A SHORT GUIDE

TO THE MEANING-TEXT LINGUISTIC THEORY

JASMINA MILIĆEVIĆ

DALHOUSIE UNIVERSITY
HALIFAX (CANADA)

Abstract

The paper presents the Meaning-Text linguistic theory, a theoretical framework for the construction of models of natural languages, called Meaning-Text Models. Since its beginnings, in the 1960's, the Meaning-Text theory has placed strong emphasis on semantics and considered natural language primarily as a tool for expressing meaning. This basic insight underlies its interest in linguistic synthesis (rather than analysis), paraphrase (synonymy of linguistic expressions, in particular of full sentences) and the lexicon. The Meaning-Text theory has always considered relations (rather than classes) to be the main organizing factor in language and has made an extensive use of the concept of linguistic dependency, in particular of syntactic dependency (vs. constituency). Thus, it has in many ways anticipated current developments in linguistics. Due to a formal character of the Meaning-Text theory and the corresponding models, the latter have been successfully applied in Natural Language Processing, in particular automatic text generation and machine translation.

The paper is organized in five sections: 1. Natural language viewed as a Meaning-Text correspondence (postulates of the theory); 2. Meaning-Text Models of natural languages (characteristics of the models: levels of linguistic representation and rules which establish correspondences between them); 3. Illustration of the linguistic synthesis in the Meaning-Text framework; 4. Summary of the main features of the Meaning-Text theory; 5. Basic Meaning-Text bibliography.

Keywords:
communicative structure, dependency, lexicon, linguistic models of natural languages, Meaning-Text linguistic theory, paraphrase, semantics, semantic/syntactic representation.

The Meaning-Text linguistic theory [= MTT] is a theoretical framework for the description of natural languages, more precisely, for the construction of models of languages—Meaning-Text models. Launched in Moscow in the 1960'/early 1970' (Žolkovskij & Mel'čuk 1967, Mel'čuk 1974), the MTT has been developed in Russia, Canada and Europe.

The MTT provides a large and elaborate basis for linguistic description and, due to its formal character, lends itself particularly well to computer applications. However, until recently it remained relatively marginal, mainly because of the fact that its philosophy is radically different from that of mainstream, i.e., generative, American linguistics. Since the last decade, the MTT has

enjoyed an increasing popularity, as witnessed by a growing number of MTT-minded publications and regularly scheduled international conferences (Paris 2003, Moscow 2005, Klagenfurt 2007). By presenting the MTT to researchers in Korea, where this theory still has not found a foothold, the present paper aims to contribute to a further dissemination of its ideas.

The structure of the paper is as follows: Postulates of the MTT (Section 1); main characteristics of Meaning-Text Models (Section 2); Illustration of linguistic synthesis in the Meaning-Text framework (Section 3); Summary of the main features of the MTT (Section 4); Basic Meaning-Text publications (Section 5).

1. Natural language viewed as a Meaning-Text correspondence

The MTT is based on the following three postulates.

Postulate 1

> Natural language is (considered as) a many-to-many correspondence between an infinite denumerable set of meanings and an infinite denumerable set of texts.

Meaning is, roughly, a linguistic content to be communicated (in R. Jakobson's terms, *something intelligible, i.e., translatable*), and *text* is any fragment of speech, of whatever length (again, in Jakobson's terms, *something immediately perceptible*). Both meanings and texts are taken to be directly accessible to the speaker, and, therefore, to the researcher; they constitute the linguistic data.

The correspondence between meanings and texts is *many-to-many* because a given meaning can be expressed by different texts (synonymy) and a given text can correspond to different meanings (ambiguity, i.e., homonymy or polysemy).

The MTT does not deal with meanings/texts in their neurological/acoustic reality, but rather with *representations* of meanings/texts, more precisely, with their descriptions by means of formal languages, devised specifically for that purpose. To represent a meaning, a formal object, called *Semantic Representation* [= SemR] is used, and, similarly, to represent a text—a *Surface-Phonological*, or *Phonetic*, *Representation* [= PhonR]; thus, Postulate 1 can be symbolically presented as follows:

$$\{SemR_i\} <=language=> \{PhonR_j\}.$$

Postulate 2

> The Meaning-Text correspondence is described by a formal device which simulates the linguistic activity of the native speaker—*a Meaning-Text Model*.

A Meaning-Text Model [= MTM] must be able to produce, for any given representation of meaning, all synonymous texts (= paraphrases) which implement it, and, conversely, to extract, from any given text, all its underlying meaning representation(s)—exactly what the native speaker can do with his/her language.

Although the inputs to and the outputs of an MTM (i.e., respectively, meanings and texts) are accessible to the speaker, the rules that link them (i.e., the correspondence itself) are not. For this reason, all an MTM can do is simulate, or approximate in the best way possible, the Meaning-Text correspondence; in other words, an MTM is a functional, rather than structural, model of language.

No strong claims can be made for the time being as to the psychological reality of such a model, because no corresponding psycholinguistic investigations have been undertaken to verify whether an MTM reflects the processes that take place in the brain of the speaker when s/he goes from meanings to texts, and vice versa. However, the philosophy of the approach is such that it is geared to what is happening in the brain and invites phycholinguistic and neurological verifications. For the same reason, the MTM admits *introspection* as one of the most important methods of linguistic investigation.

Postulate 3

> Given the complexity of the Meaning-Text correspondence, intermediate levels of (utterance) representation have to be distinguished: more specifically, a Syntactic and a Morphological level.

The two intermediate representation levels correspond to two autonomous domains of linguistic organization: sentence and word.

All levels, except the semantic one, are further split into deep- and surface-(sub)levels, the former oriented towards the meaning (= content of expression), and the latter towards the text (= form of expression). This gives us a total of seven levels of representation (of utterances): Semantic, Deep and Surface Syntactic, Deep and Surface Morphological, Deep and Surface Phonological.

To the three above postulates, the following methodological principle is added:

> The Meaning-Text correspondence should be described in the direction of synthesis, i.e., from Meaning to Text (rather than in that of analysis, i.e., from Text to Meaning).

This, of course, is not a logical necessity, the linguistic correspondence being bi-directional, but a matter of choice; this choice is guided by linguistic considerations, of which I will mention two.

1) Producing speech is an activity that is more linguistic than understanding speech. Ideally, the speaker knows what s/he wants to express and needs only purely linguistic knowledge to construct the utterance. In contrast, understanding an utterance implies having recourse to extralinguistic knowledge—logical, pragmatic, and the like—in addition to purely linguistic one. This makes the Meaning-Text correspondence easier to study in the direction of synthesis.

2) Some linguistic phenomena can be discovered only from the viewpoint of synthesis; thus, the relevance and the difficulty of studying restricted lexical co-occurrence (i.e., collocations, such as *do a favor*, *make a mistake*, *file a complaint*, etc.) become apparent only if we adopt a Meaning-to-Text perspective.

Therefore, for the MTM, the main question *is How is a meaning M expressed in the language L?*, rather than *What does the expression E of the language L mean?*

A corollary of this is that the study of paraphrases (= synonymous linguistic expressions, in particular synonymous sentences) occupies the central place in the Meaning-Text framework.

It is well known fact that synonymy is a fundamental semantic relation in natural language, equally important for its acquisition and use (Žolkovskij & Mel'čuk, 1967). Languages have extremely rich synonymic means: almost any single (relatively complex) meaning can be implemented by an astonishingly high number of synonymous surface expressions. Given this, it is not exaggerated to say that to model a language means to describe its synonymic means and the ways it puts them to use.

The MTM takes this challenge seriously; as we shall see, this theory defines meaning as the invariant of paraphrases, regards the production of speech as 'virtual paraphrasing,' i.e., as a series of choices between possible synonymous expressions of a starting linguistic meaning, and systematically uses paraphrase as the main research tool in linguistics. It is important to note that we are talking here about a fairly sophisticated type of paraphrase—lexical paraphrase, which essentially involves semantic decompositions of lexical meanings.

2. Meaning-Text Models

The characteristics of a Meaning-Text model follow directly from the postulates of the theory.
- An MTM is an equative, or transductive, rather than generative, model (Postulate 1).

- It is a completely formalized model (Postulate 2): representations of utterances and rules that manipulate them are written in formal languages.

- It is a stratificational model (Postulate 3): multiple levels of utterance representation are used and the rules are grouped into separate, self-contained components; this modular organization of the model makes the description of the mappings (between representation levels) less complex and allows for an easy modification/updating.

As already mentioned, an MTM presupposes seven levels of representation (of utterances) and consists of six sets of rules (= modules), which ensure subsequent transitions between the adjacent levels (Figure 1). Thus, the semantic module of an MTM maps a Semantic Representation [= SemR] to all corresponding, i.e., synonymous, Deep-Syntactic Representations [= DSyntR], the deep-syntactic module produces for a given DSyntR all corresponding Surface-Syntactic Representations [= SSyntR], and so forth.
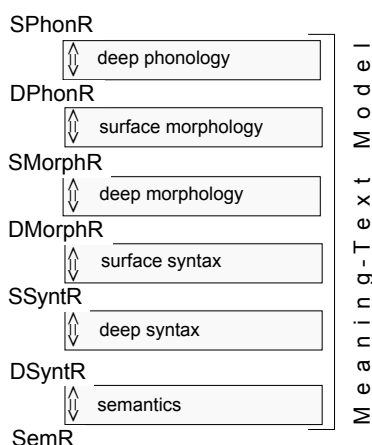

Figure 1: Architecture of an MTM

A representation (of an utterance) at a level *n* is a set of formal objects, called *structures*. Among these, a *central structure* is distinguished, which reflects the central linguistic entity of level *n*. At the semantic level, the central structure is an unordered *semantic network*, representing the propositional meaning of the utterance in terms of lexical meanings and predicate ~ argument relations between them; at the syntactic level it is an unordered *dependency tree*, representing the organization of the utterance in terms of lexical units and syntactic relations between them; at the morphological level, it is a *string of linearly ordered word-forms* which make up the utterance; and at the phonological level, it is a *string of phonemes*. Upon the central structure *'peripheral' structures* are superimposed, reflecting different characterizations of the central entity; in other words, they provide additional information—communicative, prosodic, etc.—relevant at the level *n*. Note that these structures are peripheral only in that they do not exist

independently of the central structure; as for their role in synthesis, it is by no means peripheral; thus, the Semantic-Communicative Structure is used to guide the entire process of synthesis, the Syntactic-Prosodic Structure is of crucial importance for the process of linearization, and so on.

