be represented in terms of other lexical meanings of **L**.

3) And they must be found in the process of semantic decomposition to which the entire lexical stock of **L** is necessarily submitted.

However, for the time being, I am unable to produce even one semantic primitive for any language, let alone a complete inventory of semantic primitives. Of course, I have in mind a few possible candidates: 'exist', 'time', 'space', 'information', 'contact', 'identical', 'more [than]', 'and', 'or', 'not', 'if',

'seeing' from the definition of EYES. For instance, would we call *eyes* things on the face of an inhabitant of a fantastic planet that look very much like human eyes ('open and close') but serve not for perception but only for sending out signals that are perceived by the skin? At the same time, 'eyes' should not be omitted from the definition of SEE. The same seems to be true concerning EARS/HEAR and SOUND; HEAT/LIGHT and SUN (or FIRE); SEXUAL URGE and SEXUAL ORGANS; BLOOD and RED; GRASS/LEAVES and GREEN; SUGAR and SWEET; and maybe other lexemic pairs. I do not see what harm can be done if circularity is admitted to lexicographic definitions in strictly controlled circumstances and under the provision of being always explicitly marked out. In cases like those mentioned above circular definitions reflect semantic reality better than circularity-free, but somewhat artificial definitions. – Note, however, that my claim that in certain cases circularity may be welcome does not interfere with the main line of the Meaning-Text approach. In complete conformity with Wierzbicka and her predecessors, the lexicographic definitions of the ECD do not allow circularity – except (if I am right) for a few specially singled out lexical groups.

. . .; but I would not even try to defend them against an attack or to justify my choice. (Note that except for 'information' and 'contact', these candidates are a subset of Zholkovsky's [1964c] primitives.) In the same vein, I am not ready to accept wholeheartedly, as the ultimate truth, Wierzbicka's 15 primitives, although she has tested them in insightful semantic analyses of hundreds of lexemes, grammemes and syntactic constructions of several languages. I prefer a more inductive, experimental approach. First, the whole of the lexical stock, or at least a considerable fragment of the lexical stock, of **L** must be fully described, which means, to me, that it will be semantically decomposed such that the accuracy of decompositions is checked within a paraphrasing system. Only then it will be possible, so I think, to proceed to a fruitful discussion of semantic primitives. Now I can only say that semantic decompositions as such (which will eventually lead to semantic primitives) seem to me more important than semantic primitives themselves. The trip feels more rewarding than the destination, so to speak.

For the purposes of a Meaning-Text model, i.e. for synonymic paraphrasing, semantic decomposition of lexical units need not be pushed too far, i.e. necessarily all the way to semantic primitives. There are two reasons why it is better not to decompose lexical meanings too deeply.

First, a decomposition into ultimate elements, i.e. into primitives, makes the semantic description of a semantically rather complex lexical unit too unwieldy and cumbersome, that is, in practice unmanageable. It is for this reason that Zholkovsky advocated, in his very first semantic papers, the use of INTERMEDIATE semantic units, each of which, in its turn, can be represented in terms of other intermediate units and/or semantic primitives. Theoretically, an intermediate semantic unit is simply an abbreviation, standing for a complex configuration of simpler semantic units; but in practical terms, the use of intermediate semantic units means the difference between the manageable and the unmanageable.[12]

Second, gradual decomposition allows us to indicate explicitly the DIRECT semantic links between lexical units; having recourse exclusively to semantic primitives, we would lose this possibility. Thus, if 'A' = 'x ⊕ y ⊕ z ⊕ w' and 'B' = 'y ⊕ z', where 'x', 'y', 'z' and 'w' are semantic primitives, we do not immediately see the relation between 'A' and 'B'; therefore, the decomposition 'A' = 'x ⊕ B ⊕ w' seems much preferable to me (cf. below, the Principle of Maximal Block).

In this way, we can successfully work on a MTM and pursue the goal of semantic decomposition for a vast stock of lexical units WITHOUT HAVING

[12] The technique of using intermediate expressions in order to achieve necessary economy in the complexity of messages in very typical of formal languages. At the same time, basic, or elementary, expressions ensure economy in the lexicon and grammar. An intermediate expressions is always representable in terms of simpler intermediate and/or basic expressions, so that all intermediate expressions can be reduced to basic ones. Equations relating intermediate and basic expressions are in fact TRANSFORMATION RULES of the formal language. Concerning this practice in logic, see Quine [1953: 26-27]; in programming languages, it manifests itself in macro-commands and standard functions (= sub-routines).

SEMANTIC PRIMITIVES AT OUR DISPOSAL. In a sense, the need for them is not even that urgent. This is not to say that semantic primitives would be useless for our endeavor. On the contrary, I am convinced that they would prove extremely useful: they would ensure the commensurability of our decompositions and considerably enhance the precision and rigor of our descriptions. But semantic primitives are not indispensable at the present stage of semantic research and the main thrust now, at least in my view, should rather be at the level of semantic decompositions. (Let me mention, in this connection, Miller and Johnson-Laird [1976] and Dowty [1979], two volumes where the problem of lexemic semantic decomposition is dealt with seriously.)

As far as the Meaning-Text theory is concerned, the slogan of the day should be: «More empirical work in the semantic description of lexical units in different languages» [13].

However, it turns out that a semantic description of a lexical item L is neither complete nor convincing without a full-fledged analysis and description of L's syntactic behavior and lexical cooccurrence. This is very much in keeping with F. de Saussure's famous dictum: «La langue est un système où tout se tient». Thus semantic components in L's decomposition must reflect the behavior of its syntactic actants and determine its semantically motivated lexical collocations. Therefore, we accept the necessity of a multi-faceted overall description of lexical units as entries in a very special dictionary – the Explanatory Combinatorial Dictionary.

4. *The Explanatory Combinatorial Dictionary and Theoretical Semantics*

At first glance, an Explanatory Combinatorial Dictionary, or ECD, might seem only tangentially related to semantic primitives. But to me, this is not so: according to what has been stated above, semantic primitives should be found as a result of consecutive semantic decompositions, and these should be performed within the precise framework of an ECD. Therefore, I will provide a characterization of the ECD, albeit a sketchy one.

