

Published: *CLS 37-1, The Main Session, 2001, 395-421.*

**A formal language for linguistic morphology:
Towards a coherent conceptual system**

Igor Mel'čuk
University of Montreal

—But, Doc, do you really believe terminology is that important?¹
[The catch-line of a Russian joke; see Fnt. 1
for the complete and unexpurgated text.]

1 A little history

When I started doing linguistics about half a century ago, my first impression was that of conceptual and terminological mess. The same term would be interpreted in ten different ways by different authors or even in different works by the same author; and a seemingly identical concept would go under dozen different names. This was true about most basic linguistic concepts: word, phrase, clause, sentence; morph and morpheme; syntactic structure (regularly called 'phrase-marker'); meaning; etc. The situation contrasted unfavorably with mathematics: there, you began with rigorous definitions and, based on well-defined concepts, you could reason without danger of falling into vicious circles or developing absurdities. Why was it so different in the science of language? Even the great classics (Saussure, Bloomfield, Hjelmslev, Jakobson,...), even my beloved Reformatskij (1947) were not rigorous enough to my taste—and therefore, not fully understandable. Even in newborn machine translation (where I was active) and Chomsky's generative grammar the terminological chaos was complete.

And then by chance I read about Nicolas Bourbaki, a French mathematician that never existed, but nonetheless produced a many-volume treatise on modern mathematics. Under this *nom de plume* was hiding—at the Academy of Poldavia and the University of Nancago²—a team of French mathematicians (including one Pole) lead by Jean Dieudonné, André Weil and Henri Cartan. Between 1930 and 1955 they successfully tried to create a unified metalanguage for all branches of mathematics, so that it became possible to describe algebra, geometry, calculus, number theory, probability theory, etc. using the same conceptual apparatus. As a result, mathematicians were able to reveal identical structures in seemingly very different domains and thus offer a common perspective on the whole of mathematics. At the

same time, Bourbaki's work led to a clearer and more unified terminology and notational system. Without delving deeper into the significance of Bourbaki's approach for mathematics, I can say that I immediately wanted to launch a similar enterprise in linguistics and contacted several friends and colleagues proposing a collective work with similar goals. Unfortunately, only one person accepted—Aleksandr ('Alik') Žolkovskij (now, Professor A. Zholkovsky at the University of Southern California, Los Angeles); but even he preferred working on semantics and the basics of the Meaning-Text theory. I was left alone with my beautiful idea.

If I could foresee that it would take me more than 35 years and the advent of the personal computer to finish the job, I most probably would lack the courage to start it. But I was young and extremely self-confident: I did not foresee the difficulties. I did as Napoleon suggested: 'D'abord on s'engage, et puis on voit.' When I saw where I was heading and how much I had to accomplish it was already too late; huge mental and material resources had been already invested, and I could not afford wasting all this effort. So I plodded forward. In 1982, an intermediate presentation of my project was published as a book (Mel'čuk 1982); this allowed me to see better the shape of the future work, and I proceeded to writing the final text. After many adventures and false starts (thus, a publisher who had accepted the manuscript and prepared the proofs of the first volume suddenly dropped the project without ever explaining his decision...), Mel'čuk 1993 finally appeared. It took me and the publisher³ another seven years, but in 2000 my project reached its completion: the last, 5th volume of *Cours de morphologie générale (théorique et descriptive)* [= 'Course of General Morphology (Theoretical and Descriptive)'] was printed.

2 « Cours de morphologie générale » [= CMG]

2.1 The main features of the CMG

CMG proposes a system of 248 morphological concepts (= definitions), which is organized in a strictly deductive manner:

- There is a set of *indefinibilia*, or concepts that should not be defined within the *CMG* itself. These 'undefinable' concepts include:
 - (i) All mathematical and logical concepts needed in definitions—such as *set/element/ belong to a set, string, conjunction/disjunction/implication*, etc.
 - (ii) All linguistic concepts needed in definitions, but coming from linguistic disciplines other than morphology, namely from semantics, syntax, and phonology—such as *predicate/argument, lexical unit, syntactic relation, syntactic actant, phoneme, prosodeme, tone*, etc.

(iii) Three basic concepts (see next section).

These undefinable concepts are not rigorously defined, but characterized as fully as possible and well illustrated, so that it can be hoped that the reader will be able to manipulate them without confusion.

- All the 248 concepts are defined only either in terms of the undefinable concepts or concepts that have been previously defined.

The *CMG* conceptual/terminological system possesses the following important FORMAL properties:

—No polysemy: no term refers to two concepts; in case the 'physical' signifiers of two terms coincide, a distinguishing numerical subscript is used (e.g., *reduplication*₁ vs. *reduplication*₂, etc.).

—Controlled synonymy: as a general rule, no concept is referred to by two terms; in a few cases where the use of two different but synonymous terms is imposed by pedagogical considerations, they are allowed by an explicit statement.