Each representation level thus reflects a specific aspect of utterance organization, featuring linguistic objects/relations of different nature and, consequently, making use of different formalisms. This is one of the particularities of the MTT: the accent is not on unifying, but on distinguishing.

The rules of an MTM module are of two formal types: correspondence rules and equivalence rules:
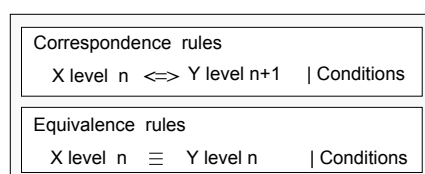
```
┌─────────────────────────────────────────────────┐
│ Correspondence  rules                           │
│    X level  n   <=>  Y level n+1     | Conditions │
├─────────────────────────────────────────────────┤
│ Equivalence  rules                              │
│    X level  n    ≡    Y level n      | Conditions │
└─────────────────────────────────────────────────┘
```

Figure 2: Two major types of rules of an MTM

Correspondence rules state correspondences between fragments of representations of two adjacent levels—e.g., lexicalization rules, operating between a SemR and a DSyntR; they represent the major part of an MTM.

Equivalence rules state equivalencies between representations of the same level—e.g., lexical-syntactic paraphrasing rules, operating between two DSyntRs.

From the substantial (= linguistic) viewpoint, these rules fall into several different types, some of which will be illustrated later.

Note that the name of an MTM module is derived from the deeper of the two representation levels between which it operates—this reflects the synthesis orientation of MTMs.

3. Linguistic Synthesis in the Meaning-Text framework

I will now illustrate the linguistic synthesis in an MTM by showing how three synonymous English sentences are synthesized, starting from their common SemR and working my way up to their DMorphRs.

At the same time, I will illustrate the way in which the conceptual and formal apparatus of the MTT is put to use in linguistic description; this is a point worth emphasizing, since the development of a coherent system of concepts and formal tools for linguistic description is one of the most important goals of the MTT.

3.1 From Semantic to Deep-Syntactic Representation: the Semantic Module

Within the MTT as such, semantic representations of utterances are considered as given and the problem of their construction is not dealt with. The MTT aims to describe only linguistic aspects of utterance production, leaving aside all its extralinguistic (conceptual, pragmatic, encyclopedic, etc.) aspects. Translating an informational content into a semantic representation involves logic, knowledge about the addressee/the situation and the extralinguistic world, and many other things. Thus, this task falls within the realm of a broader theory of human cognitive behavior and the corresponding model, which ensures the transition (Representation of the) World ⇔ (Representation of) Meaning.

Now, let me characterize the four structures that constitute the Semantic Representation, i.e., Semantic Structure, Semantic-Communicative Structure, Rhetorical Structure and Referential Structure:

$$\text{SemR} = <\text{SemS, Sem-CommS, RhetS, RefS}>.$$

The SemS, the central structure of a SemR, represents the propositional [= situational] meaning of a set of more or less synonymous sentences, i.e., paraphrases; it reflects the *paraphrasing power* of the language under consideration.

In order to discuss the SemS, I have to touch upon a seemingly simple, but in fact very difficult question: What is linguistic meaning? To grasp the meaning of an expression E, we have no other choice but to equate E with another expression E´. For instance, we will say that *know one's onions* [= E] means the same as *know one's job very well* [= E´], that *call the tune* [= E] has the same meaning as *be able to impose one's will on the person involved* [= E´], and so on. Thus, meaning turns out to be the invariant of paraphrases—the only thing that is common to all the expressions that have the same meaning. This characterization of meaning is by no means circular, since 'having the same meaning' is a simpler concept than that of 'having meaning;' in point of fact, 'having the same meaning' is a primitive (intuitive) notion, underlying all our lexical knowledge: for a native speaker it is by far easier to say whether E has the same meaning as E´ (= whether it is a paraphrase of E´) than to produce a description of the meaning of E.

In order to consider E and E´ as paraphrases, it is not necessary that their meanings be absolutely the same. Absolute synonymy is extremely rare and semantic differences between closely related expressions can be found. But, in 'ordinary' communication, these differences are by and large ignored by the speakers as irrelevant. Thus, for most speakers, it is enough that E and E´ be approximately synonymous to treat them as paraphrases.

Linguistic meaning is rather 'shallow:' it is the meaning accessible only through the speaker's knowledge of the language, without recourse to his/her extralinguistic and pragmatic knowledge or logical capabilities.

Formally, a SemS is a network—a graph whose nodes are labeled with *semantemes* [= lexical meanings of a language *L*] and whose arcs are labeled with distinctive numbers (1-6) that indicate predicate ~ argument relations linking different *arguments*, or *semantic actants*, of a predicate to this predicate. For example, the meaning of the verb *decide* is represented as a two-place predicate 'X decides that Y [should take place]', the meaning of *criticize* as a three-place predicate 'X criticizes Y for Z', and so on. For a SemS to be well formed, all argument slots of all predicate meanings it contains must be saturated (or explicitly marked as not saturated).

A semanteme is either a non-elementary meaning, i.e., such that it can be described in terms of simpler meanings, or an elementary meaning (= *seme*, *semantic primitive*), which cannot be described in such a way. The majority of meanings of *L* are of the first type. A description of a non-elementary lexical meaning in terms of simpler meanings (= its semantic decomposition) corresponds in fact to a lexicographic definition of a lexical unit having this semanteme as its signified. More on this will be said later, when the lexicon is discussed.

The first two peripheral structures, Sem-CommS and RhetS, represent, respectively, the communicative intent of the Speaker and his/her rhetorical (≈ stylistic) intent. Their purpose is to articulate the SSem into a specific *message*, by specifying the way it will be 'packaged' for communication.

The Sem-CommS specifies which part of the starting SemS will be the Theme (what is being talked about) and which the Rheme (what is communicated about the Theme), what will be presented as New and what as Given information, what will be Presupposed and what Asserted, etc. As for the RhetS, it specifies the style of the expression of the starting SemS, e.g., neutral, official, colloquial; ironic, poetic; etc.

The more the communicative/rhetoric parameters of the starting SemS are specified, the fewer is the number of (synonymous) sentences that can be synthesized out of it. Thus, by narrowing down the possibilities of expression of the starting propositional meaning, possibly to a single sentence, which would best fit into the context, the two peripheral structures reduce the 'paraphrastic power' of the starting SemS.

Note that from this perspective speaking can be viewed as 'virtual paraphrasing,' i.e., as a series of choices between equivalent ways of expressing a given meaning.

Formally, the Sem-CommS is a division of the SemS into *communicative areas* [= subnetworks], each marked with one of mutually exclusive values of *eight communicative oppositions* (Mel'čuk, 2001): Thematicity = {Theme, Rheme, Specifier}, Giveness = {Given, New}, Focalization = {Focalized, Non-Focalized}, Perspective = {Backgrounded, Foregrounded, Neutral}, Emphasis = {Emphasized, Neutral}, Assertiveness = {Asserted, Presupposed}, Unitariness = {Unitary, Articulated} and Locutionality = {Communicated, Signaled, Performed}.

In each subnetwork, one node is singled out as *communicatively dominant*: this is the node to which the entire subnetwork can be semantically reduced, i.e., the node that can function as a 'minimal paraphrase' of the subnetwork. Thus, the subnetworks 'criticize'—2→'government' and 'criticize'—2→'government', where the dominant nodes are underlined, can be reduced, respectively, to 'criticize' (*criticism* [*of the government*]) and 'government' (*the Government* [*which undergoes criticism*]).

Specification of communicatively dominant nodes (of subnetworks) is crucial for determining the syntactic structure of the sentence: it is the dominant node of the Rheme (or, less often, that of the Theme or of a Specifier) that will give the top node of the deep-syntactic tree (cf. p. 12).

The RhetS consists of a specification of stylistic labels, where appropriate.

The last peripheral structure, the RefS, consists of a set of pointers from semantic configurations to the corresponding entities in the real world; it indicates referents of corresponding meanings.

In what follows, RhetS and RefS will not be considered.

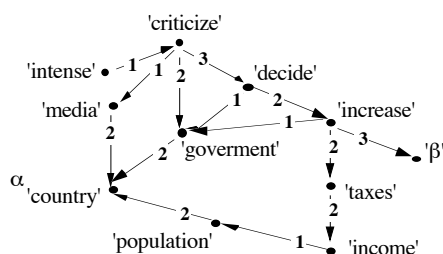Let there be the following SemS (henceforth *sample SemS*):

Figure 3: A Semantic Structure

This SemS represents two interwoven situations:

1) situation 'decision', which can be literally verbalized as 'the government [of the country α] decided to increase [by amount β] taxes on the income [of the population of α]', and

2) situation 'criticizing' (into which the first situation is, so to speak, embedded): 'the media [of α] criticized intensely the government [of α] for (having decided...)'.

Note that the meanings 'country' and 'population' are redundant and need not be explicitly expressed; cf. sentences in (1) and (2) below. The amount of the tax increase is not specified.

The inflectional meanings—in the case of English, verbal voice/mode/tense and nominal number/definiteness—should also be represented in a SemS; however, they have been omitted from the sample SemS for simplicity's sake.