In contrast to existing dictionaries, an ECD is theory-oriented: it is being compiled within the Meaning-Text linguistic theory and constitutes an integral part of a MTM's Sem-component. More specifically, an ECD contains lexical and phraseological semantic rules (see p. 78-79), i.e., an ECD entry is a lexical or phraseological Sem-rule. An ECD possesses the following three properties:

- An ECD is consistently oriented to text production (rather than text understanding). It is an ACTIVE, or synthesis-oriented, dictionary, supplying the

[13] One can of course proceed the other way around – Wierzbicka-style, that is, isolating semantic primitives first and then applying them in empirical semantic decompositions («from bottom to top»). I am only saying that within the framework of the Meaning-Text theory, with its ends and means, the way chosen is different: *per aspera* (= *decompositions*) *ad astra* (= *primitiva semantica*)! Let it also be emphasized that this specific choice does not logically follow from the postulates of the MTT: I opt for the «from top to bottom» approach because I find it more practical.

lexical means a language has for expressing a given idea. In other words, an ECD is, essentially, a description of linguistic synonymy.

- An ECD is a highly formalized dictionary. It is a sophisticated LOGICAL DEVICE, featuring entries with an explicit and precise structure.
- An ECD is not subject to any pragmatic, pedagogical or commercial considerations. It is a lexicon with PURELY SCIENTIFIC GOALS, for which we claim conformity to several rigorous criteria.

The unit of lexicographic description in the ECD is a *lexical unit*: a *lexeme* or a *phraseme* (i.e., a word or an idiomatic expression taken in one well-specified sense). Each lexical unit has its own dictionary entry, and each dictionary entry corresponds to one lexical unit. Related lexical units having an identical signifier and sharing non-trivial semantic components in their signifieds are grouped into *vocables* (which reflect polysemy; cf. the remark on p. 100). Thus, the English vocable *improve* includes six lexemes:

improve 1a. 'The value of the quality of X becomes higher'
[*The weather suddenly improved*]

improve 1b. 'X causes that Y improves 1a' [14]
[*A wave of hot air from the Atlantic suddenly improved the weather*]

improve 1c. 'X's having practiced Y causes that X's execution of Y improves 1a'
[*Jim is steadily improving at soccer*]

improve 1d. 'The health of a sick person X improves 1a . . .'
[*Jim is steadily improving*]

improve 2. 'X causes that the market value of a piece of real estate Y becomes higher'
[*Jim improved his house by installing indoor plumbing*]

improve 3. 'X creates a new Y' by improving 1b Y . . .'
[*Jim has drastically improved upon that translation*]

I cannot go into the details of semantic relations among lexical units within a vocable, but it seems obvious that these relations are relevant to semantic decomposition and consequently to semantic primitives.

An ECD entry, i.e. the description of a lexeme, consists of three main zones (divided into sub-zones, which I will ignore here).

The *Semantic Zone* of an entry contains the DEFINITION (of the head lexeme), written in a special semantic metalanguage (according to a few formal principles, see below). To put it differently, the definition of a lexeme represents its semantic decomposition and, therefore, the definitions are of primary interest. An important property of an ECD definition is that it uses, in a very essential way, variables for semantic actants (= for arguments of the functors appearing in it); these variables appear both in the definiendum and the defi-

[14] Note the use of ungrammatical expression *causes that*. In our semantic metalanguage, i.e., in semantic decompositions, we admit a few clumsy and even ugnrammatical expressions, if this is unavoidable for the sake of semantic precision.

niens, the former being what is called a PROPOSITIONAL FORM. For example, the definition of the English verbal lexeme *help* 2 [*Jack helped Mary (to) finish her studies with his generous gifts of money*] would be written as follows:

(18) *X helps Y to do Z with W* = 'Y doing Z, X causes that X's resources W help 1 Y to Z'

The lexeme *help* 1 [*Trade helps the industry to develop*] is defined, in its turn, as in (19):

(19) *W helps Y to do Z* = 'W is a factor contributing to Y'_sZ_{-ing} taking place, Z being desirable to Y or to people in general'

Let it be emphasized that the formulations we see in the right-hand side of definitions (18) and (19) correspond to the semantic networks (= semantic decompositions) appearing in (9)-(11). The only difference is one of presentation: in the ECD definitions we use a less formal way of writing, with an eye to better readability.

I will return to the question of ECD definitions at the end of this section.

The *Syntactic Zone* of an entry contains the GOVERNMENT PATTERN, which specifies, for each semantic actant of the head word (X, Y, Z, . . .), the corresponding deep-syntact actant (I, II, III, . . .) and all surface means for expressing the latter in the text. For example, for *help* 2 we have the following Government Pattern:

Government Pattern

X = I	Y = II	Z = III	W = IV
1. N	1. N	1. V_{inf} 2. *to*V_{inf} 3. Adv_{dir} 4. $Prep_{dir}$ 5. *with*N	1. *with*N 2. *by*V_{ing}

1) *If* Z = 'move or travel in the direction α

2) $C_{3.5}+C_{4.1}$: undesirable *then* (III=L('α') *and* $C_3 = C_{3.3,3.4}$) *is possible Kathleen helped Peter (to) finish his work with her judicious advice. Kathleen helped Peter with his work. Kathleen helped Peter down (out of the room) with a strong kick in the bottom.*

[Adv_{dir} and $Prep_{dir}$ stand for directional adverbs and prepositions, respectively: *up, out, into, across*, . . .; Rule 1 means that, for instance, instead of *help John to climb up the stairs* one can say *help John up the stairs.*]

The *Lexical Cooccurrence Zone* of an entry contains LEXICAL FUNCTIONS, which were introduced above (p. 75). To the example given there, I will add a few more:

Magn (*naked*)	= *stark*	$Oper_1$ (*blow*)	= *deal* [ART~]
Magn (*patience*)	= *infinite*	$Oper_1$ (*support*)	= *lend* [~]

Magn (*thin* [person]) = *as a rake*	$Oper_1$ (*order*)	= *give* [ART ~]
	$Real_2$ (*demands*)	= *meet* [*the* ~]
	$Real_2$ (*exam*)	= *pass* [ART ~]
	$Real_2$ (*hint*)	= *take* [ART ~]

The syntactic and lexical cooccurrence zone of an ECD entry correspond to the cooccurrence conditions shown in Sem-rules (9)-(11). Thus the reader can see now that an ECD entry is but a semantic rule of the MTM, presented – for the sake of convenience – less formally.

