Published in: *Bulag 33 - International Symposium on Data and Sense Mining Machine Translation and Controlled Languages, and Their Application to Emergencies and Safety Critical Domains*, 2009, Besançon: Presses Universitaires de Franche-Compté, 1-10.

Functional Linguistic Models: A Step Forward in the Study of Man

Igor Mel'čuk
University of Montreal, Montreal, Canada
igor.melcuk@umontreal.ca

Abstract

The paper considers the importance of functional modeling of natural language and presents a comprehensive linguistic theory—Meaning-Text Theory [MTT], put forth in late 1960's and aimed at the construction of functional models of language, or Meaning-Text models [MTMs]. The three postulates of MTT are sketched out, followed by a sample of linguistic representations of all levels that are distinguished; then a few rules from all MTM's main modules are given. Two important linguistic phenomena are considered: paradigmatic and syntagmatic lexical choices, dealt with by using semantic decompositions and lexical functions [LFs]; LFs describe, in a systematic way, semantic derivations and collocations. A short characterization of a special type of dictionary, Explanatory Combinatorial Dictionary, presupposed by MTT and underlying MTMs, is given. Two possible ways for the verification of functional models in linguistics are suggested: computer text processing and psycholinguistic experimentation. A conclusion: functional models of languages may contribute significantly to the scientific progress by allowing for a better understanding of human thought.

1. Introduction: Models in sciences

How does one know what happened a few minutes after the Big Bang? Nobody was there, and no observation device could exist at the moment of the creation of the Universe. However, we know a lot, and this is made possible by abstract cosmological models—systems of equations, which, based on logic and known physical laws, relate to each other various facts observable hundreds of millions of years after the Big Bang. From these models, scientists can reach conclusions about the state of the Universe at different stages of its existence.

This is but one example among many that could be cited. In countless cases, a researcher who is in no position to directly observe the internal structure of an object or a phenomenon has recourse to a model. Exaggerating a bit, hard science is mainly the construction of models. This has been well known at least since Galileo. 'There is no scientist who does not reason in terms of models—even if he does not admit this to others or to himself' [7, p. 4].

Linguistics, which has natural language as its object, is in the same position as cosmology. Language, an extremely complex system of rules, is encoded in the brains of speakers and thus inaccessible to direct observation: linguists cannot open the skulls or penetrate the brain with electrodes at their will. The only solution is the recourse to models. Chomsky's Generative-Transformational Grammar has solidly implanted the idea of modeling in linguistics; already in [33] we find several articles that discuss the topic in a precise way. Consider as well the following statement by J. Molino: 'Morphology—as other domains of linguistics and as language in general—can be described only by models' [32, p. 29]. At the same time, intensive work in different branches of computational linguistics has contributed heavily to this trend. Today we can take it for granted that modeling is fully accepted in linguistics. However, it remains to be established what types of linguistic models are the most promising and to make the notion of model more specific. In fact, he term model is ambiguous; in order to eliminate confusion, a rigorous definition must be proposed.

Let there be an entity E (an object or a system ob objects); E functions in the sense that it receives observable inputs and produces for them corresponding observable outputs. The researcher is interested in the functioning of E (rather than in its internal structure, which is in any case not observable). To describe E, he constructs a functional model of E —that is, M.

> M is a functional model of E if, and only if, M is a system of symbolic expressions created by the researcher to describe the functioning of E, that is, to associate with the given inputs the same outputs as E does.

The model M is functional in two senses: 1) it seeks to represent the functioning of Y and 2) it does so by using mathematical functions, so that X is actually a very complex function in the mathematical sense.

2. Meaning-Text theory [MTT] and functional models of natural language: Meaning-Text Models [MTMs]

The MTM is characterized here in three steps: the postulates of MTT, which underlies MTMs, the linguistic representations it uses, and examples of some basic types of MTM rules.

2.1. The MTT's postulates

More than 40 years ago, the work started on the development of a functional model of natural language: MTM. The project was begum in Moscow in the 1960's by the present author, together with several colleagues, principally—A. Zholkovsky and Ju. Apresjan (see, e.g., [39], [13] - [15, pp. 12-101]; a concise overview of MTMs is offered in [10]). The linguistic theory underlying MTMs, is known as Meaning-Text theory; it is based on the following three postulates:

Postulate 1

> A natural language is a system of rules that describe the correspondence between a denumerable set of meanings and a denumerable set of texts.

Meanings (in the technical sense of the term) appear as formal symbolic objects called Semantic Representations [SemRs], and texts—as Phonetic Representations [PhonRs]. Postulate 1 can then be expressed in symbolic form as (1):

(1) $\{SemR_i\} \Leftarrow\!\textbf{language}\!\Rightarrow \{PhonR_j\} \mid i \neq j,\ 0 < i, j \leq \infty$

Logically, the Meaning-Text correspondence (i.e., "⇐**language**⇒") is bidirectional and represents equivalence; yet in linguistics it should be studied and described in the Meaning-to-Text direction: natural language is mainly about speaking, not understanding. Linguistic synthesis, or text production, is much more important for linguistics than analysis, or text understanding. The meaning-to-text orientation of linguistic research and description gives absolute priority to the study of synonymy, in particular—of linguistic paraphrase [31].

Postulate 2

> The Meaning-Text correspondence in (1) is described by a logical device, or system of rules, which constitutes a functional model of language: an MTM.

Thus, an MTM takes meanings, or SemRs, as its inputs, and produces texts, or PhonRs, as its outputs—in the same way that native speakers do. It is in this sense that an MTM is a mathematical function:

$$\mathbf{f}(SemR) = \{PhonR_j\}.$$

Applied to a SemR, it produces the set of all (nearly) synonymous PhonRs that correspond to it. (An MTM can be also used in the inverse direction: taking texts as inputs and extracting meanings from them. Here, however, only the Meaning-to-Text direction is considered.)