The propositional meaning represented by the sample SemS can be communicatively organized in many different ways; let us consider just two of those: two SemRs obtained from the sample SemS by assigning it two different Sem-CommSs, and some of their respective realizations. In both cases, only the Theme ~ Rheme communicative opposition is used. (Dominant nodes are underlined.)
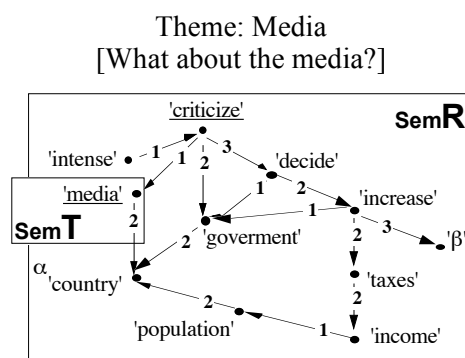
Theme: Media
[What about the media?]

Figure 4: SemR [1], underlying sentences in (1)

(1) a. [*The media*]**T** [*harshly criticized the Government for its decision to increase income taxes*]**R**.
   b. [*The media*]**T** [*seriously criticized the Government's decision to raise income taxes*]**R**.
   c. [*The media*]**T** [*leveled harsh criticism at the Government for its decision to increase income taxes*]**R**.

Theme: Government's decision to raise taxes
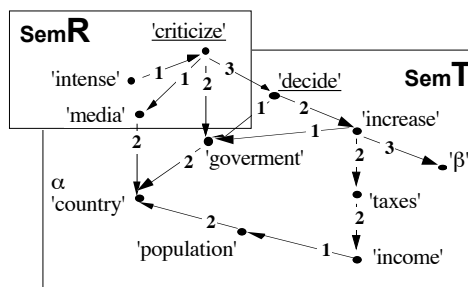[What about the Governments' decision?]

Figure 5: SemR [2], underlying sentences in (2)

(2) a. [*The Government's decision to increase income taxes*]**T** [*was severely criticized by the media*]**R**.

> b. [*The Government's decision to raise income taxes*]**T** [*drew harsh criticism from the media*]**R**.
>
> c. [*The Government's decision to increase income taxes*]**T** [*came under harsh criticism from the media*]**R**.

Let me point out two major differences between the two sets of realizations, which stem directly from their having different underlying Sem-CommSs.

First, in (1), 'criticize' is realized both as a three-actantial and a two-actantial expression: in (1a) and (1c), we have, respectively, *X criticizes Y for Z* and *X levels criticism at Y for Z*, while in (1b) the realization is *X criticizes Z's Y*, i.e., *Z's Y* (= *Government's decision*) is taken as one actant of 'criticize'. In (2), only the two-actantial realization of 'criticize' is possible.

Second, in (1), the dominant node of the Theme, which is the semantic actant (= SemA) **1** of 'criticize', is realized as the deep-syntactic actant (= DSyntA) **I** of the verb (or of the light verb construction) which corresponds to 'criticize', so that the resulting sentences are communicatively neutral. In (2), the dominant node of the Theme is the SemA **3** of 'criticize' and it is realized as the DSyntA **I** of the verb (or, again, of the light verb construction) corresponding to 'criticize'; to achieve this, the conversion is required—in (2a), conversion is realized by passivization of the main verb, and in (2b/c) by the choice of an appropriate light verb—so that the resulting sentences are communicatively marked with respect to those in (1).

The remaining differences between the two sets of realizations as well as between the sentences within each set (e.g., different lexical expressions of intensification) are more superficial, i.e., they are brought to bear in subsequent stages of synthesis, and will be described in due course.

Let us now see how sentences are represented at the Deep-Syntactic level. But before I start talking about deep syntax, I would like to characterize those aspects of syntactic representations that are common for both the deep- and the surface-syntactic levels.

Unlike most mainstream approaches, which use different variants of phrase-structure formalism to represent syntactic structures, MTT uses only *dependencies*: the syntactic structure of a sentence, at both deep and surface levels, is represented in terms of a set of *binary dependency relations* defined over the set of lexical units L_i making up the sentence.

Informally, we can say that (in the syntactic structure of the sentence) the lexical unit L_2 syntactically depends on the lexical unit L_1 (L_1—synt$\rightarrow L_2$) if the linear positioning of L_2 is determined with respect to L_1. In other words, if L_2 depends on L_1, the rule for the linear positioning of L_2 has to mention L_1, while the inverse is not true.

Syntactic dependency, in our case, an immediate (or direct) dependency, is a relation of strict hierarchy; it is *antireflexive* (L cannot depend on L, which means that L cannot be linearly positioned or inflected with respect to itself), *antisymmetric* (if L_1 depends on L_2, then L_2 cannot depend on L_1) and *antitransitive* (if L_1 depends on L_2 and L_2 depends on L_3, then L_1 cannot depend directly on L_3; this is actually the condition of the unicity of the Synt-governor and of the presence of one and only one top node top node of the dependency tree, see below).

The syntactic structure of a sentence appears as a *tree*: a connected two-dimensional graph, with lexical units labeling the nodes and dependency relations labeling the arcs (= branches).

The nodes of a syntactic dependency tree are not linearly ordered: since word order is a means the language uses to express syntactic relations (e.g., in an English declarative sentence the subject of the Main Verb is positioned to the left of the Main Verb, etc.), it cannot be present in the syntactic structure.

One node of the tree is singled out as the *top node* (= absolute head, or root): it does not depend on any lexical unit in the syntactic structure, while all other units depend on it, directly or indirectly. In the so-called Standard Average European languages, the top node of a dependency tree corresponding to a sentence is a finite verb.

Syntactic structures at both deep and surface levels consist of the same formal elements, i.e., they are unordered rooted dependency trees, but they feature different substantive (= linguistic) elements, i.e., different types of lexical units and syntactic relations. (The same dichotomy is valid for all deep vs. surface representation levels.)

A Deep-Syntactic Representation is made up of four structures: Deep-Syntactic Structure, Deep-Syntactic-Communicative Structure, Deep-Syntactic-Prosodic Structure, and Deep-Syntactic-Anaphoric Structure:

DSyntR = <DSyntS, DSynt-CommS, DSynt-ProsS, DSynt-AnaphS>.

The nodes of the DSyntS, the central structure of the DSyntR, are labeled *with deep lexical units* subscripted for all *meaning-bearing inflections* and its branches are labeled with names of *deep-syntactic dependency relations*.

Only a deep, or semantically full, lexical unit, i.e., a unit that corresponds directly to a (configuration of) meaning(s) in the SemS, can appear as a label on a node of a DSyntS; this means that substitute pronouns and 'structural words'—auxiliaries, governed prepositions, dummy subject pronouns, and the like—do not appear at this level of representation.

A deep lexical unit can be a *lexeme* (= a word taken in one specific sense), a *(full) phraseme* (= an idiom), e.g., RED HERRING, EAT HUMBLE PIE, etc., or the name of a *lexical function*.

Lexical functions [= LF] represent one of the most important discoveries of the Meaning-Text linguistics. They are formal tools used to model *lexical relations*, i.e., *restricted lexical co-occurrence* (= collocations) and *semantic 'derivation*.' I cannot offer here a serious presentation of LFs and will limit myself to a minimum of explanations.

An LF corresponds to a meaning whose expression is phraseologically bound by a particular lexeme L (= argument, or keyword, of the LF). For instance, the LFs Magn and S_1, which correspond, respectively, to the meanings 'intense/very' and 'person/object doing [what is denoted by] L', have different lexical expressions (= elements of value) contingent on the keyword:

Magn(*wind*) = *strong, powerful*

Magn(*rain*$_{(N)}$) = *heavy, torrential // downpour* [the symbol " // " identifies a fused value of an LF—a value expressing together the meaning of the keyword and that of the FL]

Magn(*rain*$_{(V)}$) = *heavily, cats and dogs*

S_1(*crime*) = *author, perpetrator* [of ART ~] // *criminal*

S_1(*kill*) = *killer*

Lexical functions can be classified along three axes.

1) According to their capacity to appear in the text alongside the keyword, LFs fall into *syntagmatic* functions (normally appear with the key word) and *paradigmatic* functions (normally do not appear with the key word). Roughly speaking, syntagmatic functions of L correspond to characterizers of L, while paradigmatic functions correspond to derivatives of L (in broad sense, including synonyms and antonyms). The LF Magn illustrates the former type, and the LF S_1—the latter.

2) According to their generality/universality, LFs can be *standard* (general and universal) *or non standard* (neither general nor universal).

A standard FL is general in that it is applicable to a high number of keywords and has many elements of its possible values. It is universal in that it is valid cross-linguistically. Magn and S_1 are such LFs. Standard LFs are used to formulate lexical-syntactic paraphrasing rules (see p. 21).

A non-standard LF is applicable to few keywords, possibly only one, and has few expressions; cf. the following non-standard LF, which applies only to the lexeme YEAR and has just one value:

a YEAR that has 366 days = *leap* [~].

3) According to their formal structure, LFs are divided into *simple* functions (such as, again, Magn and S_1) and *complex* functions, a complex LF being a combination of syntactically related simple LFs (cf. the LF CausPredPlus below).

Here are the LFs realized in sentences in (1) and (2):

Magn(*criticize*) = *bitterly*, *harshly, seriously, severely, strongly* // *blast*

Magn(*criticism*) = *bitter*, *harsh, serious, severe, strong*

[Verb meaning 'cause to become more']

CausPredPlus(*taxes*) = *increase, raise*

[Action noun for the situation described by the keyword; Q stands for 'quasi', since *criticism* is not a 'pure' action noun of *criticize*; in the representations below, I ignore this subtlety]

QS_0(*criticize*) = *criticism*

QS_0(*decide*) = *decision*

[Light (= semantically empty) verbs linking the keyword to its DSyntAs **I** and **II**]

Oper$_1$(*criticism*) = *level* [~ at N | N denotes a person], *raise* [~ against N], *voice* [~]

Oper$_2$(*criticism*) = *come* [under ~], *draw* [~ from N], *meet* [with~]

Deep lexical units (in the deep-syntactic structure of the sentence) do not correspond one-to-one to the surface lexemes (of the actual sentence): in the transition towards surface syntax, some deep lexical units may get deleted or pronominalized, and some surface lexemes may be added.