Since fragment of a Russian ECD [Mel'čuk and Zholkovsky 1984] and of a French ECD [Mel'čuk *et al.* 1984, 1988] are now available, I will limit myself to the above remarks, plus two additional sets of data: first, a series of lexical entries for English; and second, a brief discussion of lexicographic definition, since it is so essential for the problem of semantic primitives.

Six Lexical Entries for English. I will cite four lexical entries for different senses of the verb ESCAPE and two lexical entries for corresponding nouns (= *nomina actionis*). Notice that these entries do not cover all senses of the verb and of the noun; they were chosen for illustrative purposes only. Notice also that there are many important details that cannot be explained so that I have to rely on examples and the reader's goodwill.

Lexical entries for the verb ESCAPE and the corresponding noun have been written by Ian Mackenzie and are reproduced here with his kind permission.

To clarify the intention behind the sample entries and to justify, if only superficially, our division of the verb ESCAPE into lexicographic senses (= into separate lexemes), let me point out the following: The major borderline – within the limits of the senses considered – is drawn between those senses of ESCAPE that do not presuppose pursuit (E. 1a, 1b, 1c) and the one that does (E.2). In the first group we further distinguish the following three senses: roughly, 'to escape from where one is confined by someone' (= E. 1a), 'to escape from where one is trapped by something' (= E. 1b) and 'to escape from where one is threatened' (= E. 1c). All those senses feature observable differences in their government patterns and lexical functions. Note, for instance, such a subtle distinction as the one between *Pilot McQueen escaped all their missiles by sheer chance – he wasn't even aware that they had been fired at him*, where we have ESCAPE 1c ['the missiles simply missed him or else he moved out of the dangerous zone'], and *A brilliant pilot, McQueen escaped from all their missiles*, with ESCAPE 2, which governs *from* ['he outmaneuvered the pursuing missiles']; cf. also *McQueen escaped their bullets* vs. **McQueen escaped from their bullets* [because bullets cannot pursue you].

ESCAPE, verb

I.1a. *X escapes from Y through Z* = *X*, being kept by Y^1 against X's will in place or state Y^2, such that Y^1s intent is to thwart any attempt by X to move out of Y^2, intentionally moves out of Y^2 via Z, thereby becoming free.
[NB: Y^1 and Y^2 represent what can be called a split variable: Y^1 are jailers,

and Y^2 is jail; you can *escape from your jailers* or *from your jail*, but you cannot express both in the same phrase.]

Government Pattern

X = I	Y = II	Z = III
1. N	1. *from* N 2. N	1. $PREP_{traj}$ N

$PREP_{traj}$ is a preposition which can denote the trajectory of something moving; e.g. *through [the open window], out of [the back door], across [the steppe], over [the mountains], by way of [the river].*

1) $C_{2.2}$: semi-archaic; not felicitous in spoken usage: *?He escaped his sleeping jailer* or *?He escaped Alcatraz*. It is likely to be used in formal written style, and especially in situation where there is overlap with ESCAPE I.4b, i.e., where emphasis lies on X's ending an unwanted or unpleasant experience; e.g., *He vowed to escape the cold and fifthy cell.*

C_1 : *McQueen escaped.*

C_1+C_2 : *McQueen escaped from his guards ⟨from Stalag XIV⟩*

C_1+C_3 : *Joe Bonano escaped through his cell window.*

$C_1+C_2+C_3$: *Joe Bonano escaped from San Quentin through a storm sewer.*

Lexical Functions

$Syn_{\supset}$: flee; obtain X's freedom

$Syn_{\cap}$: break out [of], run away [from]; reach freedom

S_0 : escape I.1a [noun]

S_1Perf : escapee

$S_{1\cap}Perf$: runaway, fugitive

S_1Able_1 : escape artist 2

S_2^{usual} : guard, jailer, turnkey [= Y^1]; place of confinement, jail, prison, prison camp, concentration camp, lockup, dungeon, . . . [= Y^2]

S_3 : escape route[15]

A_1 : runaway, fugitive, fleeing

Bon : daringly

$Qual_1$: kept, imprisoned, guarded, . . .

One may escape from a jail, prison, or prison camp by either *stealthy* or *overt*

[15] The expression *escape route* in the entries for *escape* I.1a, *escape* I.1c and *escape* I.2 should be represented either by three different lexical units (*escape route* 1a, 1b and 2) or as one lexical unit with disjunctive meaning ('S_3(escape I.1a, or I.1c, or I.2)'). The choice must be made according to the naturalness of that or this complete lexicographic description of the expression in question.

means. Stealthy means typically include *filing* or *sawing through prison bars*, *cutting through barbed wire* with wirecutters, *digging tunnels* to the outside, *hiding in a vehicle* leaving the prison, *walking out in disguise* or obtaining *the connivence of the guards*. Overt means could entail the *taking of hostages* or a *violent incursion of armed accomplices*.

Examples

He escaped from custody.. Dreyfus did not escape from Devil's Island; he was finally released as a result of mounting public outrage. A plot by at least six inmates to use a crossbow to kill a tower guard or incinerate the tower and then escape from Trenton State Prison over a homemade bridge has been thwarted, state correction officials said today. He managed to escape from the miner's cabin while the kidnappers were in the kitchen. Three more East Germans have just daringly escaped over the Berlin wall in a homemade balloon. Four o'clock had come and gone with still no sign of little Billy, and Martha's head was awhirl with visions of the lion that had escaped from Riddington Zoo the previous night.

I.1b. *X escapes from Y through Z* = X, being trapped in Y, moves out of Y via Z.