—Mutual substitutability: in each definition, each term can be replaced by its own definition, and each definition can be replaced by the corresponding term—*salva significatione*.

—Maximal block principle: in a definition, each fragment that corresponds to a concept must be replaced with the corresponding term, so that a concept is always defined in a strict stepwise fashion.

SUBSTANTIALLY, the definitions of the *CMG* follow two principles:

—For each prescientific notion *N*, the *CMG* tries to define as general a concept as possible, in order to cover all cases judged to be subsumed under *N*. Then all varieties of *N* are defined as particular cases of the concept in question. In other words, the adopted policy is to have first the 'maximal' concept available and to subdivide it afterwards.

—Morphological concepts are defined using a calculus of logical possibilities. Such an approach ensures an exhaustive coverage of the domain considered.

The goals of the *CMG* are:

- 1) to stabilize morphological terminology;
- 2) to offer a working morphologist an inventory of variegated, well-sharpened tools for morphological descriptions of languages—in such a way that Mr. A's statements are easily and unambiguously understood by Mr. B;
- 3) to prepare the ground for typological studies in the morphology domain: all the concepts introduced are universal and thus allow for easy comparison of the phenomena found in different languages.

The *CMG* is a combination of a concise linguistic encyclopedia, a typology research monograph, and a high-level manual.

2.2 Concepts introduced in the *CMG*

The three basic (undefinable) concepts are the *signified*, the *signifier*, and the *syntactics* (= a set of data on cooccurrence of a pair <signified; signifier> with other such pairs). Linguistic sign is then defined as a triplet <signified; signifier; syntactics>. Thus, the radical of the French verb CHANTER is represented as

< [to] sing' ; /šãt/ ; Σ = verb, transitive, 1st conjugation, ...>.

The German circumfix *ge-...-t*, as in *ge+frag+t* ('asked-PAST.PART'), appears as

< _{PAST.PART}' ; /ge+...+t/ ; Σ = circumfix, verbal,
used with regular verbs without inseparable prefixes and not in *-ieren*,
ge- being positioned before the radical but after the separable prefix, ...>

The concept of representability is introduced: a sign *s* is representable in terms of signs *s*₁, *s*₂, ..., *s*_n if and only if $s = \oplus\{s_1, s_2, \dots, s_n\}$, where \oplus is the meta-operation of linguistic union (\oplus puts together signs and components of signs according to their syntactics and general rules of the language). The concepts of representability and quasi-representability (= representability in the signified or the signifier only) allows one to define the elementary sign (= not representable and not quasi-representable) as well as suppletive signs (quasi-representable in their signifieds) and phraseologized signs (quasi-representable in their signifiers). And so forth, until all the relevant concepts are obtained.

2.3 The structure and the contents of the *CMG*

CMG consists of seven parts (grouped in 5 volumes):

- Part I gives a general outline of the Meaning-Text approach that underlies the *CMG*. It introduces then all initial concepts, among which the central is *wordform*—the main object of morphology, defined as a particular type of linguistic sign. Grammatical meanings—inflectional (*grammemes*) and derivational (*derivatememes*)— are defined as well.

Since a wordform is a linguistic sign, which has a signified, a signifier and syntactics, the next three parts deal with these components taken separately.

- Part II discusses morphological SIGNIFIEDS. More specifically, it defines and describes inflectional categories, such as nominal number, case, tense, mode, voice, etc., and their grammemes. Then main derivatememes are presented as well.

- Part III deals with morphological SIGNIFIERS. It proposes a typology thereof: signifier entities—segments and suprasegmentals vs. signifier operations—alternations, reduplications₁, and conversions₁.

- Part IV covers morphological SYNTACTICS; its central concepts are grammatical gender and nominal class.

After the three components of linguistic signs have been studied separately, signs as such can be considered.

- Part V is dedicated to morphological signs of all possible types. First of all, morphs (radicals and affixes) are considered in connection with morphemes; then also apophonies, reduplications₂, conversions₂, etc. A full typology of affixes is proposed. Some minor but important issues are dealt with at length: suppletion, zero and empty signs, possible relations between linguistic signs.

As the next step, the *CMG* sets out to demonstrate how the introduced concepts 'work' in concrete morphological descriptions, that is, in morphological models of particular languages. (A morphological model is a system of formal rules that synthesizes and analyzes wordforms.)

- Part VI contains three complete morphological models: Spanish conjugation, Hungarian declension, and Bafia conjugation.

The last step is the study of the way a morphological description is done. Therefore,

- Part VII touches on the principles of morphological description, i.e. on metamorphology. The principles discussed include the Principle for postulating a zero sign, the Principle of the more general description (it is used in the cases of non-uniqueness of descriptive solutions), etc.

3 An example: A definition of voice and a calculus of grammatical voices

In order to illustrate the contents of the *CMG*, I will present here the definition of the inflectional category of voice and then the calculus of its possible grammemes. (Of course I cannot and will not explain all of the intermediate concepts.)