Twelve Deep-Syntactic Relations [= DSyntRel] are distinguished:

- Six actantial DSyntRels (**I**, **II**, **III**, ..., **VI**) and one DSyntRel for representing direct speech, which is a 'variant' of the DSyntRel **II**.

- Two Attributive DSyntRels: **ATTR**$_{\text{restr(ictive)}}$ and **ATTR**$_{\text{qual(ificative)}}$.

- One Appenditive DSyntRel (**APPEND**), which links the Main Verb to 'extra-structural' sentence elements, such as sentential adverbs, interjections, addresses, etc.

- Two coordinative DSyntRels: **COORD** and **QUASI-COORD**, the latter being used in special constructions of type *He came on November 2nd*—**QUASI-COORD**→in *the evening*—**QUASI-COORD**→*at 8 o'clock*, where each following conjunct is an 'elaboration' of the preceding one.

Each relation represents a family of syntactic constructions of particular languages. Thus, DSyntRel **I** represents predicative constructions (with transitive verbs in the active voice), DSyntRel **II** represents main object constructions, the two attributive DSyntRels represent modifier constructions, etc. DSyntRels are deemed to be universal, i.e., sufficient for describing the deep syntax of any language.

The DSynt-CommS is the syntactic counterpart of the Sem-CommS. It is also formed of markers of communicative oppositions, such as Thematicity, etc., attached to particular DSynt-subtrees. (Of course, these are different sets of oppositions with respect to those appearing in the SemS, even though they may share the same names.) DSynt-CommS provides the information necessary to control linearization and prosodization in the subsequent stages of synthesis.

The DSynt-ProsS represents the semantic prosody of the sentence; it consists of a set of markers of meaning-bearing prosodies: declarative vs. interrogative; neutral vs. ironic vs. indignant; and so on.

The DSynt-AnaphS represents the links of co-referentiality between nodes of the DSyntS: a pair of co-referential nodes is linked by a dashed double-headed arrow; it provides the information necessary for performing ellipsis and pronominalization in the subsequent stages of synthesis.

The following DSyntSs correspond to SemR [1], underlying sentences in (1); inflectional subscripts are not shown, except for the voice of the main verb.
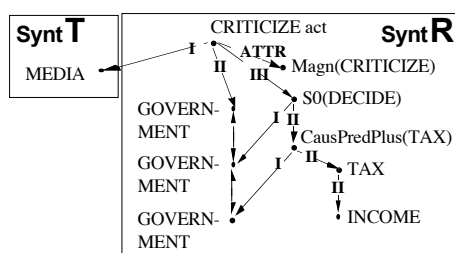
[**Prosody: declarative, neutral**]

Figure 6: DSyntR of sentence (1a)

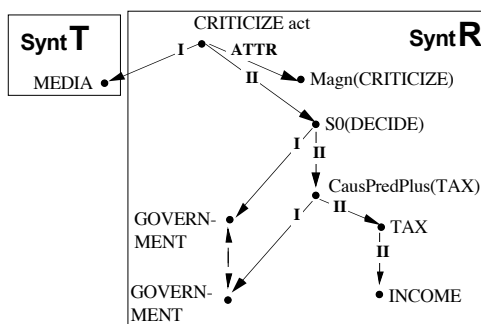[**Prosody: declarative, neutral**]

Figure 7: DSyntR of sentence (1b)
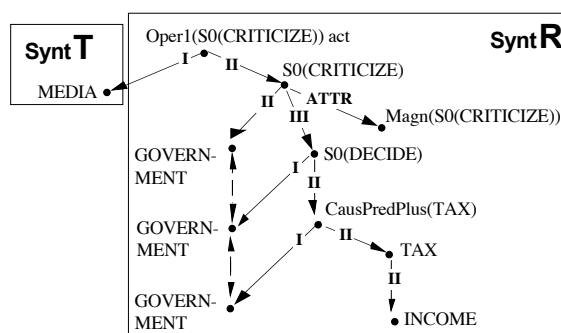
[Prosody: declarative, neutral]

Figure 8: DSyntR of sentence (1c)

In the three DSyntRs above, MEDIA is the syntactic Theme and all other lexical units are in the syntactic Rheme; the top node—CRITICIZE in (1a-b) and the LF $Oper_1(S_0(CRITICIZE))$ in (1c)—realizes the dominant node of the Rheme of the corresponding SemR.

Let us turn now to the rules of the Semantic Module, mapping SemRs to their corresponding DSyntRs: first to the correspondence rules, and then to the equivalence rules.

Semantic correspondence rules are of the following major types:

1) *Lexicalization rules*: they map (configurations of) semantemes to the corresponding deep-lexical units; these rules are further divided into lexemic, phrasemic and lexical-functional rules, depending on what their right-hand side is (for illustrations, see below).

2) *Morphologization rules*: they map (configurations of) semantemes to the corresponding full grammatical (= inflectional and derivational) values, such as nominal number, verbal mood/tense, etc., attached to the deep-lexical units.

3) *Arborization rules*: they construct the deep-syntactic tree, based on semantic/semantic-communicative dependencies between semantemes and the lexicographic properties of the corresponding lexical units (for illustrations, see below).

4) *Communicative rules*: they construct the Communicative Structure at the deep-syntactic level of representation.

5) *Prosodic rules*: they construct the Prosodic Structure at the deep-syntactic level of representation.

Since the mapping between SemRs and DSyntRs crucially relies on the information in the lexicon, I have to say a few words about the lexicon presupposed by the MTT—the *Explanatory Combinatorial Dictionary*, or ECD. (A heavy reliance on the lexicon is one of the distinguishing features of the MTT; for more on this, see Section 4.) I start with a general characterization of an

ECD, followed by a description of the structure of an ECD entry; the illustrations use the data from the entry for the lexeme CRITICIZE.

An ECD has the following main properties.

1) It is a theoretical lexicon, i.e., a lexicon anchored in a linguistic theory and used primarily as a research tool in linguistics. The elaboration of an ECD is guided by concerns of logic/coherence and is based on rigorous principles/methods of lexicographic description (see, for instance, Mel'čuk *et al*., 1995: 72-111).

2) It is semantics-based and production-oriented (in accordance with the MTT). This means that the definition of a lexical unit L provides the basis for the specification of all other features of L (hence the qualifier *explanatory*), and that the description of L supplies all the information necessary to use L correctly—in order to express a given meaning.

3) It is *combinatorial*: it provides a detailed description of the syntactic and lexical co-occurrence of L.

4) It is systematic: the elaboration of an ECD is carried out by semantic fields (rather than by alphabetical order), so that semantically related lexical units have similar descriptions; all links existing between the definition and the restricted co-occurrence of L are made explicit, etc.

5) It is exhaustive in the following sense: the description of a given L contains all information necessary to use L in all possible contexts.

6) It is formalized: it uses sophisticated metalanguages (language of definitions, lexical functions, etc.) allowing for rigorous lexicographic descriptions.

An ECD entry, which corresponds to a single lexeme or a single phraseme, has a standardized structure; it comprises three major zones: *semantic zone* (≈ definition of L), *syntactic co-occurrence zone* (≈ Government Pattern of L) and *lexical relations zone* (lexical functions of L).

1) Semantic zone

An ECD definition is made up of two parts:

- *definiendum*, which (in the case of Ls corresponding to predicative meanings) is a propositional form featuring L and its semantic actants, represented by variables ('X', 'Y', 'Z', ...);

- *definiens*, or the definition proper, which is the semantic decomposition of L—its paraphrase in terms of simpler constitutive meanings. (A meaning 'm_1' is simpler than a meaning 'm_2' if 'm_1' is used to define 'm_2', but the inverse is not possible; for instance, the meaning 'express

an opinion' is simpler than the meaning 'criticize', since 'criticize' ≈ 'express a particular kind of opinion in a particular way', but 'express an opinion' ≠ 'criticize in a particular way'.)

The definition of the lexeme CRITICIZE follows; the double bar separates the presuppositional part of the definiens (leftmost) from the assertorial part; the distinctive numbers identifying word-senses are taken from the *Longman Dictionary of Contemporary English*:

(definiendum)   (definiens)

'X criticizes Y for Z' ≈ 'Y having done$^2_1$ Z which X considers$_2$ bad$_2$ for Y or other people$_1$,
and X believing$_3$ that X has good$^1_1$ reasons$^1_2$ for considering$_2$ Z bad$_2$, ||
X expresses$^3_1$ X's negative$^1_1$ opinion$_1$ of Y because of Z(Y),
specifying what X considers$_2$ bad$_2$ about Z,
with the intention$_2$ to cause^2 that people$_1$ (including Y) do not do$^2_1$ Z.'

The definiens and the definiendum are fully synonymous and must be substitutable in all contexts, with the preservation of semantic content (even if some stylistic rules are violated). In other words, each component of the definiens must be necessary and all its components sufficient to specify the entire range of use of L (adequacy principle).

Defining lexical units by semantic decomposition into simpler meanings allows us to avoid vicious circles, a problem which plagues most of the existing dictionaries. (To my knowledge, the only other lexicographic approach which consistently uses the principle of semantic decomposition is that of A. Wierzbicka, who was actually the first researcher to introduce it, some 40 year ago.) Ultimately, this technique leads to the identification of semantic primitives, i.e., 'atomic' meanings of a language.

Definitions are written in a natural language (for an English ECD, it is English, for a French ECD, it is French, etc.) that has been 'treated' is such a way as to be free of lexical ambiguity and synonymy. This means that all lexical meanings used in definitions are disambiguated by means of distinctive numbers (cf. the above definition), and that among synonymous lexical meanings (e.g., 'consider X as Y', 'regard X as Y', 'find X [to be] Y') only one is admitted in definitions. A definition written in such a language is equivalent to a semantic network.

2) Syntactic co-occurrence zone

This zone comprises the *Government Pattern* (≈ subcategorization frame) of L, which specifies:

- the diathesis of L, i.e., the correspondence between L's SemAs and its DSyntAs;
- all surface expressions that implement the DSyntAs of L.