Government Pattern

X = I	Y = II	Z = III
1.N	1. *from* N	1. *through* N 2. *by way of* N 3. *via* N

C_1 : *Houdini escaped*
C_1+C_2 : *Houdini escaped from the padlocked coffin.*
$C_1+C_2+C_3$: *I escaped from the rubble of the collapsed building through a still intact storm sewer.*

Lexical Functions

$Syn_{\cap}$: break out [of]
S_0 : escape I.1b [noun]
professional-
$S_1Able_{\supset}$: escape artist 1
$Qual_1$: trapped, confined, entangled, . . .
F_1=XcannotE.
from Y: //There is no escape I.1b [noun] [for N = X / from N = Y] [*There was no escape for the whale; There was no escape from the Sacramento River*]

Examples

We finally escaped from the fishnet in which our bodies had become entan-

gled. I certainly hope the whale escapes [in regard to a humback whale which swan up the Sacramento River and appeared trapped and unable to get back to the ocean]. Houdini needed only 7 minutes to escape from a coffin bound with three chains, six padlocks, and immersed in ten feet of water. A rocket must be traveling more than 25000 miles an hour before it can escape from the earth's gravity.

I.1c. *X escapes from Y by way of Z to W* = X, being in place Y^1 where Y^2 is, such that (something related to) Y^2 threatens or can threaten X and such that it is possible that X will not be able to move away from Y^1 before the threat by Y^2 to X is realized, succeeds in intentionally moving out from Y^1 via Z to place W, thereby causing that the threat not be realized.

Government Pattern

X = I	Y = II	Z = III	W = IV
1. N	1. *from* N	1. $PREP_{traj}$ N	1. *into* N 2. *to* N 3. *for* N

1) $C_{4.1,2}$: place W is reached
2) $C_{4.3}$: place W is not necessarily reached
3) $C_{2.2}+C_{4.1}$ } : impossible
4) $C_{4.3}$ without $C_{2.2}$ }

C_1 : *Dith Pran escaped.*
$C_1+C_{2.1}$: *Dith Pran escaped from Cambodia ⟨from the Khmer Rouge⟩.*
$C_1+C_{2.2}$: *Dith Pran escaped Cambodia ⟨the Khmer Rouge⟩.*
C_1+C_3 : *Dith Pran escaped across the Mekong.*
C_1+C_4 : *Dith Pran escaped into (to) Thailand.*
$C_1+C_{2.1}+C_{4.1}$: *Dith Pran escaped from Cambodia into Thailand.*
$C_1+C_{2.2}+C_{4.3}$: *Dith Pran escaped Cambodia for Thailand..*
$C_1+C_2+C_3$: *The crew escaped from the sinking ship along the lifeline to the safety of the Coast Guard cutter.*
Impossible: **They escaped Cambodia into Thailand* (3).
**They escaped for Thailand* (4).

Lexical Functions

$Syn_{\supset}$: flee
S_0 : escape I.2 [noun]
$S_1Perf_{\subset}$: refugee
S_3 : escape route
$S_{4\cap}$: refuge; sanctuary
$A_{1\cap}$: fleeing
$Magn_{['moveaway']}$: frantically

Miraculously, a large number of people escaped from the burning Hindenburg to the safety of the nearby tanker. The first party had run into a Soviet ambush; later it was learned that all of the men had escaped but that almost a ton of arms and ammunition had had to be abandoned. Since the Soviet government ultimately controls all employment and education, once a person is on the wrong side of the law, once his name gets on the wrong list, he has nowhere to escape to. It was there on a winter day 44 years ago that 86 Jews gathered to discuss how to escape German-occupied France for neutral Switzerland. The U.S. counterintelligence service, concerned about Barbie's extensive knowledge of American intelligence operations, helped him and his family escape to Bolivia. Two police officers, identified as Isidoro Chantal and Moises Gonzalez, escaped from the shootout [a shootout in which 21 policemen were killed by a drug gang]. The attackers fired at least nine shots, and then escaped from the scene in a green Peugeot.

I.2. *X escapes from Y by way of Z to W* = X, being pursued by Y^1, which is in place Y^2, succeeds in intentionally moving out from Y via Z to place W, thereby causing that Y^1's pursuit fail.

Government Pattern

X = I	Y = II	Z = III	W = IV
1. N	1. *from* N	1. $PREP_{traj}$ N	1. *into* N 2. *to* N

C_1 : *The antelope escaped.*
C_1+C_2 : *The antelope escaped from the lion.*
C_1+C_3 : *Dith Pran escaped across the Mekong.*
C_1+C_4 : *Dith Pran escaped (in)to Thailand.*
$C_1+C_2+C_4$: *Dith Pran escaped from Cambodia into Thailand; Manuel escaped from Nogales, Texas, to the city of the same name across the Mexican border.*
$C_1+C_2+C_3+C_4$: *McQuenn escaped from Germany across the Alps into Switzerland.*

Lexical Functions

$Syn_{\supset}$: flee
S_0 : escape I.2 [noun]
S_1 : fugitive
$S_{1\subset}$: refugee
$S_{2\cap}$: oppressor; persecutor
S_2 : pursuer
S_3 : escape route
$S_{4\cap}$: refuge, sanctuary
A_1 : fugitive, fleeing
$Magn_{['moveaway']}$: frantically

Examples

When Jose saw the Immigration officers enter the restaurant, he quickly escaped through the kitchen. The grizzly sow pulled Hosick down from the tree, puncturing his boot with one fang; Fisher escaped to a treetop, where horsepackers found him after meeting Hosick on the trail. Then, he said, the tanks moved in: his men held them off until midnight, when he and his soldiers escaped through irrigation canals to the hills. He headed for Tijuana in order to escape from the FBI who were hot on his tail.

ESCAPE, noun.

I.1a. *X's escape from Y through Z* = S_0 (*X escapes I.1a from Y through Z*).