The Meaning-Text correspondence is many-to-many: one SemR can correspond to an astronomical number of PhonRs (several million; there is incredibly rich synonymy: see, e.g., [14]), and one PhonR can express many SemRs (ambiguity). Because of this, Postulate 3 is needed.

Postulate 3

> To successfully describe the Meaning-Text correspondence, two intermediate levels of representation are needed: Synt(actic)R, corresponding to sentences, and Morph(ological)R, corresponding to wordforms, i.e., to derivation and inflection.

As a result, an MTM has the following architecture:

(2) $\{SemR_i\} \Leftrightarrow \{SyntR_k\} \Leftrightarrow \{MorphR_l\} \Leftrightarrow \{PhonR_i\}$

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

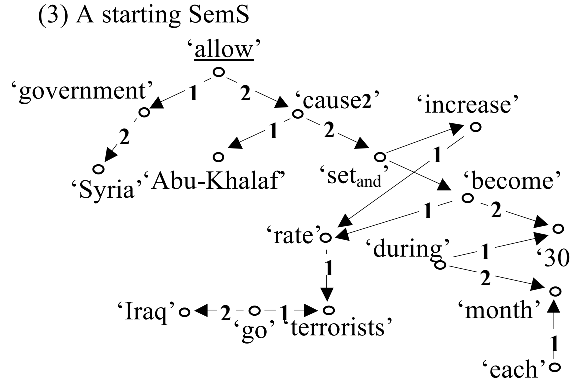
The boldfaced expressions are the names of MTM's major components, or modules.

Representations of all levels, except for the semantic level, are each subdivided into deep, or meaning-geared, and surface, or form-geared, sublevels. Including the final—phonetic/graphic—level, this gives us a total of seven representations. To avoid unnecessary complications, we will use in this discussion only graphic representations of actual sentences.

2.2. Linguistic representations in MTT

To make clearer the basic ideas of an MTM clearer, I will supply some examples of linguistic representations (and in the next subsection, a few rules relating them). Due to lack of space, many approximate, incomplete descriptions will be used and many explanations foregone.

(3) A starting SemS

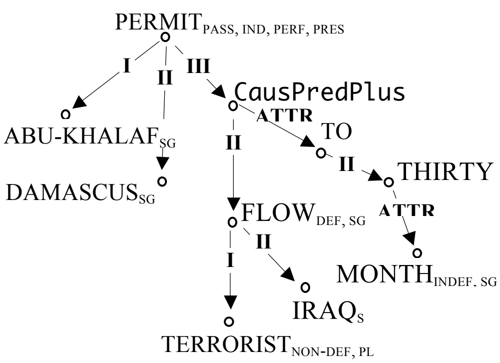


This is one of the four components of a SemR—a Semantic Structure [SemS], the main structure in a SemR; formally, it is a network whose nodes are labeled with semantemes (meanings of disambiguated lexical units [LUs]) and arcs are labeled with numbers used to distinguish the arguments of a predicate.

The other three components of a SemR—Sem-Comm(unicative)S, Rhet(orical)S and Ref(erential)S—are not shown. In what follows, only the main structures of each representation are given.

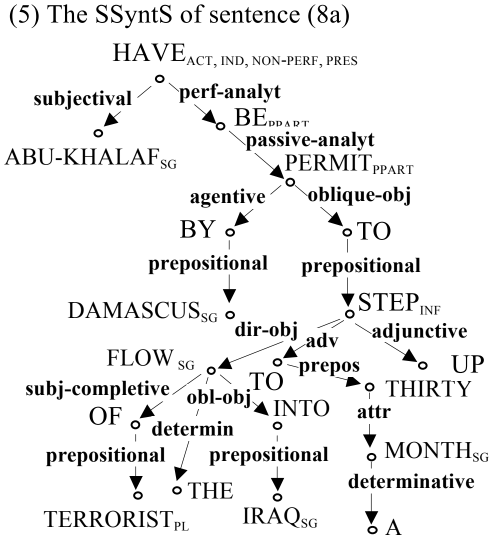
The SemS in (3) can be verbalized by a huge number of sentences, of which only three are shown in (8).

(4) The DSyntS of sentence (8a)



A DSyntS is a dependency tree; its nodes are labeled with full lexical units (including idioms and lexical functions—see below, **3.3**), and its branches, with the names of 12 universal DSynt-relations; see [21]. (Structural, i.e., 'auxiliary,' LUs are not present here.)

(5) The SSyntS of sentence (8a)



A SSyntS is also a dependency tree; but its nodes are labeled with all the actual lexemes of the sentence, and its branches, are labeled with the names of language-specific SSynt-relations. (Their number seems to be about 50 per language.)

(6) The DMorphS of sentence (8a)

[The symbols "|," "||" and "|||" stand for pauses of different duration: they represent some elements of the DPros(odic)S.]

ABU-KHALAF$_{SG}$ HAVE$_{IND, PRES, 3SG}$ BE$_{PPART}$ PERMIT$_{PPART}$
BY DAMASCUS$_{SG}$ ||
TO STEP$_{INF}$ UP |
THE FLOW$_{SG}$ OF TERRORIST$_{PL}$ INTO IRAQ$_{SG}$ |
TO THIRTY A MONTH$_{SG}$ |||

A DMorphS is a string of lexemes supplied with all the values of their inflectional categories, i.e., grammemes.

(7) The SMorphS of sentence (8a)

{A.-K.}+{SG} {HAVE}+{IND.PRES}+{3SG} {BE}+{PPART}
{PERMIT}+{PPART} {BY} {DAMASCUS}+{SG} ||
{TO} {STEP}+{INF} {UP} {THE} {FLOW}+{SG} {OF}
{TERRORIST}+{PL} {INTO} {IRAQ}+{SG} |
{TO} {THIRTY} {A} {MONTH}+{SG} |||

A SMorphS is a string of groups of morphemes that are fed to the SMorph-module of an MTM, to produce actual sentences, such as in (8a).