The lexeme CRITICIZE has a rather complex Government Pattern [= GP]: its SemAs can be mapped onto DSyntAs not in just one way (which is the prototypical case), but in two different

ways; this is what I referred to earlier as a three- or two-actantial realization of the predicate 'criticize'. Thus, the GP of this lexeme has two modifications, which amounts to saying that there are two different GPs:

	GP 1			**GP 2**	
Diathesis	X = I	Y = II	Z = III	X = I	Z = II
Surface expressions of DSyntAs	1. N	1. N	1. *for* N 2. *for* V$_{ger}$	1. N	1. N
		oblig			**oblig**

Figure 9: Government Patterns of CRITICIZE

GP 1

The SemAs X, Y and Z of CRITICIZE correspond, respectively, to its DSyntAs **I**, **II** and **III**. The DSyntAs **I** and **II** are expressed on the surface as prepositionless nouns, and the DSyntA **III** can be expressed as a noun/a gerund introduced by the preposition FOR. E.g.: *John* [= **I**] *criticized Mary* [= **II**] *for her rude behavior* ⟨*for having behaved rudely*⟩ [= **III**].

GP 2

The SemA X of CRITICIZE corresponds to its DSyntA **I**, and its SemA Z corresponds to its DSyntA **II**. Both DSyntAs are expressed on the surface as prepositionless nouns. E.g., *John* [= **I**] *criticized Mary's rude behavior* [= **II**].

GP 1 of CRITICIZE is used to construct the DSyntR of the sentence (1a), while that of the sentence (1b) is constructed by using GP 2.

3) Lexical relation zone

This zone contains the data on 'semantic derivation' and restricted lexical co-occurrence of L, described by means of lexical functions. For each L, LFs specify its paradigmatic lexical correlates—roughly, (quasi-)synonyms and (quasi)derivatives of L, and its syntagmatic lexical correlates—lexical units which form collocations with L; cf. some lexical relations of CRITICIZE:

QSyn: *attack, disapprove, reproach*
QAnti: *praise, congratulate*
QS$_0$:   *criticism*
S$_1$:   *critic*
A$_1$:   *critical* [of N]
Magn: *bitterly, harshly, seriously, severely, strongly*
Magn^{quant}:   *all the time, relentlessly, without stopping*
AntiMagn: *half-heartedly, mildly*
AntiVer: *unjustly, without reason*

I will now cite three semantic correspondence rules used to construct the DSyntRs of the sentences (1a) and (1b) from the SemR [1]; as for the DSyntR of the sentence (1c), it is obtained from the DSyntR of the sentence (1a) by an equivalence (paraphrasing) rule, which will be given later.

Note that in our case all lexemic rules are trivial—the semanteme 'criticize' maps onto the lexeme CRITICIZE, etc.; that is why they will not be illustrated. Examples of more complex lexemic rules can be found, for instance, in Kahane & Mel'čuk (1999).

All the rules need conditions of application; however, they will not be indicated.
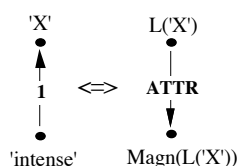

Figure 10: A lexical-functional rule

This lexical-functional rule maps the semanteme 'intense' to the LF Magn, which later will be implemented by one of lexically distributed elements of its value. (The other elements of the subnetwork/the subtree shown in the above rule constitute the context of the rule. The translation of the semantic arc **1** to the corresponding deep-syntactic branch **ATTR** is taken care of by Arborization rule 2.)
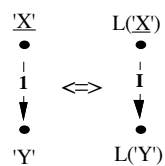

Figure 11: Arborization rule 1

Arborization rule 1 maps the SemRel **1** linking a communicatively dominant predicate '$\underline{X}$' with its argument 'Y' to the DSyntRel **I**; cf. *the government* [= L('Y')] *decided* [= L('X')]; *the Government's* [= L('Y')] *decision* [= L('X')]; *the decision* [= L('X')] *of the Government* [= L('Y')].
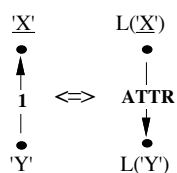

Figure 12: Arborization rule 2

Arborization rule 2 maps the SemRel **1** linking a predicate 'Y' with its communicatively dominant argument '<u>X</u>' to the DSyntRel **ATTR**; cf. *criticize* [= L('X')] *Magn (= harshly)* [= L('Y')]; *Magn (= harsh)* [= L('Y')] *criticism* [= L('X')].

Equivalence rules of the Semantic Module—actually, paraphrasing rules—are of the following two types:

1) semantic equivalence rules, which establish the equivalence between (fragments of) two SemRs (these rules will not be illustrated);

2) lexico-syntactic equivalence rules, formulated in terms of lexical functions, which establish the equivalence between (fragments of) two DSyntRs.

Lexical-syntactic equivalence rules allow us to construct, starting from a given DSyntR, equivalent DSyntRs, i.e., DSyntRs that express the same meaning, but feature different lexical and structural elements. One such rule is given in Figure 13:

$$\text{Oper}_1(S_0(L_{(V)}))$$
$$L_{(V)} \quad \equiv \qquad \begin{array}{c} \circ \\ \shortmid \\ \text{II} \\ \downarrow \\ \circ \\ S_0(L_{(V)}) \end{array}$$

Figure 13: A Deep-Syntactic Equivalence Rule

By applying this equivalence rule to the DSyntR of (1a), we get an equivalent DSyntR—that of (1c): *The media harshly **criticized*** [= L] ... ≡ *The media **leveled*** [= Oper_1] *harsh **criticism*** [= $(S_0(L))$] ...

Although the DSyntR of (1c) can be produced directly from the SemR [1] by correspondence rules, it is easier to get it from a lexically/structurally simpler DSyntR of (1c) by the above equivalence rule.

There are about 50 rules of this type, valid cross-linguistically; they constitute the *paraphrasing system* of an MTM, see Žolkovskij & Mel'čuk, 1967; Mel'čuk, 1992; Milićević, in press.

3.2 From Deep- to Surface-Syntactic Representation: the Deep-Syntactic Module

A Surface-Syntactic Representation is a set of four structures: Surface-Syntactic Structure, Surface-Syntactic-Communicative Structure, Surface-Syntactic Prosodic Structure and Surface-Syntactic Anaphoric Structure:

SSyntR = <SSyntS, SSynt-CommS, SSynt-ProsS, SSynt-AnaphS>.

The central structure, SSyntS, is a dependency tree whose nodes are labeled with actual lexemes of the sentence (with the same morphological subscripts as in the DSyntS) and whose branches are labeled with names of language-specific surface-syntactic dependency relations. Thus, a deep- and a surface-syntactic structure differ in the following two respects: 1) lexically, the former features only semantically full lexemes, while the latter contains all lexemes which will be present in the output sentence (both full and structural words, as well as substitute pronouns); 2) syntactically, a DSyntS uses only universal dependency relations, whereas in a SSyntS these are specific dependency relations, which have to be established empirically for each language. For surface-syntactic relations in French, see Mel'čuk & Iordanskaja in Polguère, ed. (to appear); English surface-syntactic relations are described in Mel'čuk & Pertsov (1987) and Mel'čuk in Polguère, ed. (to appear); an inventory of surface-syntactic relations for Russian can be found in Apresjan *et al.* (1989) and (1992).

The three peripheral structures have similar organization and fulfill similar roles as their counterparts at the deep-syntactic level.

Sentences in (1) have the following SSyntRs (again, the inflectional subscripts are not indicated, except for the voice of the main verb).
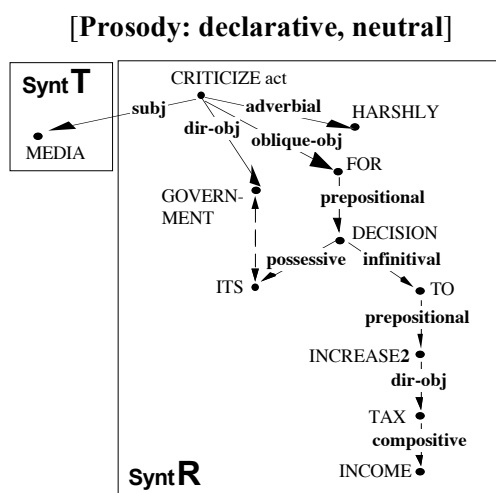
[**Prosody: declarative, neutral**]

Figure 14: SSyntR of sentence (1a)
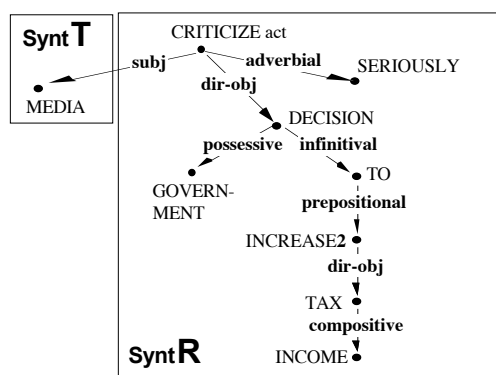
[Prosody: declarative, neutral]

Figure 15: SSyntR of sentence (1b)

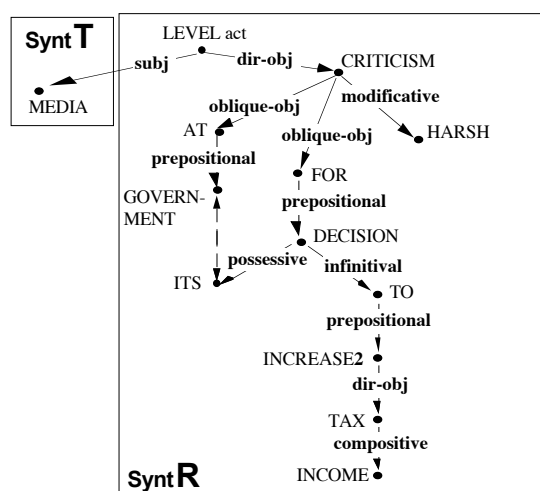**[Prosody: declarative, neutral]**

Figure 16: SSyntR of sentence (1c)

The rules of the Deep-Syntactic Module, which map DSyntRs onto their corresponding SSyntRs, are of the following major types:

1)   Phrasemic rules 'expand' DSynt-nodes labeled with full phrasemes into corresponding SSynt-subtrees and compute values of LFs present in the DSyntR.