Government Pattern

X = I	Y = II	Z = III
1. *of* N 2. *-'s* N 3. A_{poss}	1. *from* N	1. $PREP_{traj}$ N

C_1 : *the escape of Casanova* ⟨*Casanova's/his escape*⟩
C_2 : *the escape from Alcatraz*
C_3 : *the daring escape across the steppe*
C_1+C_2 : *Casanova's escape from his cell* ⟨*from his guards*⟩
C_1+C_3 : *his escape through the cell window* ⟨*by way of the prison roof*⟩
C_2+C_3 : *the daring escape from Stalag XIV by way of a tunnel dug under the latrine*
$C_1+C_2+C_3$: *Casanova's escape from his cell out of a hole he had made in the wall*

Lexical Functions

LFs S_1 Perf, $S_{1\cap}$ Perf, S_1 $Able_1$, S_2^{usual}, S_3, A_1: ↑ ESCAPE I.1a [verb]

$Syn_{\supset}$: flight
$Syn_{\cap}$: jailbreak
V_0 : escape I.1a [verb]
$Magn^{quant}$: mass
Bon : daring
$Magn_{1quant}$+Bon : great
$Oper_1$: make [ART ~] [*He made his escape from Stalag XIV; He made a daring escape from Stalag XIV*]
$LiquFunc_0$: foil [ART ~] [*They foiled his escape*]

Examples

The spectacular escape from the dungeons of Venice was described by Casa-

nova himself in a book published in 1797. The author greatly exaggerates the disrupting effect on the Nazi war effort caused by the «great escape» of several score Allied POW's. The escape of Asian animals from hobby ranches is the latest twist in the drastic evolution of Texas rangeland in 125 years of European settlement. The six inmates have been charged with planning an escape and face disciplinary action. In addition, information about the escape plot has been turned over to prosecutors for possible criminal charges.

I.1b. *X's escape from Y through Z* = S_0 (*X escapes I.1b from Y through Z*)

Government Pattern

X = I	Y = II	Z = III
1. *of* + N 2. -'*s* N 3. A_{poss}	1. *from* N	1. *through* N 2. *by way of* N 3. *via* N

C_1 : *the escape of Houdini (Houdini's/his escape)*
C_1+C_2 : *Houdini's escape from a padlocked coffin*
C_1+C_3 : *Houdini's escape via a secret trap door in the side of the padlocked coffin*
$C_1+C_2+C_3$: *his escape from a padlocked coffin via a secret trap door in its side*

Lexical Functions

V_0 : escape I.1b [verb]
S_1^{usual} $Able_{1C}$: escape artist 1
X cannot $Oper_1$
E. from Y : There is no ~ [for N = X from N = Y] [*There was no escape from the Sacramento River; There was no escape for the whale*]

Examples

To this day, the secrets of Houdini's greatest escapes remain undisclosed. The whale's escape from the Sacramento River was accompanied by joyful cheers from the shore. There is as yet no escape from the political quagmire in South Africa. The escape of a rocket from the earth's gravitational field requires an enormous amount of fuel.

These illustrations make it possible to get a fairly good idea of the nature of the lexical entries of an ECD, which constitute the framework for semantic decompositions, which, in their turn, constitute the framework for our search for semantic primitives. This section ends with a few remarks on the lexicographic definition in the ECD.

The following presentation does not claim originality: the ideas it contains have been advanced more than once by different people, some of them already by Zholkovsky in the early sixties, some by Apresjan and Wierzbicka, as well as by other researchers. I will simply expound them in a systematic manner.

What do we want from a definition in an ECD? Clearly, it should reflect the linguistic intuition of native speakers; but this is a very informal requirement that is hard to check. More formally, the definition of a lexical unit L should satisfy the following three conditions:

- it should ensure an accurate presentation of the PARAPHRASTIC POTENTIAL OF L;
- it should ensure an accurate presentation of its DENOTATIONAL POTENTIAL;
- it should ensure an accurate presentation and explicit demonstration of the SEMANTIC LINKS of L with the the related lexical units of the language.

(Whenever we speak about «accurate presentation», we mean 'accurate from the viewpoint of native speakers'.)

I will take these three conditions in turn.

1. The *paraphrastic potential* of the given lexical unit L is the whole set of paraphrases which can replace L in a text (or within a lexicographic definition). In this connection, the «vertical» (= paradigmatic) and the «horizontal» (= syntagmatic) aspects of paraphrasing should be distinguished.

From the paradigmatic viewpoint, the definition of L – together, of course, with the rest of L's dictionary entry – must supply all the lexical means necessary to replace L in all possible paraphrases, clearly stating the semantic differences between L and its eventual substitutes. In other words, the definition must guarantee the correct and exhaustive SELECTION of lexical material related to L. Thus the definition of *escape* I.1c must allow one to replace *escape* I.1c with *flee* in some context (*Dith Pran escaped Cambodia* ≅ *Dith Pran fled Cambodia*), but not in others (*Dith Pran barely escaped* ⟨**fled*⟩ *Cambodia*) [16].

From the syntagmatic viewpoint, the definition of L must specificy, first of all, its non-restricted lexical cooccurrence, i.e., that lexical cooccurrence which is determined semantically (restricted lexical cooccurrence being covered by lexical functions). Thus *graft* [corruption] should not be defined as 'the practice of obtaining money unlawfully of unfairly . . .' (*Longman Dictionary of Contemporary English*), since *practice* and *graft* are different in their free lexical cooccurrence: *these various practices* vs. **these various grafts, *three counts of the practice of obtaining money* vs. *there counts of graft* etc. A better definition, in this respect, would be 'obtaining money unlawfully or unfairly . . .', the free cooccurrence of gerunds being much closer to that of *graft* than that of *practice*. To sum up, the definition of L must account for L's FREE (i.e. semantically motivated) COMBINATION with other lexical unit in the text.

[16] The combinability of *escape* I.1c with *barely* is guaranteed by the central component 'succeed' in the definition of the former ('barely succeeded' is correct), while the central component in the definition of *flee* is 'move out', which does not combine with 'barely' (*barely moved out').

2. The *denotational potential* of L is the set of objects or situations to which L can be applied. The definition of L must include all the components necessary to allow the use of L when appropriate and to disallow it when inappropriate. It should not, however, embody the entire amount of knowledge a speaker possesses in connection with L: much of this pertains to the THING denoted by L, and not to L itself. The lexicographic definition must most of all avoid including encyclopedic information about the real world; the aim of semantic decomposition is restricted, in our view, to stating relations among words and to applying words to things. Therefore, definitions in the ECD do not try to reflect naive concepts the speakers have of objects and events; these definitions aim at naive concepts the speakers have of words and their usages. For instance, an English speaker knows that cups are normally used with saucers, but this fact is irrelevant for the lexicographic definition of *cup*: a cup will be still called a cup by an English speaker, even if (say, in an exotic culture) it is not meant to be put on a saucer. (For a thorough discussion of all problems related to the lexicographic definitions of such English lexemes as *cup, mug, saucer* and the like, see Wierzbicka [1985: 100-103], which presents such a wealth of ideas and data that it is simply impossible to summarize it here.)