(8) **a**. *Abu-Khalaf has been permitted by Damascus to step up the flow of terrorists into Iraq to 30 a month.*

b. *The Syrian government let Abu-Khalaf increase the quantity of terrorists slipping into Iraq up to 30 per month.*

c. *Abu-Khalaf has the permission of the Syrian ruler to raise the number of Al-Qaeda fighters going to Iraq to 30 each month.*

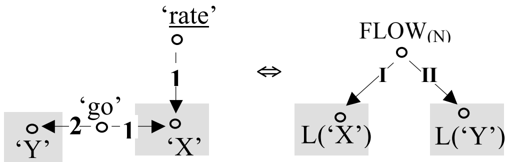
2.3. Linguistic rules in an MTM

Three major classes of linguistic rules will be illustrated: semantic, syntactic and morphological. These rules describe the mapping of a linguistic representation of the level *n* onto the representation of the level *n+1*. The rules presented below are all used in transitions between the representations in (3) through (8a).

2.3.1. Semantic Meaning-Text rules. It is impossible to give examples of every type of Sem-rules here; I will limit myself to just three.

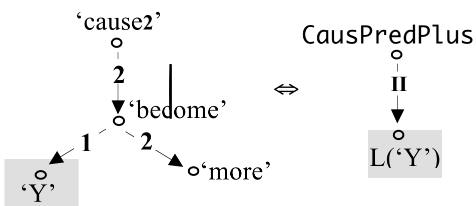
[“L(‘α’)” means ‘lexical unit that expresses the meaning ‘α’; shading marks the context of the rule.]

Individual Sem-Lexemic rule



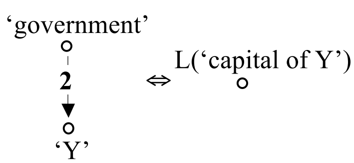
The meaning ‘rate of Xs that go to Y’ can be expressed by the noun FLOW [of Xs (in)to Y]. This is, roughly speaking, (a part of) a lexicographic entry for $\text{FLOW}_{(N)}$.

Individual Sem-Lexical-Functional rule



The meaning ‘cause Y to become more [than Y was]’ can give rise to the lexical function (see **3.2**) CausPredPlus, whose value is specified for its second argument, i.e., its DSynt-actant **II**, in the dictionary (***step up*** *the flow*; ***widen*** *the gap*; ***solidify*** *the ties*, etc.)

General Sem-Metonymic rule

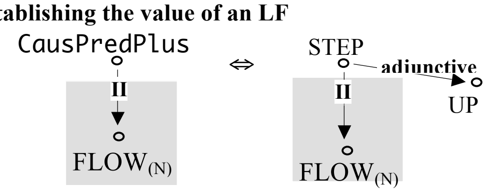


The meaning ‘government of country Y’ can be expressed by the name of the capital of Y (‘government of Russia’ ⇔ MOSCOW, ‘government of USA’ ⇔ WASHINGTON, etc.).

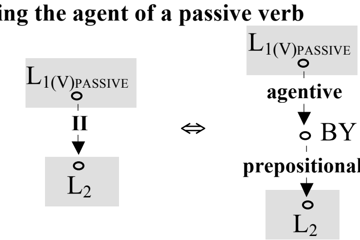
2.3.2. Syntactic Meaning-Text rules. For syntactic rules, it is necessary to present DSynt- and SSynt-rules; I will give two rules of each type.

DSyntS ⇔ SSyntS (Deep Syntax)

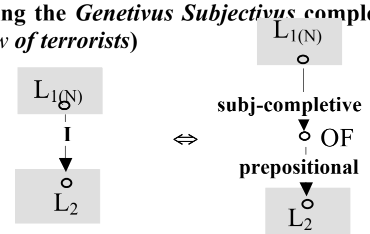
Establishing the value of an LF



Expressing the agent of a passive verb



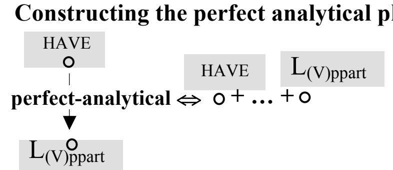
Expressing the *Genetivus Subjectivus* complement of a noun (e.g., *flow **of terrorists***)



SSyntS ⇔ DMorphS (Surface Syntax)

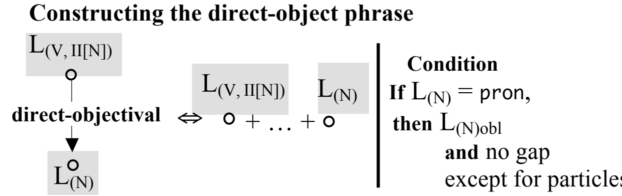
The symbol “+” specifies linear order, and “…”—a possible gap.

Constructing the perfect analytical phrase



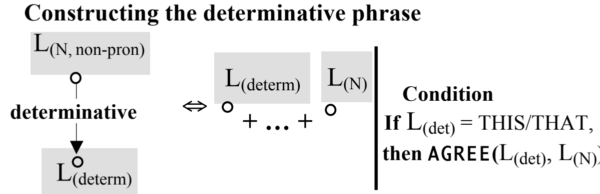
(***Having****, as everybody knows,* ***written*** *to Father*)

Constructing the direct-object phrase



(***step*** *up the* ***flow***; *Have you* ***seen*** *(just)* ***him****?*)

Constructing the determinative phrase



(***the laws***; ***these books***; ***a flow***)

For more on surface syntax, see [28] and [25].