2)   Deep-Syntactic rules proper construct the surface-syntactic tree, based on deep-syntactic dependencies and lexicographic properties of lexical units (among other things, they introduce structural words).

3)   Pronominalization rules introduce substitute pronouns replacing some of the co-referential expressions in the DSynt-tree, according to the general rules of the language.

4)   Ellipsis rules perform all sorts of ellipsis, e.g., equi-deletion and conjunction reduction, again according to the general rules of the language.

5)   Communicative rules construct the Communicative Structure at the surface-syntactic level of representation.

6)   Prosodic rules construct Prosodic Structures at the surface-syntactic level of representation.

Here are six phrasemic (more precisely, lexical-functional) rules that compute the values of LFs for the SyntSs (1a-c):

SSyntS (1a)

1) Magn(CRITICIZE) <=> *harshly*; 2) CausPredPlus(TAXES) <=> *increase*

SSyntS (1b)

3) Magn(CRITICIZE) <=> *seriously*; 4) CausPredPlus(TAXES) <=> *raise*

SSyntS (1c)

5) Oper$_1$(CRITICISM) <=> *level*; 6) Magn(CRITICISM) <=> *harsh*

In all these cases, the choice among possible elements of the value of a LF is unconstrained, i.e., in each SSyntS, each LF could have been realized by selecting any other element of its value (cf. p. 19). This does not mean, however, that the selection of a value for a LF is always trivial— it can be constrained by syntactic/communicative, stylistic, and perhaps other factors. For instance, the choice of the value *level* for Oper$_1$(CRITICISM) is excluded if the third actant of *level* does not denote a person, cf. (3a) vs. (3b). Similarly, if two LFs in the same SSyntR have identical realizations, the resulting sentence is stylistically unacceptable, cf. (4).

(3) a. *The media raised harsh criticism against the Government for its decision to impose higher taxes.*
   or
   The media leveled harsh criticism at the Government for its decision to impose higher taxes.
   b. *The media raised harsh criticism against the Government's decision to impose higher taxes.*
   vs.
   *T*he media leveled harsh criticism at the Government's decision to impose higher taxes.*

(4) $^?$*The media raised harsh criticism against the Government for its decision to raise taxes.*

Two deep-syntactic rules used to construct the SSyntS of (1a) and (1b) are presented below.
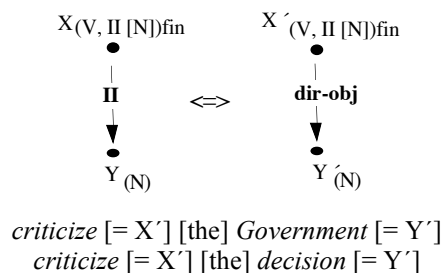
$$X_{(V, \text{II [N])fin}} \qquad X'_{(V, \text{II [N])fin}}$$

$$\text{II} \qquad \Longleftrightarrow \qquad \text{dir-obj}$$

$$Y_{(N)} \qquad\qquad Y'_{(N)}$$

criticize [= X´] [the] *Government* [= Y´]
criticize [= X´] [the] *decision* [= Y´]

Figure 17: DSynt-rule 1

DSynt-rule 1 maps the DSyntRel **II** to the **direct-objectival** SSyntRel, according to the Government Pattern of the governing member of the relation (= $X_{(V, \text{II(N))fin}}$; this notation means 'finite verb whose DSyntA **II** is realized in the SSyntS as a prepositionless noun').

$$X'_{(V, \text{III[Prep])}}$$

$$X_{(V, \text{III[Prep])}} \qquad \text{oblique-obj}$$

$$\text{III} \qquad \Longleftrightarrow \qquad Z'_{(Prep)}$$

$$\text{prepositional}$$

$$Y_{\alpha} \qquad\qquad Y'_{\alpha}$$

$$\alpha = N, V_{\text{ger}}$$

criticize [= X´] *for* [= Z´] [*the*] *decision* [= Y´]
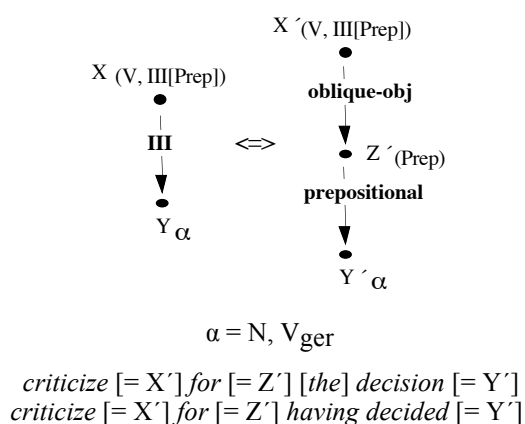*criticize* [= X´] *for* [= Z´] *having decided* [= Y´]

Figure 18: DSynt-rule 2

DSynt-rule 2 maps, again according to the Government Pattern of the governing member (= $X_{(V, \text{III[Prep])}}$), the DSyntRel **III** to this particular SSynt-configuration; more precisely, it introduces a preposition with two corresponding branches to express the DSyntA **III** of the verb that controls, in the SSyntS, a prepositional object.

After the above illustrations, the dichotomy deep- vs. surface-levels in syntax should be clearer to the reader. In the transition Sem=>DSynt, we make the syntactic choices relative to the content of the utterance we want to produce (e.g., the choice of an intensifying modifier or of an appropriate actantial relation), while the syntactic choices made in the transition DSynt=>SSynt concern the form in which this content is expressed (e.g., the choice between *criticize* **harshly** and *criticize* **severely** or between *criticize* [...] **for the decision** and *criticize* [...] **for having decided**).

3.3 From Surface-Syntactic to Deep-Morphological Representation: the Surface-Syntactic Module

The Deep-Morphological Representation (of the sentence) consists of Deep-Morphological Structure and Deep-Morphological Prosodic Structure:

DMorphR = <DMorphS, DMorph-ProsS>.

The DMorphS is a string of fully ordered lexemes subscripted with all inflectional values (= both semantic and syntactically induced inflections). The DMorph-ProsS consists of a specification of semantically induced prosodies (carried over from syntactic representation levels), and of syntactically induced ones—pauses, breath-groups and contours.

Sentences in (1) have the following DMorphRs (the only element of the DMorph-ProsS indicated in the representations below are the pauses: '|' stands for a minor pause, '||' for a longer one, and '|||' indicates a sentence-final pause):

Sentence (1a)

THE MEDIA$_{pl}$ | HARSHLY CRITICIZE$_{act, ind, past, 3sg}$ THE GOVERNMENT$_{sg}$ ||
FOR ITS$_{sg}$ DECISION$_{sg}$ | TO INCREASE$_{inf}$ INCOME$_{sg}$ TAX$_{pl}$ |||

Sentence (1b)

THE MEDIA$_{pl}$ || SERIOUSLY CRITICIZE$_{act, ind, past, 3sg}$
THE GOVERNMENT$_{sg, possessive}$ DECISION$_{sg}$ | TO RAISE$_{inf}$ INCOME$_{sg}$ TAX$_{pl}$ |||

Sentence (1c)

THE MEDIA$_{pl}$ | LEVEL$_{act, ind, past, 3sg}$
HARSH CRITICISM$_{sg}$ AT THE GOVERNMENT$_{sg}$ || FOR ITS DECISION$_{sg}$ |
TO INCREASE$_{inf}$ INCOME$_{sg}$ TAX$_{pl}$ |||

Let me now indicate the major types of rules of the SSynt-module:

1) Linearization rules determine the actual linear order of words in the sentence, based on syntactic, communicative and prosodic information in the SSyntR (for examples of such linearization rules, see Mel'čuk, 1967).

Local (and semi-local) linearization rules build elementary phrases, cf. (5a), and combine them into complex phrases, cf. (5b). Global linearization rules determine the order of complex phrases within the clause and the order of clauses within the sentence.

(5) a. [*the Government's*]$_{elementary.ph.}$[*decision*]$_{elementary.ph.}$
   [*to increase*]$_{elementary.ph.}$ [*taxes*]$_{elementary.ph.}$
  b. [[*the Government's decision*]$_{complex ph.}$ [*to increase taxes*]$_{complex ph.}$]$_{complex ph.}$

2) Morphologization rules compute, for each lexeme, its syntactically induced morphological values (= inflections imposed by agreement and government, such as verbal

number/person, nominal case, etc.), adding them to semantically full inflections, already computed by morphologization rules of the Semantic module.

3) Prosodization rules compute syntactically induced prosodies (breath groups, pauses and contours).

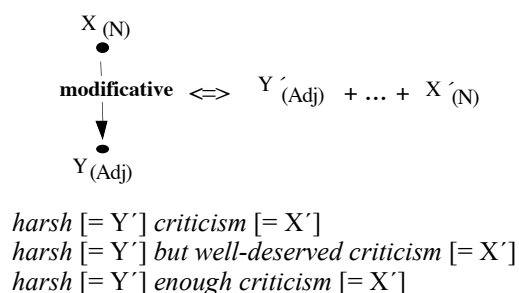Here is a local linearization rule used to construct a fragment of the DMorphR of sentence (1c).

$$X_{(N)}$$

modificative $\Longleftrightarrow$ $Y'_{(Adj)}$ + ... + $X'_{(N)}$

$$Y_{(Adj)}$$

harsh [= Y´] *criticism* [= X´]
harsh [= Y´] *but well-deserved criticism* [= X´]
harsh [= Y´] *enough criticism* [= X´]

Figure 19: SSynt-rule 1

This rule stipulates that an adjectival lexeme Y of the common type that depends on a nominal lexeme X via the **modificative** SSynt-relation has to be linearly positioned to the left of X; the notation + ...+ means that a particular postposed dependent of Y may intervene between Y and X. Let it be emphasized that all I have said about linearization rules is extremely approximate; thus, the above rule has to include all conditions which license the anteposing of the adjective and all constraints on the gap between Y and X.