Note, however, that in many cases the definition of L must contain a non-distinctive semantic component – if it is justified by the existence (in **L**) of lexical units that formally include L and unquestionably have this semantic component. Thus, the definition of *snow* should include the component 'white' – not because snow is white (since there is no substance of another color which could be called *snow*, 'white' is non-distinctive in the definition), but because there are English expression like *snowy* 'pure white', *white as snow, snow-[white]*, and *snow* 3 'cocaine powder'. At the same time, the definition of *rice*, rice being also white, should not include the component 'white' since there are no supporting English expressions (no **ricey white, *rice-white* or something similar). This criterion for the inclusion of non-distinctive semantic components in the definitions can be called the CRITERION OF LINGUISTIC RELEVANCE. It was advanced in Apresjan [1969] and has been used since then in the lexicographic research within the framework of the MTM. (This is related to the requirement concerning explicit indication of semantic links between lexical items, see below, p. 100, item 5.)

At this juncture, an interesting question might be asked: Can we, without delving into encyclopedic information, FULLY define EVERY word in a language? Might it not be that, if sticking to the criterion of linguistic relevance, the lexicographer will be obliged to leave certain lexemes under- (or even un-) defined? Let me attempt an answer.

Not all lexical meanings lend themselves equally well to semantic decomposition. Already 25 years ago, Zholkovsky insisted on the essential distinction between *abstract* and *concrete* vocabulary. Abstract lexemes can be described only in terms of other lexemes; there is no way of representing non-verbally, e.g., by a drawing such meanings as 'support', 'manage', 'escape', or 'procrastinate'. On the other hand, concrete lexemes, i.e. names of specific objects, sub-

stances, actions, states etc., such as *horse, gun, sand, jump, write, sleep,* can be (although in many cases not fully) represented by drawings or specimens. Abstract meanings are essentially verbal; concrete meanings are essentially directly associated with complexes of visual, acoustic, gustatory etc. perceptions. The technique of semantic decomposition is especially powerful and successful if applied to abstract lexical meanings. It is, however, not that efficient with respect to concrete lexical meanings. [17]

This is not to say that concrete vocabulary defies semantic decomposition. Most names of artifacts (= things created by people for a specific use), of prominent natural phenomena, of plants, animals and substances important in human life, etc. can be, and are, quite successfully described via semantic decompositions (Wierzbicka [1985] provides ample evidence of this fact). I believe, however, that many names of natural species, of (even culturally important) substances, of colors, of minerals and the like cannot be fully described SEMANTICALLY, i.e. exclusively in terms of simpler lexical meanings. I mean here words having a clearly terminological character (although they are part of common language). I do not see how we can lexicographically define *camembert* as opposed to *brie* (and all other cheeses); how to decompose 'chocolate'; how *malachite* could be fully described in an English (and not geological) dictionary; and what is – in semantic terms! – the difference between 'whiskey' and 'cognac'.

True, all objective properties of things that are perceived by humans and used for distinguishing these things can be described verbally. It is possible to write a text which will be detailed enough to identify *camembert* and to distinguish it from *brie*. But this text would be a description of this particular type of cheese, that is, of the cheese itself, not of the lexical meaning 'camembert'. In other words, if we care about the distinction between SEMANTIC descriptions of lexical meanings and ENCYCLOPEDIC descriptions of things denoted by the corresponding lexemes, then I think we have to admit that numerous concrete lexemes cannot be fully described semantically, i.e., as to their meanings. From this it follows that these meanings cannot be reduced to semantic primitives, whatever these latter be. (Saying this, I follow Apresjan [1974: 71, footnote 5], and *passim*. Cf. also remarks concerning the lexicographic description of the Russian noun *zolotnik* 'slide valve [in a steam engine]' in Ščerba [1940: 68].)

At the same time, «stubborn» concrete meanings should not be left completely undecomposed. Thus, we should state that 'chocolate' is a solid sweet substance that is dark brown (and has a slightly bitter taste), while 'malachite' is a stone that is intensively green and has designs on its surface if polished (although these statements are by no means exhaustive and distinctive defini-

[17] Concrete lexical meanings are very much like proper names, which obviously cannot be represented by semantic decompositions; a proper name has a REFERENT (= thing it refers to) but does not have a MEANING (= cannot have a definition). Along the same lines, concrete lexical meanings can also be compared to elements of a nomenclature (which can be scientifically defined within the corresponding system, but can hardly have purely linguistic definitions).

tions). In this manner, many concrete lexical meanings of «terminological» nature must be represented by PARTIAL decompositions: their *genus* is stated, as well as some *differentiae specificae*; but because of an irreducible residual, they have to be considered as wholes.- Such meanings may be called *quasi-primitives*, and there are hundreds or rather thousands of them! [18].

3. The *semantic links* of L are the relations between the meaning of L and those of other lexical units L'_1 which are perceived by native speakers. If L and certain L'_i are felt as being semantically related, this fact should be explicitly reflected in their definitions. Thus we cannot accept for ESCAPE I.1a and *escape* I.1c the definitions proposed in the *Heritage Illustrated Dictionary of English Language*: 'to break loose from confinement' and 'to succeed in avoiding (capture, danger, of harm)' – since they do not explicitly show the semantic relatedness of these two senses of ESCAPE. In an ECD, the definitions of ESCAPE I.1a, I.1b, I.1c and I.2 share the central component '(intentionally) move out', see pp. 89-92.

So that is what we want from a lexicographic definition: it must be good for paraphrasing (giving all paraphrastic relatives of L and predicting L's lexical combinability), for denotation (that is, for applying L to entities of the extralinguistic world) and for explaining semantic links within the lexical system.

The above three requirements state WHAT a definition should or should not contain; the following six principles state HOW a definition should be written. Let me present them in four groups:

A. The defining language.
B. A particular definition.
C. The system of definition within a vocable.
D. The system of vocables within a lexical field.