2.3.3. Morphological Meaning-Text rules. DMorph- and SMorph-rules are illustrated below.

DMorphS ⇔ SMorphS (Deep Morphology)

PL ⇔ {PL}

PAST ⇔ $\text{A}^{\text{I}\Rightarrow\text{æ}}$ (apophony of the *sing* ~ *sang* type)

DMorphS ⇔ SMorphS (Deep Morphology)

{PL} ⇔ /z/ | after $/\text{C}_{[+\text{voiced}],\ [-\text{hushing}/-\text{whistling}]}/$

/s/ | after $/\text{C}_{[-\text{voiced}],\ [-\text{hushing}/-\text{whistling}]}/$

/ɪz/ | after $/\text{C}_{[+\text{hushing}/+\text{whistling}]}$

$\mathbf{A}^{\mathbf{I \Rightarrow æ}}$ ⇔ /ɪ/ ⇒ /æ/

There is no need to enter into the details of Meaning-Text morphology since the interested reader has [17] and [22] at his disposal.

To conclude this section, a general architecture of an MTM is presented in Fig. 1 (see next page).

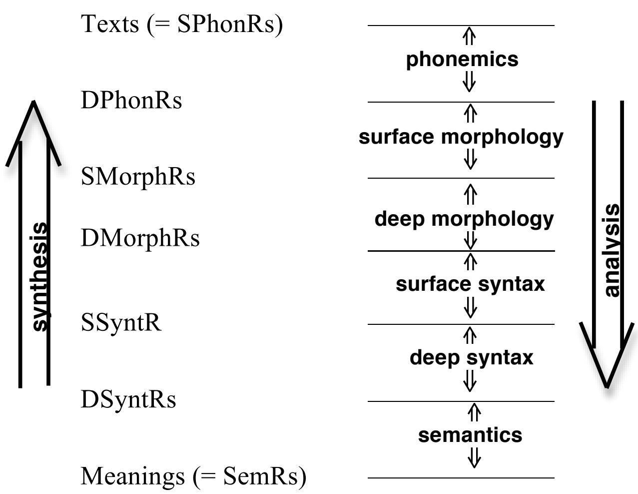


Figure 1. The General Structure of a Meaning-Text Model

Note that **conceptics** (= the module describing the correspondence between a conceptual representation of reality and the SemR of the utterance) and **phonetics** (= the module repsonsible for the correspondence between the SPhonR and actual articulated sound) remain outside of an MTM. They belong to a more general model of linguistic behavior—a Reality-Speech model.

3. Modeling two important linguistic phenomena: Paradigmatic and syntagmatic lexical choices

As we have known since F. de Saussure and R. Jakobson, any linguistic activity is carried out by the Speaker along two axes: paradigmatic, where the selection of linguistic units happens, and syntagmatic, where the units selected are combined to produce utterances. To highlight certain advantages of MTMs, one particular aspect of this activity can be considered: lexical choices. This is reputedly one of the hardest nuts for any linguistic theory to crack.

3.1. Paradigmatic lexical choices

Two types of paradigmatic lexical choices are known: free choices, where an LU L is selected directly and only for its meaning; and restricted choices, where an LU L is selected instead of another LU L´ as L´s function (L replaces L´ in the text.) Free choices are modeled by means of semantic decompositions, and restricted choices, by means of semantic derivations.

3.1.1. Semantic decomposition. The sentences (9a) and (9b) are synonymous:

(9) **a**. *John is sure that Mary is in town.*
b. *John has no doubt that Mary is in town*.

This means that 1) they are mutually substitutable in texts; 2) their negations are also synonymous: *John is not sure that P* ≡ *John has doubt that P*; 3) both sentences are not factive and therefore can be continued by *…but this is not true*. But what type of information should a Speaker have in his brain about the lexemes SURE and DOUBT in order to be able to manipulate them as he actually does? We cannot know for sure; but we can propose a plausible model.

Following A. Zholkovsky, A. Bogusławski and A Wierzbicka, MTT proposes that the meanings of these lexemes (like all lexical meanings, i.e., all semantemes) consist of simpler meanings—in other words, that meanings are decomposable. (Note that these simpler meanings are decomposable into even simpler meanings and so forth, until semantic primitives, or meaning atoms, are reached [16].) Consider the following dataset (the asterisk indicates an unacceptable continuation; the verb BELIEVE is taken in the sense ‘hold a belief’):

(10) **a**. *I believe that Alan has come, but I am not sure.*
b. *I am sure that Alan has come, *but I do not believe this.*
c. *I believe that Alan has come,*but I have some doubts.*
d. *I am sure that Alan has come.* ≡ *I have no doubt that Alan has come.*
e. *I am not sure that Alan has come.* ≅ *I have doubt that Alan has come.*

To enable a human (or a computer) to compute all the expressions in (10) and to establish their acceptability and synonymy, it is sufficient to represent the meanings under consideration as follows:

(11) **a**. *X is sure that P* ≡ ‘Holding the belief «P takes place», X is not prepared to admit that P does not take place’.
b. *X has doubt that P* ≡ ‘Not holding the belief «P takes place», X is prepared to admit that P does not take place’.

With these definitions, one obtains for sentences in (10) the following semantic decompositions:

(12) **a**. ‘I believe that A. has come, but holding the belief «A. has come», I am prepared to admit that A. hasn’t come’.

b. ‘**Holding the belief «A. has come»**, I am not prepared to admit that A. hasn’t come, *but **I do not believe that A. has come**’ (a contradiction, marked in boldface)

c. ‘**I believe that A. has come**, *but **not holding the belief «A. has come»**, I am prepared to admit that A. hasn’t come’ (another contradiction)

d. ‘Holding the belief «A. has come», I am not prepared to admit that A. hasn’t come’. ≡ ‘Holding the belief «A. has come» [this semantic component is the result of double negation], I am not prepared to admit that A. hasn’t come’.

e. ‘Holding the belief «A. has come», I am prepared to admit that A. hasn’t come’. ≅ ‘Not holding the belief «A. has come», I am prepared to admit that A. hasn’t come’.