A fairly complete set of surface syntactic rules for English can be found in Mel'čuk & Pertsov (1987); note, however, that in these rules different linguistic aspects—linearization, morphologization and prosodization—were not as strictly separated as is the case now.

4. Summary of the main features of the MTT

I will now summarize the main features of the MTT and the corresponding models, providing a very broad and sketchy comparison with the mainstream linguistics.

1) Globality, descriptive orientation

The MTT is global in that it studies all aspects of linguistic organization together—lexical, semantic, syntactic, morphological and phonological; the globality of the MTT approach follows from the fact that this theory adopts a Saussurian view of language as a system in which everything holds together (*un système où tout se tient*) and which cannot be described in a satisfactory way if not studied in its entirety.

It puts emphasis on describing particular languages (Fr. *langues*), rather than on explaining the human language faculty in general (Fr. *langage*). However, in doing so, the MTT strives to describe languages in a unified way—using conceptual and formal apparatus that is valid cross-linguistically. The development of a coherent system of concepts and formal tools (representational languages: semantic networks, syntactic trees, lexical functions, etc.) to be used for linguistic description is thus one of the main goals of the theory (cf. in particular Mel'čuk, 1982 and Mel'čuk, 1993-1999). This in its turn entails a high degree of universality of the approach.

2) <u>Semantic bases and synthesis orientation, essential role of the paraphrase and of communicative organization</u>

The main tenet of the MTT is that in language 'everything starts with semantics,' i.e., that language is used above all to express meanings. Synthesis orientation and the interest for the paraphrase follow naturally from this basic insight. These features set the MTT apart from virtually all current linguistic approaches, which favor the study of syntax and are oriented towards analysis.

The MTT recognizes the crucial role of the paraphrase in language: meaning is taken to be invariant of paraphrases and speaking—virtual paraphrasing (cf. p. 8). Paraphrase—in particular, lexical paraphrase—is used as the main research tool in lexicology and elsewhere. Thus, from the perspective of the MTT, one of the most important tasks of modern linguistics is a development of a theory of paraphrase.

The MTT takes into account, in an essential way, communicative aspects of utterance organization; in this respect, it is close to the Prague School and Halliday's Systemic Linguistics. Communicative information (Theme/Rheme, Given/New, Focalized/Neutral, etc.) is specified in the starting Semantic Representation and is used to drive the synthesis, in particular the process of lexicalization. In other words, communicative organization of a starting semantic structure efficiently controls the production of paraphrases.

3) <u>Strong emphasis on the lexicon</u>

Given its semantic and synthesis orientation, with the dominant role of lexical paraphrasing, the MTT is crucially interested in describing lexical resources of the language, in particular its paraphrastic [= synonymic] means.

The entire process of synthesis in an MTM relies on the information stored in the lexicon. Thus, the mapping from semantic to deep-syntactic representations is carried out essentially

using lexical data—semantic description (= lexicographic definitions) of lexical units, as well as their syntactic and lexical co-occurrence, etc. Therefore, the lexicon—*The Explanatory Combinatorial Dictionary* —is the core component of the semantic module of an MTM.

Lexicographic work and the development of principles/methods of lexicographic description have occupied a central place in the MTT framework from the very beginning (cf. Mel'čuk & Zholkovsky, 1984, which presents the cumulative results of 20 years of lexicographic work, as well as Mel'čuk *et al*., 1984, 1988, 1992 and 1999); in contrast, the Western linguistic tradition considered lexicography as unworthy of linguist's attention and has started to recognize its relevance only recently.

4) Relational approach to language: the use of dependencies at all levels of linguistic description

The MTT considers relations, rather than classes, to be the main organizational factor in language, at all levels. Thus, the MTT uses hierarchical relations between elements of the corresponding structures on both the semantic and the syntactic level of representation:

- semantic dependencies (predicate ~ argument relations);
- communicative dependencies (communicative dominance relations);
- syntactic dependencies (specifying the dependent element whose syntactic behavior is controlled by another—governing—element, rather than the way these elements 'belong together,' i.e., what constituency does);
- morphological dependencies (control of inflectional forms).

This accent on dependencies, again, sets the MTT apart from mainstream linguistics, as a completely relational approach.

5) Formal character

As most current linguistic theories, the MTT is formal: it uses formal languages to represent linguistic data and to write rules that manipulate them. What distinguishes the MTT in this respect, however, is the richness and relative complexity of the formalisms used (see 6).

6) Stratificational and modular organization of MTMs

According to the MTT, linguistic modeling should reflect multidimensionality and heterogeneity of natural language. Consequently, MTMs

- have stratificational (= multi-representational) and modular organization, and
- use different formal languages for different levels of representation.

In contrast, most modern linguistic approaches look for homogeneous linguistic representations (cf., for instance, Head-Driven Phrase-Structure Grammar, where the only formalism used to represent linguistic objects are feature structures).

7) Implementability: the MTT lends itself well to computer applications

Being practically minded and formal, the MTT is predisposed to be applied—in Natural Language Processing, machine translation and computer-aided language teaching. Owing to their synthesis orientation, MTMs are particularly well suited for language generation and translation. They have been (partially) implemented in a number of applications, ranging from generation of weather forecasts/statistical reports, text reformulation/summarizing to machine translation; see 5.7 below.

5. Basic MTT Bibliography

This section offers a representative sample of MTT publications, under eight headings: general studies, collective volumes and conference proceedings, the four core linguistic disciplines, computational linguistics and NLP applications, applied linguistics.

A fairly complete MTT bibliography up to 1992 can be found in Mel'čuk *et al.* (1992), 59-93.

5.1 General Studies

Kahane, S. (2003). The Meaning-Text Theory. In: Agel, V, Eichinger L.M., Eroms, H-W., Hellwig, P. Heringer H.J. & Lobin, H., eds., *Dependency and Valency. An International Handbook of Contemporary Research*, vol. 1. Berlin/New York: De Gruyter, 546-570.

Mel'čuk, I. (2003). Levels of Dependency in Linguistic Description: Concepts and Problems. In: Agel, V, Eichinger L.M., Eroms, H-W., Hellwig, P. Heringer H.J. & Lobin, H., eds., *Dependency and Valency. An International Handbook of Contemporary Research*, vol. 1. Berlin/New York: De Gruyter, 188-229.

— (1997). *Vers une linguistique Sens-Texte. Leçon inaugurale*. Paris: Collège de France.

— (1985). Semantic Bases for Linguistic Description (Meaning-Text Linguistic Theory). In: Marino, M., Pérez, L. , eds., *The Twelfth LACUS Forum 1985*, 41-87.

—(1974). *Opyt teorii lingvističeskix modelej Smysl- Tekst*. Moskva: Nauka. [*Towards a Theory of Meaning-Text Linguistic Models*; reprinted in 1999, Moskva: Jazyki russkoj kul'tury]

Polguère, A. (1998). La théorie Sens-Texte. *Dialangue* 8-9: Université du Québec à Chicoutimi, 9-30.

Žolkovskij, A. & Mel'čuk, I. (1967a). O sisteme semantičeskogo sinteza. II: Pravila preobrazovanija [On a System of Semantic Synthesis (of texts). II: Paraphrasing Rules]. *Naučno-texničeskaja informacija 2, Informacionnye processy i sistemy*, 17-27.

— (1967b). O semantičeskom sinteze [On Semantic Synthesis (of texts)]. *Problemy kibernetiki* 19, 177-238. [French translation: Sur la synthèse sémantique (1970). *TA Informations* 2, 1-85].

— (1966). O sisteme semantičeskogo sinteza. I: Stroenie slovarja [On a System of Semantic Synthesis (of texts). I: Building the Dictionary]. *Naučno-texničeskaja informacija* 11, 48-55.

— (1965). O vozmožnom metode i instrumentax semantičeskogo sinteza [On a Possible Methodology and Tools for Semantic Synthesis (of texts)]. *Naučno-texničeskaja informacija* 5, 23-28.

5.2 Collective Volumes and Conference Proceedings

Apresjan, Y. & Iomdin, L., eds. (2005). *Proceedings of the Second International Conference on the Meaning-Text Theory, Moscow, June 21-23, 2005*. Moscow: Languages of the Slavic Culture Publishing House.

Grossmann, F. & Tutin, A., réds. (2003). *Les collocations: analyse et traitement*. Université Stendhal-Grenoble 3.

Kahane, S. & Nasr, A., eds. (2003). *Proceedings of the First International Conference on the Meaning-Text Theory, Paris, June 16-18, 2003*. Paris: Ecole Normale Supérieure.

Polguère, A., ed. (to appear). *Syntactic Dependency in Linguistic Description*. Amsterdam/Philadelphia: Benjamins.

Steele, J., ed. (1990). *Meaning-Text Theory: Linguistics, Lexicography and Implications*. Ottawa etc.: University of Ottawa Press.

Wanner, L., ed. (to appear). *Selected Lexical and Grammatical Topics in Meaning-Text Theory. In Honor of Igor Mel'čuk*. Amsterdam/Philadelphia: Benjamins.

— (1998). *Recent Trends in Meaning-Text Theory*. Amsterdam/Philadelphia: Benjamins.

— (1996). *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins.

5.3 Semantics and Lexicology

Alonso Ramos, M. (2004a). Elaboración del Diccionario de colocaciones en español y sus aplicaciones. In: P. Bataner & J. de Cesaris, eds., *De Lexicographia. Actes del I Symposium internacional de Lexicografía*, Barcelona, IULA-Edicions Petició, 149-162

Apresjan, Ju. (2000). *Systematic Lexicography*. Oxford: Oxford University Press.