(For a detailed presentation see Mel'čuk [1988b].)

A. The Defining Language

1. Univocity Principle.

The defining language must not contain any ambiguous or synonymous terms.

According to this principle, the lexical units within a definition must be accompanied by distinctive numbers that identify the intended senses, so that every expression in a definition has only one meaning. Conversely, every meaning must be expressed in a definition by the same expression.

B. A Particular Definition

2. Adequacy Principle.

[18] Quasi-primitives should also include a number of lexical meanings that are only half-known to the average speaker: special parts of mechanisms (*crank*), exotic animals and plants (*armadillo*), historic terms (*abbot*), scientific concepts (*spin*), etc. Half a century ago L. V. Ščerba insisted that such words should not be fully defined in a language dictionary – see the remark on p. 98. Cf. an interesting discussion of cases of incomplete lexical knowledge in Wierzbicka [1985: 217 ff.].

In the definition of a lexical unit L, each component must be necessary, and the set of all components must be sufficient for the definition to identify L uniquely in all contexts.

This principle is related, in an obvious way, to the requirement of mutual substitutability of the defined and the defining expression in all possible contexts.

3. Decomposition Principle

The definition of L must contain only terms that are semantically simpler than L.

Semantic decompositions have been discussed in detail, and there is no need to return to this problem here.

4. Maximum Block Principle

If a definition contains a free phrase which is composed of lexical units $L_1 + L_2 + \ldots + L_n$ and is semantically equivalent to a lexical unit L (so that 'L_1' $\oplus$ 'L_2' $\oplus \ldots \oplus$ 'L_n' = 'L'), then this phrase must be replaced by L.

This principle ensures GRADUAL decomposition (into «immediate semantic constituents») and thus makes definitions which would otherwise be too long and cumbersome more manageable.

C. The System of Definitions within a Vocable

5. Semantic Bridge Principle

The definitions of any two lexical units of the same vocable must be explicitly linked: whether by a semantic bridge (i.e., they share a sufficiently important semantic component occupying the same structural position in both definitions) of by a sequence of semantic bridges.

Actually, this can be considered as a definition of the concept «vocable»: a vocable is the set of lexical units having the same signifier and linked – directly or indirectly – by semantic bridges. A vocable in an ECD corresponds to what is a polysemous entry in traditional dictionaries.

D. The System of Vocables within a Lexical Field [19]

6. Uniformity Principle

Two vocables belonging to the same lexical field must be presented, everything else being equal, according to the same schema; or more precisely, the related lexical units of either vocable should be described in a parallel fashion:
(i) their definitions must be formulated as similarly as possible;
(ii) they must appear in the same order within each vocable;

[19] A lexical field is the set of vocables whose main lexical units are linked by a semantic bridge.

(iii) the semantic distances between them must be represented identically.

I am convinced that semantic research must be based, in the most essential way, on a dictionary of the ECD type, with its definitions written according to the proposals that have been formulated in this section. Not that semantics can be fully reduced to the lexicon: there is also the semantics of grammatical values, of syntactic constructions, of prosodies, as well as referential semantics. But the lion's share of semantic problems, including the most challenging and most promising of these, in the first place – these of lexical semantics, lies within the realm of lexicography. This is especially true concerning semantic primitives.

5. *Semantic Primitives: Conclusions*

Having completed the foregoing lengthy presentation, I am now in a position to formulate a few short conclusions (of a rather speculative character).

1. The existence and nature of semantic primitives: Within the Meaning-Text framework, semantic primitives of language **L** are ELEMENTARY LEXICAL MEANINGS (= semantemes) of **L** [20]. They will be arrived at by way of systematic semantic decompositions carried out on the lexical stock of **L**, in accordance with the MTM and, more specifically, the ECD. Thus they are thought of as the final result of a purely empirical research.

2. The role of semantic primitives in actual semantic research: Given their nature, I believe that successful lexicographic descriptions can be carried out in the absence of a previously defined set of semantic primitives. Nevertheless, semantic primitives are very important for the final standardization and systematization of these descriptions. To put it in a nutshell, I believe that work on a dictionary of the ECD type is the task of the day; semantic primitives will be an outcome of this work.

3. The number of semantic primitives: For the time being, I am unable to say anything conclusive about it. I am inclined to agree with Wierzbicka that semantic primitives are few in number (15, for Wierzbicka) – but, because of the above-mentioned quasi-primitives I am not sure what their actual number might be. I think that the total number of primitives and quasi-primitives for a complete lexicon of a natural language could be around a few hundred; «pure» primitives may number a few dozen – or less.

4. The universality of semantic primitives. As I have already stated, I do not think that they are ultimate units of human thought (but even if they were, I

[20] Thus, in complete agreement with Wierzbicka, I require that semantic primitives be natural semantemes, and not artificially constructed concepts or something similar. I think, however, that among intermediate expressions used in the lexicographic definitions (= semantic decompositions) artificial expressions are unavoidable – in case we try to observe all the principles for definitions and to obtain elegant, compact and easily graspable descriptions. For instance, Russian does not have a word for '(to) cause' (all Russian verbs denoting causation denote specific kinds of causation); therefore it turned out necessary to introduce the fictitious verb *kauzirovat'* 'cause', which appears in thousands of definitions. See Apresjan [1974: 74] concerning the necessity of fictitious intermediate expressions in lexicographic definitions (he analyzes the fictitious lexeme *norm* = 'state of affairs perceived by the majority of speakers as most likely in a given situation').

would have doubts about their universality). I conceive of semantic primitives simply as ELEMENTARY LEXICAL MEANINGS OF A PARTICULAR LANGUAGE and I do not know whether they will be the same for all languages. In fact, this is another empirical question that should be answered by future investigations.

Acknowledgements

The ideas presented in this paper have been discussed, on an infinite number of occasions, with A. Zholkovsky, Ju. Apresjan and A. Wierzbicka; I acknowledge here my general indebtedness to them and express my warmest gratitude for sharing so generously with me their intellectual endeavor. The actual text has been read and commented upon by Ju. Apresjan, M. Dillinger, L. Iordanskaja, G. Hirt, I. Mackenzie, N. Pertsov, O. Rambow, and M. Yaeger-Dror, who I am happy to thank for their judicious remarks and suggestions.