In (10e), the two sentences are roughly synonymous, but less so than the two sentences in (10d): the second sentence in (10e) expresses a slightly higher level of uncertainty than the first one. This is borne out by the decomposition in (12e): in both sentences, I am prepared to admit that A. has not come; however, the first one expresses the belief «A. has come», while the second one does not. Thus, the proposed decompositions allow for a formal and coherent description of the data in (10). Semantic decompositions constitute an important descriptive tool in MTT.

3.1.2. Semantic derivation. Suppose a speaker wants to talk about the person who is at the wheel of a car; he needs the lexeme DRIVER. But a person ‘driving’ a locomotive is called an ENGINEER, the one ‘driving’ a plane, a PILOT and a ship, a CAPTAIN. Similarly, a person who manages a farm is a FARMER, and the one managing a shop, is a SHOPKEEPER. On the other hand, the client of a restaurant is a CLIENT or PATRON, that of a hospital, a PATIENT, of a prison, a PRISONER or INMATE, of a theater, a SPECTATOR, of a school, a STUDENT, etc. We see that there are regular semantic links of the type CAR ~ DRIVER, LOCOMOTIVE ~ ENGINEER, PLANE ~ PILOT, SHIP ~ CAPTAIN: they remind one of derivation, except that in this case there can be no formal similarity between the members of the pair. For the purposes of formal modeling of text production such lexical relations should be explicitly specified. MTT proposes to do that by means of lexical functions [LFs], which are introduced in the next subsection; more precisely, these are paradigmatic LFs.

3.2. Syntagmatic lexical choices

Restricted lexical cooccurrence is a serious problem for any lexicographic description. Thus, English says *MAKE a mistake* but *DO a favor*, *TAKE an action*, but *BE ENGAGED in an activity*, *HOLD influence over N*, but *GIVE N orders*, etc. Similarly, we have *a GREAT achievement*, *a DRASTIC action*, *a FIRM believer*, *SOLID grounds*, *a HARSH protest*, *quick AS LIGHTNING*, *rain HARD*, etc. MTT puts forth the following hypothesis:

> In the majority of cases, restricted (i.e., synchronically arbitrary) lexical cooccurrence manifests itself in the expression of a limited number of very abstract, ‘nearly-grammatical’ meanings.

Thus, in the first series of examples above, this meaning is ≈ ‘do’, and in the second, ≈ ‘very/intense’.

The crucial fact is that such a meaning corresponds to a function (in the mathematical sense): the lexical unit is its argument and the cooccurrents themselves, (elements for which the appropriate cooccurrent(s) must be selected of) the value, so that we have $\mathbf{f}(L) = \{L_1, L_2, ..., L_n\}$. For the above examples, two such functions are proposed—these are simple standard lexical functions (see next section):

— A support verb $Oper_1$:

$Oper_1$(*mistake*) = *make* [ART ~]
$Oper_1$(*favor*) = *do* [ART ~]
$Oper_1$(*rage*) = *be* [*in* ART ~]
$Oper_1$(*action*) = *take* [ART ~]
$Oper_1$(*activity*) = *be engaged* [*in* ART ~]

— An intensifier Magn:

Magn(*achievement*) = *great*
Magn(*action*) = *drastic*
Magn(*believer*) = *firm*
Magn(*quick*) = *as lightning*
Magn(*dispute*$_V$) = *hotly*

The expressions described by these LFs are nothing other than collocations: the collocation’s base, selected independently by the Speaker for its meaning, corresponds to the argument of the LF, and the collocate, selected as a function of the base, is one of the elements of its value. (For more on collocations in the MTT framework, see [1], [2] and [20].)

3.3. Lexical functions

Three properties of standard lexical functions [LFs] prove to be especially important:

- LFs are not numerous: between 50 and 60
- LFs are cross-linguistically valid

• LFs are equally convenient for the description of both paradigmatic and syntagmatic restricted lexical choices. In other words, they allow for a homogeneous and systematic description of semantic derivations and collocations.

Several examples of LFs will make the corresponding notion clearer. (For more on LFs, see [39], [36], [18], [19], [23, pp. 275*ff*] and [24].)

Paradigmatic LFs

1. Action/property noun S_0

S_0(*accept*) = *acceptance* S_0(*intrude*) = *intrusion*
S_0(*capable*) = *capacity* S_0(*angry*) = *anger*

2. Patient noun S_2

S_2(*award*) = *recipient* S_2(*shoot*) = *target*
S_2(*sell*) = *merchandise* S_2(*talk* [to]) = *addressee*

3. Active possibility adjective $Able_1$

$Able_1$(*harm*) = *harmful* $Able_1$(*rebellion*) = *restive*
$Able_1$(*war*) = *bellicose* $Able_1$(*vomit*) = *queasy*

Syntagmatic LFs

1. Positive evaluation adjective Bon

Bon(*contribution*) = *valuable* Bon(*service*) = *quality*
Bon (*idea*) = *promising* Bon(*weather*) = *fine, nice, lovely, …*

2. Support verbs $Oper_i$, $Func_i$ and $Labor_{ij}$

$Oper_1$(*apology*) = *offer* [ART ~]
$Oper_2$(*apology*) = *receive* [ART ~]
$Func_1$(*support*) = *comes* [from N_X]
$Func_1$(*support*) = *goes* [to N_Y]
$Labor_{12}$(*inheritance*) = *leave* [N_Y as ~ to N_Z]
$Labor_{12}$(*inheritance*) = *receive* [N_Y as ~ from N_X]