— (1974). *Leksičeskaja semantika. Sinonimičeskie sredstva jazyka*. Moskva: Nauka. [English translation: 1992. *Lexical Semantics: User's Guide to Contemporary Russian Vocabulary*. Ann Arbor: Karoma Publishers.]

Iomdin, L. (1972). Ob odnom klasse anglijskix kauzativnyx glagolov, obrazovannyx sposobom konversii [On a Class of English Causative Verbs Formed by Conversion]. *Mašinnyij perevod i prikladnaja lingvistika* 15, 99-142.

Kahane. S & Polguère, A. (2001). Formal Foundations of Lexical Functions. In: Daille, B. & Williams, G., eds., *Workshop on Collocations, ACL 2001*, Toulouse, 8-15.

Mel'čuk, I. (2001). *Communicative Organization in Natural Language*. Amsterdam/Philadelphia: Benjamins.

— (1996). Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In: Wanner, L., ed. (1996), 37-102.

— (1995). The Future of the Lexicon in Linguistic Description and the Explanatory Combinatorial Dictionary, In: I.-H. Lee (ed.), *Linguistics in the Morning Calm 3 (Selected Papers from SICOL-1992)*, Seoul, 181-270.

— (1992). Paraphrase et lexique: la théorie Sens-Texte et le *Dictionnaire explicatif et combinatoire*. In: Mel'čuk *et al.*, 1992, 9-59.

Mel'čuk, I. *et al.* (1984, 1988, 1992, 1999). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I-IV*. Montreal: PUM.

Melčuk, I. A., Clas, A. & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.

Mel'čuk, I. & Zholkovsky [= Žolkovskij], A. (1984). *Tolkovo-kombinatornyj slovar´ russkogo jazyka. Opyty semantiko-sintaksičeskogo opisania russkoj leksiki* [*Explanatory Combinatorial Dictionary of Russian. Towards a Semantic and Syntactic Description of Russian Lexicon*]. Wienna: Wiener Slawistischer Almanach.

Milićević, J. (in press). *La paraphrase. Modélisation de la paraphrase langagière*. Bern: Peter Lang.

Polguère. A. (2003). Étiquetage sémantique des lexies dans la base de données DiCo. *T.A.L.* 44: 2, 39-68.

— (2000a). Towards a Theoretically Motivated General Public Dictionary of Semantic Derivations and Collocations for French. In: *Proceedings of EURALEX'2000*, Stuttgart, 517-527.

5.4 Syntax

Alonso Ramos, M. (2004b). *Las construciones con verbos de apoyo*. Madrid: Visor.

Beck, D. (2002). *The Typology of Parts of Speech Systems: the Markedness of Adjectives*. Routledge Outstanding Dissertations in Linguistics.

Gerdes. K. (2002). *Topologie et grammaires formelles de l'allemand*. Thèse de doctorat. Université Paris 7, Paris.

Iordanskaja, L. (1963). Svoistva pravil'noj sintaksičeskoj struktury (na materiale russkogo jazyka) [Properties of a Well-Formed Syntactic Structure (Illustrated from Russian)]. *Vorprosy jazykoznania* 4, 102-112.

Kahane, S. (2001). Grammaires de dépendance formelles et Théorie Sens-Texte. Tutoriel. In: *Actes de TALN 2001, Tours,* 17-76.

Korel'skaja, T. (1975). *O formal'nom opisanii sintaksičeskoj sinonimii* [*On a Formal Description of Syntactic Synonymy*]. Moskva: Nauka.

Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. Albany: State University of New York Press.

Mel'čuk, I. & Pertsov, N. (1987). *Surface Syntax of English. A Formal Model within the Meaning-Text Framework*. Amsterdam/Philadelphia: Benjamins.

Žolkovskij A. (1971). *Sintaksis somali (glubinnye i poverxnostnye struktury)* [*Syntax of Somali (Deep- and Surface Structures)*]. Moskva: Nauka.

5.5 Morphology

Mel'čuk, I. (2006). *Aspects of the Theory of Morphology*. Berlin/New York: Mouton de Gruyter.

— (1993-2000). *Cours de morphologie générale*, vol. 1-5. Montreal/Paris: PUM/C.N.R.S. Éditions.

— (1982). *Towards a Language of Linguistics. A System of Formal Notions for Theoretical Morphology*. München: Fink.

5.6 Phonology

Yoo, H-Y. (2002). Ordre des mots et prosodie: essai de description et de formalisation pour le grec moderne. Thèse de doctorat. Université Paris 7, Paris.

5.7 Computational Linguistics and NLP Applications

Apresjan Ju. *et al*. (2003). ETAP-3 Linguistics Processor: a Full-Fledged Implementation of the MTT. In: Kahane, S. & Nasr, A., eds. (2003), 279-288.

— (1992*). Lingvističeskii processor dlja složnyx informacionnyx system* [*A Linguistic Processor for Complex Information Systems*]. Moskva: Nauka.

— (1989). *Lingvističeskoe obespečenie sistemy ÈTAP-2* [*Linguistic Software for the System ETAP-2*]. Moskva: Nauka.

Apresjan, Ju. & Tsinman, L. (1998). Perifrazirovanie na kompjutere [Paraphrasing on the Computer]. *Semiotika i informatika* 36, 177-202.

Boguslavskij, I., Iomdin. L. & Sizov. V. (2004). Multilinguality in ETAP-3. Reuse of Linguistic Ressources. In: *Proceedings of the Conference Multilingual Linguistic Ressources. 20th International Conference on Computational Linguistics, Geneva 2004*, 7-14.

Boyer, M. & Lapalme, G. (1985). *Generating Paraphrases from Meaning-Text Semantic Networks*. Montreal: Université de Montréal.

CoGenTex (1992). *Bilingual Text Synthesis System for Statistics Canada Database Reports : Design of Retail Trade Statistics (RTS) Prototype. Technical Report 8*. CoGenTex Inc., Montreal.

Iordanskaja, L., Kim, M., Kittredge, R., Lavoie, B. & Polguère, A. (1992). Generation of Extended Bilingual Statistical Reports. In: *COLING-92*, Nantes, 1019-1022.

Iordanskaja, L., Kim, M. & Polguère, A. (1996). Some Procedural Problems in the Implementation of Lexical Functions for Text Generation. In: Wanner, L., ed., (1996), 279-297.

Iordanskaja, L., Kittredge, R. & Polguère, A. (1991). Lexical Selection and Paraphrase in a Meaning-Text Generation Model. In: Paris, C. L., Swartout, W. R. & Mann, W. C., eds., *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Boston: Kluwer, 293-312.

Iordanskaja, L. & Polguère, A. (1988). Semantic Processing for Text Generation. In: *Proceedings of the First International Computer Science Conference-88, Hong Kong, 19-21 December 1988*, 310-318.

Kahane, S. & Mel'čuk, I. (1999). Synthèse des phrases à extraction en français contemporain (Du graphe sémantique à l'arbre de dépendance). *T.A.L.*, 40:2, 25-85.

Kittredge, R. (2002). Paraphrasing for Condensation in Journal Abstracting. *Journal of Biomedical Informatics* 35: 4, 265-277.

Kulagina, O. & Mel'čuk, I. (1967). Automatic Translation: Some Theoretical Aspects and the Design of a Translation System. In: Booth, A.D., ed., *Machine Translation*. Amsterdam: North-Holland Publishing Company.

Lareau, F. (2002). *La synthèse automatique de paraphrases comme outil de vérification des dictionnaires et grammaires de type Sens-Texte*. Mémoire de maîtrise. Département de linguistique et de traduction, Université de Montréal.

Mel'čuk, I. (1967). Ordre des mots en synthèse automatique des textes russes. *T.A. Informations*, 8:2, 65-84.

Mel'čuk, I. & Wanner, L. (to appear). Syntactic Mismatches in Machine Translation. To appear in *Machine Translation*.

— (2001). Towards a Lexicographic Approach to Lexical Transfer in Machine Translation (Illustrated by the German—Russian pair). *Machine Translation* 16, 21-87.

Nasr, A. (1996). *Un modèle de reformulation automatique fondé sur la théorie Sens-Texte : applications aux langages contrôlés*. Thèse de doctorat en informatique. Paris: Université Paris 7.

Polguère, A. (2000b). A 'Natural' Lexicalization Model for Language Generation. In: *Proceedings of the Fourth Symposium on Natural Language Processing 2000 (SNLP'2000)*. Chiangmai, May 10-12, 2000, 37-50.

— (1990). *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*. Thèse de doctorat. Montréal: Université de Montréal.

5.8 Applied Linguistics

Alonso Ramos, M. (2002). Un vacío en la enseñanza del léxico del español como lengua extranjera. In A. Braasch & C. Povlsen, eds., *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002,* Copenhagen, vol. 2, 551-561.

Anctil, D. (2005). Maîtrise du lexique chez les étudiants universitaires : typologie des problèmes lexicaux et stratégies de résolution de problèmes lexicaux. Mémoire de maîtrise. Montréal: Université de Montréal.

Gentilhomme, Y. (1992). Panorama sur le *Dictionnaire explicatif et combinatoire* : retombées pégagogiques. In: Mel'čuk *et al*. (1992), 95-120.

Milićević, J. & Hamel, M.-J. (to appear). Un dictionnaire de reformulation pour les apprenants du français langue seconde. In: *Actes du 30ᵉ colloque de L'Association Linguistique des Provinces Atlantiques*, 2-3 novembre 2005. Moncton.

Polguère. A. (2000c). Une base de données lexicales du français et ses applications possibles en didactique. *Revue de Linguistique et de Didactique des Langues* 21, 75-97.

— (2004). La paraphrase comme outil pédagogique de modélisation des liens lexicaux. In: Calaque, E. & David. J., réds., *Didactique du lexique : contextes, démarches, supports*. Collection « Savoirs en pratique ». Bruxelles: De Boeck, 115-125.

Tremblay, O. (2004). Une approche structurée de l'enseignement/l'apprentissage du lexique en français langue maternelle basée sur la lexicologie explicative et combinatoire. Mémoire de maîtrise. Montréal: Université de Montréal.

Acknowledgment