Fall 1986-Fall 1987
Newark, Delaware, USA
Mysore, Karnataka, India
Toronto, Ontario, Canada
Montreal, Quebec, Canada.

REFERENCES

Apresjan, Jurij D. [1969], *O jazyke dlija opisanija značenij slov* [On a Language for Describing Word Meanings], in «Izvestija Akademii Nauk SSSR, Serija lit. i jazyka» 28/5, pp. 415-428.

— [1974], *Leksičeskaja semantika. Sinonimičeskie sredstva jazyka* [Lexical Semantics. Synonymic Means of Natural Language], Moscow, Nauka.

— [1977], *Concerning One Rule in the Amalgamation of Lexical Meanings*, in «Linguistics», 185, pp. 5-25.

— [1980], *Tipy informacii dlja poverxnostno-semantičeskogo komponenta modeli «Smysl ⟷ Tekst»* [Types of Information Needed for Surface-Semantic Component of the Meaning-Text Model], Vienna, Wiener Slawistischer Almanach.

Bolinger, D. [1977], *Meaning and Form*, London and New York, Longman.

Dowty, D. R. [1979], *Word Meaning and Montague Grammar. The Semantics of Verbs and Times in Generative Semantics and in Montague's PLQ*, Dordrecht, D. Reidel.

Fodor, J. D. [1980], *Semantics: Theories of Meaning in Generative Grammar*, Cambridge, MA, Harvard University Press.

Hofmann, Th. [1974], *Semantic Studies IV. In Support of Semantic Atoms*, in «Languages Sciences», n° 32, pp. 12-18.

Kay, C., and M. Samuels [1975], *Componential Analysis in Semantics: Its Validity and Applications*, in «Transactions of the Philological Society», pp. 49-81.

Lyons, J. [1977], *Semantics*, I, Cambridge etc., Cambridge University Press.

Masterman, Margaret [1961, *Semantic Message Detection for Machine Translation, Using an Interlingua*, [Paper 36, presented at the First Intern. Conference on Machine Translation of Languages . . .], Teddington, England, National Physical Laboratory.

Mel'čuk, I. A. [1974], *Opyt teorii lingvističeskix modelej «Smysl ⟺ Tekst»*, [Outline of the Theory of Meaning-Text Linguistic Models], Moscow, Nauka.

— [1981], *Meaning-Text Models: A Recent Trend in Soviet Linguistics*, in «Annual Review of Anthropology», vol. 10, pp. 27-62.

— [1982], *Lexical Functions in Lexicographic Description*, in *Proceedings of the Eighth Annual Meeting of the Berkeley Linguistic Society*, pp. 427-443.
— [1988a], *Dependency Syntax: Theory and Practice*, Albany, N.Y., The State University of New York Press.
— [1988b], *Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria*, in «Intern. Journal of Lexicography», vol. 1, n. 3, pp. 165-188.
Mel'čuk, I. A. (with a lexicographic team), *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques*, Vol. I, 1984; vol. II, 1988, Montréal, Presses de l'Université de Montréal.
Mel'čuk, I. A. and N. V. Pertsov [1987], *Surface Syntax of English: A Formal Model Within the Meaning-Text Framework*, Amsterdam-Philadelphia, Benjamins.
Mel'čuk, I. A. and A. K. Zholkovsky [1984], *Explanatory Combinatorial Dictionary of Modern Russian, Semantico-Syntactic Study of Russian Vocabulary*, Vienna, Wiener Slawistischer Almanach.
Mel'čuk, I. A. and A. K. Žolkovskij [1970], *Toward a Functioning Meaning-Text Model of Language*, in «Linguistics», 57, pp. 10-47.
Miller, G. A. and Ph. M. Johnson-Laird [1976], *Language and Perception*, Cambridge, MA, Belknap Press of Harvard University Press.
Nida, E. A. [1975], *Componential Analysis of Meaning. An Introduction to Semantic Structures*, The Hague-Paris, Mouton.
Quine, W. V. O. [1953], *From a Logical Point of View*, Cambridge, MA, Harvard University Press.
Ščerba, L. V. [1940-1958], *Opyt obščej teorii leksikografii* [Outline of a General Theory of Lexicography], in L. V. Ščerba, *Izbrannye raboty po jazykoznaniju i fonetike*, I, Leningrad, Izdatel'stvo LGU, pp. 54-91.
Wierzbicka, A. [1972], *Semantic Primitives*, Frankfurt am Main, Athenaum.
— [1980], *Lingua Mentalis: The Semantics of Natural Language*, Sydney etc., Academic Press.
— [1985], *Lexicography and Conceptual Analysis*, Ann Arbor, MI, Karoma.
— [1987], *English Speech Act Verbs: A Semantic Dictionary*, Sydney etc., Academic Press.
Žolkovskij, A. K. [1961], *O pravilax semantičeskogo analiza* [Rules for Semantic Analysis]. Paper presented at the Conference on Information Processing, Machine Translation and Automatic Text Reading, Moscow, Institute for Scientific Information.
— [1964a], *Predislovie* [Introduction], in *Mašinnyj perevod i prikladnaja lingvistika*, vol. 8, pp. 3-16.
— [1964b], *O pravilax semantičeskogo analiza* [Rules for Semantic Analysis], *Idem*, pp. 17-32.
— [1964c], *Leksika celesoobraznoj dejatel'nosti* [The Vocabulary of Purposeful Activity], *Idem*, pp. 67-108.
Žolkovskij, A. K., N. N. Leont'eva and Ju. S. Martem'janov [1961], *O principialnom ispol'zovanii smysla pri mašinnom perevode* [On the Essential Use of Meaning in Machine Translation], in *Mašinnyj perevod*, vol. 2, pp. 17-47. NB: English translations of the four above articles are found in: V. Ju. Rozencvejg (ed.), *Essays on Lexical Semantics*, vol. 1, 1974, Stockholm, Scriptor.
Žolkovskij, A. K. and I. A. Mel'čuk [1967], *O semantičeskom sinteze* [On Semantic Synthesis], in «Problemy kibernetiki», vol. 19, pp. 177-238.