3. Realization verbs $Real_i$, $Fact_i$ and $Labreal_{ij}$

$Real_1$(*duty*) = *discharge* [N_X's ~]
$Real_1$(*appointment*) = *keep* [ART ~]
$Real_2$(*treatment*) = *respond* [to ART ~]
$Fact_0$(*film*) = *is playing, is in the theaters*
$Fact_1$(*river*) = *empties* [into N_X]
$Fact_2$(*bomb*) = *falls* [on N_Y]
$Labreal_{12}$(*artillery*) = *hit* [N_Y with ~]
$Labreal_{21}$(*invitation*) = *take up* [N_X on $A_{poss\text{-}X}$ ~]

4. Locative/temporal preposition Loc_{in}

Loc_{in}(*list*) = *on* [ART ~] Loc_{in}(*end*) = *at* [ART ~]
Loc_{in}(*program*) = *on* [ART ~] Loc_{in}(*diagram*) = *in* ART ~]
Loc_{in}(*holiday*) = *on* [~] Loc_{in}(*socialism*) = *under* [~]
Loc_{in}(*past*) = *in* [ART ~] Loc_{in}(*NUM o'clock*) = *at* [~]

The nature of this paper does not allow us to touch on several interesting aspects of LFs: standard *vs.* non-standard LFs, simple *vs.* complex LFs, configurations of LFs, fused elements of value of LFs, etc.

3.4. Correlations between the meaning and the collocates of an LU

The proposed description of lexical meaning and restricted lexical cooccurrence leads to sharpening of lexicographic definitions. Take, for instance, the noun APPLAUSE. In [12], it is defined as 'the sound of many people hitting their hands together and shouting, to show that they have enjoyed something'. This definition would be OK, if it weren't for the LFs Magn/AntiMagn of the noun APPLAUSE: *deafening*, *rapturous* vs. *thin*, *scattered*, etc. These adjectives indicate that the applause is gradable: the strength and frequency of hitting hands together is (roughly) proportional to the approval/enjoyment by the applauder. Therefore, the definition of APPLAUSE (and that of the verb APPLAUD) must be corrected:

X applauds Y ≡ 'X claps hands to express X's approval/ enjoyment of Y, the strength and frequency of clapping being proportional to X's approval/enjoyment'.

In this way, the definitions and collocations are buttressing each other; on this topic, see [9] and [8]. Such a link is vital for the description of language, a system 'où tout se tient' [Saussure].

3.5. The Explanatory Combinatorial Dictionary

All types of information about an LU L necessary to ensure the correct use of L in any context are stored in a dictionary of a special type: the Explanatory Combinatorial Dictionary [ECD]. It is explanatory since each L receives in it a semantic decomposition (L's 'explanation'); it is combinatorial because it specifies for each L its syntactic and lexical cooccurrence. In the framework of MTT, such a lexicon plays the central role: L's lexical entry contains all the semantic data concerning L and all the combinatorial data that are used by MTM grammatical rules. In this sense, MTT is lexically based.

Since the ECD has been described in detail in numerous publications ([13, pp. 110-140], [26], [27], [29], [30]), suffice it here to state its defining properties:

General properties of the ECD

• the ECD is a theoretical lexicon elaborated within the framework of a full-fledged linguistic theory (MTT)
• the ECD is a formalized lexicon, written in terms of several lexicographic metalanguages
• the ECD is complete at the level of each entry

Specific properties of the ECD

• the ECD is an active dictionary, supplying all the data in the direction Meaning-to-Text

• the ECD is semantic: the definition (= SemR) of the headword L underlies and determines L's entry

• the ECD fully covers L's restricted lexical cooccurrence—in terms of LFs

• the ECD treats lexemes and idioms in the same way: all of them constitute headwords of the corresponding entries

• each ECD's entry describes a (monosemous) LU; LUs related to each other by polysemy are united within a superentry, called a vocable.

4. Validation of functional models in linguistics

An MTM of a natural language is speculative by its very nature: analyzing speakers' behavior, the linguist observes the associations between (understood) meanings and (perceived) texts and makes inferences as to the underlying representations and formal rules that relate them. The question naturally arises: How can we validate the model being proposed? At the time being, at least two experimental techniques are available:

• Computerization of linguistic MTMs and the use of the resulting systems in all branches of Natural Language Processing: machine translation, text generation, question-answering, etc. (For some applications of MTT in natural language processing, see, for instance, [3] - [5] and [11], as well as the collections [35] - [37]).

• Psycholinguistic experimentation, which could shed precious light on the psychological reality of the basic oppositions and descriptive formal objects put forth in MTT (see [6]). Thus, psycholinguistic experiments may contribute to our knowledge of whether (or to what extent) it is correct to insist, as MTT does, on the following five oppositions:

1) Linguistic synthesis (Meaning ⇒ Text) *vs.* linguistic analysis (Text ⇒ Meaning).

2) Static linguistic knowledge *vs.* dynamic procedural knowledge. The information of the first type is specified in linguistic rules, which constitute the MTM itself; the information of the second type is embodied in algorithmic rules that manage the application of linguistic rules.

3) Linguistic representations of various levels *vs.* modules of the MTM that relate them.

4) Semantic representation, which targets meaning of sentences, ignoring their 'physical' organization, *vs.* syntactic representation, which targets the structure of sentences, ignoring their meaning.

5) The lexicon, where all data concerning individual LUs are stored, *vs.* grammar, which present the information about classes of LUs and 'grammatical' signs (affixes, apophonies, conversions and meaningful syntactic constructions).

It seems crucial to know whether the actual behavior of speakers is based on these oppositions. Thus, we need to know more (much more!) on the psychological and neurological differences between speakers' encoding of the and decoding of texts, on the way the strictly linguistic data are stored in the brain in contrast to procedural knowledge, etc.

At the same time, the psychological correlates of MTT's descriptive formal objects are no less interesting:

— The SemS: Is meaning represented in the brain by networks similar to those of MTT? Do speakers use semantic decompositions? If so, exactly how? Is it true that the production of a sentence begins with the shallowest and quite approximate SemS available (as MTT has it)?

— The SyntS: Is a sentence represented in the brain by a dependency tree similar to that of MTT?

— The lexicon: Is the storage of lexical information in the brain isomorphic to what is presupposed by the ECD of MTT? What are neurological differences between encoding the meaning of sentences *vs.* encoding the meaning of LUs in the lexicon? What are the mechanisms allowing the speaker to apply the descriptions of LUs in his brain to the starting SemS in order to produce the DSyntS of the future sentence? How is the interaction between the starting SemS and the sentence under production carried out?

The studies into the acquisition of language by children and adult learners, into aphasic disorders, into diachronic developments, etc. also could contribute their share to the acceptance/rejection of a given functional model.

To sum up: Functional models in linguistics, including MTMs, do not lack ways and means of validation.

5. Conclusion: The value of functional models in linguistics

A functional model of a natural language is of high practical utility in at least three technological and social domains, which are:

• Natural language processing (see section **4**).

• Teaching and learning languages (as an example of the application of MTT to teaching linguistics, see [34]).

• Manufacturing reference books, such as dictionaries of all types, pedagogical grammars, and manuals.

The formal character of MTMs and their orientation ('How is such-and-such a thought expressed in such-and-such a language?') are especially valuable in this connection.

The theoretical impact of MTMs of natural languages appears to me even more important. Scientific progress until today has been basically addressing the problems of the physical universe: matter and energy. Since *Homo sapiens* started speaking, we have developed new means of transportation (including spacecrafts), enhanced our physical strength manifold (remember the H bomb!), improved our organs of perception (electronic microscopes and radiotelescopes), widened our communication abilities (electronic media, the Internet), etc. We have penetrated the atom and the depths of the Universe; we know a lot about the origins of our world and the structure of our genes. But we have as yet made no comparable headway in the mastery of information (in the scientific sense)—this evasive 'substance,' which is so central to life in general and to the life of humans in particular. We do not know enough about the workings of our brain, while the enhancement of the brain remains task number one for today's science. Facing the challenges of the 21st century, the humanity badly needs good models of human thinking and reasoning (and, why not, of human emotions). This seems to be well understood by the international scientific community, and the majority of scientists would probably be in agreement with such a program.

However, strangely enough, people tend to forget — or disregard?—this vital fact:

> The only reliable key to human thinking, in all its complexity, is natural language.

Without an understanding of how language is functioning in our psyche, there will be no good understanding of information processing by the human brain. That is why I believe that functional models of language, and MTMs in particular, nowadays have acquired quite a special significance. Linguistics must take a place of honor among the 'hard' sciences, and functional models, which embody the typical scientific approach to complex phenomena, will make their contribution.

Acknowledgments

I most cordially thank D. Beck, J. Milićević and L. Wanner, who have read and commented upon the preliminary versions of this paper.

6. References

Presenting a complete list of relevant publications here is out of the question. Only a few (mainly those explicitly related to the Meaning-Text approach) have been selected.

[1] M. Alonso Ramos, "Elaboración del Diccionario de colocaciones del español", in: P. Bataner and J. de Cesaris, eds, *De lexicographia. Actos del I Symposium internacional de lexicografía*, IULA/Petició, Barcelona, 2004, pp. 149-162.

[2] M. Alonso Ramos, *Las construcciones con verbo de apoyo*, Visor Libros, Madrid, 2004.

[3] Ju. Apresjan, I. Boguslavsky, L. Iomdin, A. Lazurskij, L. Mitjushin, V. Sannikov, and L. Tsinman, "ETAP-2: The Linguistics of a Machine Translation System", *META*, 37: 1, 1992, pp. 97-112.

[4] Ju. Apresjan, I. Boguslavsky, L. Iomdin, A. Lazurskij, V. Sannikov, V. Sizov, and L. Tsinman, "ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT", in: S. Kahane and A. Nasr, eds., *Meaning-Text Theory. First International Conference on Meaning-Text Theory [MTT'03]*, École Normale Supérieure, Paris, 2003, pp. 279-288.

[5] Ju. Apresjan and L. Cinman, "Formal´naja model´ perefrazirovanija predloženij dlja sistem pererabotki tekstov na estestvennyx jazykax [A Formal Model of Sentence Paraphrasing for Natural Language Text Processing Systems]", *Russkij Jazyk v Naučnom Osveščenii*, No. 2 [= 4], 2002, pp. 102-146.

[6] Ju. Apresjan and I. Mel'čuk, "Computer Implementation of a Linguistic Model and the Problem of its Psychological Reality", *Folia Linguistica*, 24: 3-4, 1990, pp. 432-446.

[7] P. Auger, "Les modèles dans la science", *Diogène*, nº 52, 1965, pp. 3-15.

[8] L. Iordanskaja, "Lexicographic Definition and Lexical Cooccurrence: Presuppositions as a 'No-Go' Zone for the Meaning of Modifiers", in: K. Gerdes, T. Reuther, L. Wanner, eds, *Meaning-Text Theory 2007 [MTT'07]*, Wiener Slawistischer Almanach, München/Wien, 2009, p. 00-00.

[9] L. Iordanskaja and A Polguère, "Hooking Up Syntagmatic Lexical Functions to lexicographic Definitions", in: Ju. Apresjan and L. Iomdin, eds, *East-West Encounter, Second International Conference on Meaning-Text Theory [MTT'05]*, Jazyki slavjanskoj kul´tury, Moskva, 2005, pp. 170-186.

[10] S. Kahane, "The Meaning-Text Theory", in: Agel *et al.*, eds, *Dependency and Valency. An International Handbook of Contemporary Research*, vol. 1, Berlin—New York, W. de Gruyter, 2003, pp. 546-570.

[11] F. Lareau and L. Wanner, "Towards a Generic Multilingual Dependency Grammar for Text Generation", in: T. Holloway King and E. Bender, eds, *Proceedings of the GEAF 2007 Workshop, CSLI Studies in Computational Linguistics Online*, http://csli-publications.stanford.edu/GEAF/2007/papers/geaf07lareauwanner.pdf

[12] *Longman Dictionary of Contemporary English*, www.ldoceonline.com

[13] I. Mel'čuk, *Opyt teorii lingvističeskix modelej "Smysl⇔* Tekst" [Outline of a Theory of Meaning-Text Type Linguistic Models]. Nauka, Moskva, 1974 [1995].

[14] I. Mel'čuk, "Meaning-Text Models: A Recent Trend in Soviet Linguistics", *Annual Review of Anthropology*, v. 10, 1981, pp. 27-62.

[15] I. Mel'čuk, *Dependency Syntax: Theory and Practice*, The SUNY Press, Albany (N.Y.), 1988.

[16] I. Mel'čuk, "Semantic Primitives from the Viewpoint of the Meaning-Text Linguistic Theory", *Cuaderni di semantica*, 10, 1989, pp. 65-102.

[17] I. Mel'čuk, *Cours de morphologie générale, vol. 1 - 5*, Les Presses de l'Univ. de Montréal, Montréal, 1992 - 2000.

[18] I. Mel'čuk, "Phrasemes in Language and Phraseology in Linguistics", in: M. Everaert, E.-J. van der Linden, A. Schenk and R. Schreuder, eds, *Idioms. Structural and Psychological Perspectives*, Lawrence Erlbaum, Hillsdale (N.J.)—Hove, 1995, pp. 167-232.

[19] I. Mel'čuk, "A Tool for the Description of Lexical Relations in the Lexicon", 1996, in: [35, pp. 37-102].
[20] I. Mel'čuk, "Collocations and Lexical Functions", in: A. Cowie, ed., *Phraseology. Theory, Analysis, and Applications*, Clarendon Press, Oxford, 1998, pp. 23-53.
[21] I. Mel'čuk, "Actants in Semantics and Syntax. I/II: Actants in Semantics/in Syntax", *Linguistics*, 42: 1, 2004, pp. 1-66; 42: 2, 2004, pp. 247-291.
[22] I. Mel'čuk, *Aspects of the Theory of Morphology,* Mouton de Gruyter, Berlin/New York, 2006.
[23] I. Mel'čuk, *Explanatory Combinatorial Dictionary*, in: G. Sica, ed., *Open Problems in Linguistics and Lexicography*, Polimetrica, Monza (Italy), 2006, pp. 225-355. Also: http://www.polimetrica.com/?p=productsList&sWord=lexicography
[24] I. Mel'čuk, "Lexical Functions", in: H. Burger, D. Dobrovol'skij, P. Kühn and N. Norrick, eds, *Phraseology. An International Handbook of Contemporary Research*, de Gruyter, Berlin/New York, 2007, pp. 119-131.
[25] I. Mel'čuk, "Dependency in Natural Language", 2009, in: [35, pp. 1-110].
[26] I. Mel'čuk *et al*., *Dictionnaire explicatif et combinatoire du français contemporain : Recherches lexico-sémantiques I – IV*, Les Presses de l'Univ. de Montréal, Montréal, 1984-1999.
[27] I. Mel'čuk, A. Clas and A. Polguère, *Introduction à la lexicologie explicative et combinatoire*, Duculot, Louvain-la-Neuve, 1995.
[28] I. Mel'čuk and N. Pertsov, *Surface Syntax of English. A Formal Model within the Meaning-Text Framework*, Benjamins, Amsterdam, 1987.
[29] I. Mel'čuk and A. Polguère, *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*, De Boeck, Bruxelles, 2007.
[30] I. Mel'čuk and A. Zholkovsky, *Explanatory Combinatorial Dictionary of Modern Russian*, Wiener Slawistischer Almanach, Vienna, 1984.
[31] J. Milićević, *La paraphrase. Modélisation de la paraphrase langagière*, P. Lang, Bern, 2007.
[32] J. Molino, "Où en est la morphologie ?", *Langages*, nº 78, 1985, pp. 5-40.
[33] E. Nagel, P. Suppes and A. Tarski, eds, *Logic, Methodology and Philosophy of Science*, Stanford University Press, Stanford (CA), 1962.
[34] A. Polguère, *Lexicologie et sémantique lexicale. Notions fondamentales*, Les Presses de l'Univ. de Montréal, Montréal, 2008.
[35] A. Polguère and I. Mel'čuk, eds, *Dependency in Linguistic Description*, Benjamins, Amsterdam/Philadelphia, 2009.
[36] L. Wanner, ed., *Lexical Functions in Lexicography and Natural Language Processing*. Benjamins, Amsterdam/Philadelphia, 1996.
[37] L. Wanner, ed., *Recent Trends in Meaning-Text Theory*. Benjamins, Amsterdam/Philadelphia, 1997.
[38] L. Wanner, ed., *Selected Lexical and Grammatical Issues in the Meaning–Text Theory: In honour of Igor Mel'čuk*, Benjamins, Amsterdam/Philadelphia, 2007.
[39] A. Žolkovskij and I. Mel'čuk, "O semantičeskom sinteze" [On Semantic Synthesis], *Problemy kibernetiki*, v. 19, 1967, pp. 177-238.