I start with the definition of the central underlying concept: diathesis.

|| The diathesis of a lexical unit L is the correspondence between L's Semantic Actants [= SemAs] and L's Deep-Syntactic Actants [= DSyntAs].

For instance, the English verb [*to*] LIKE and the French verb PLAIRE

(≈ [*to*] like) have the same SemAs, but different diatheses:

‘X experiences pleasant feelings caused by a contact of X with Y’

[to]	LIKE

	PLAIRE

Now **voice** can be readily defined:

Voice is an inflectional category whose grammemes specify such modifications of the diathesis of a lexical unit L that do not affect the propositional meaning of L.

For a binary diathesis, there are 11 possible modifications (= derived variants), obtainable by the following three operations:

- permutation of L's DSyntAs (with respect of L's SemAs);
- suppression of L's DSyntAs;
- referential identification of L's SemAs.

To illustrate these modifications, I have first to introduce the conventions to be respected when determining the numbers of DSyntAs of L:

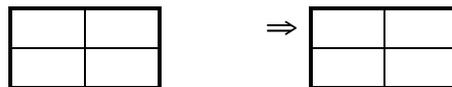
1. The DSyntAs of L are numbered consecutively: e.g., the numbering *I, III is disallowed.
2. The numbering of DSyntAs of L must begin at I or II: e.g., the numbering *III, IV is disallowed.
3. The DSyntAs of L must be numbered without repetitions: e.g., the numbering *II, II, III is disallowed.

Here are logically possible binary diatheses obtained by mechanical application of the three above operations:

The shadowed variants are 'illegitimate:' they either violate numbering conventions for DSyntAs (the asterisked ones) or else they coincide with some other variant, already on the list. The subtraction of illegitimate variants gives us 12 logically possible grammemes of voice for a prototypical binary diathesis.

Grammmemes of voice:

- 1) 'active' : zero modification of the basic diathesis
 ('John is-combing Mary's hair')



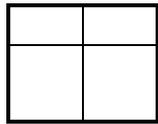


(1) Lat. *Xenophōn+Ø agricultur+am lauda+ba+tt+Ø*

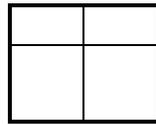
‘Xenophon [I, Subj, NOM] praised [the] agriculture [II, DirO, ACC)].

2) ‘FULL PROMOTIONAL PASSIVE’ : bilateral permutation of DSyntAs, which produces a diathesis converse with respect to the basic one

(‘Mary's hair is being-combed by-John’)



⇒



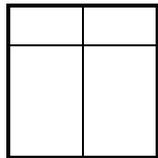
(2) Lat. *A Xenophōnt+e agricultur+a lauda+ba+tt+ur*

lit. ‘By Xenophon [II, AgCo] [the] agriculture was praised’.

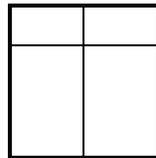
3) ‘PARTIAL DEMOTIONAL PASSIVE’ : permutation—or, more precisely, pure demotion, see below—of the DSyntA I only (to III), with

the DSyntA II retained in place (‘[“It”] is-combing

Mary's hair by John’)



⇒



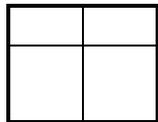
(3) Ukr. *Mnoju bu+l+o spleče+no cju sum+u*

lit. ‘By-me [III, AgCo, INSTR] [“it”] was paid this sum [II, DirO, ACC)] =

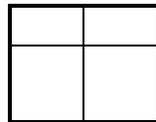
‘I paid this sum’.

4) ‘FULL DEMOTIONAL PASSIVE’ : pure demotion of both the DSyntAs I and II

(‘[“It”] is-combing by-John at-Mary's hair’)



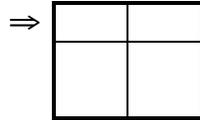
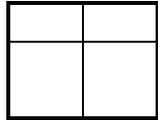
⇒





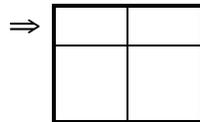
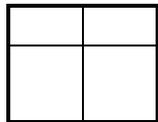
I did not find actual examples of this voice.

- 5) (SUBJECTLESS SUPPRESSIVE) : suppression of the DSyntA I, i.e. what should become, at the SSynt-level, the Subject
 (["They"] are-combing Mary's hair²)



- (4) Est. *Ehitata+kse sild+a*, lit. 'Build bridge [II, DirO, PART(itive)]².

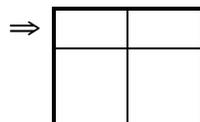
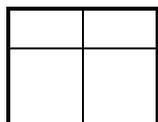
- 6) (OBJECTLESS SUPPRESSIVE) : suppression of the DSyntA II, i.e. of what should become, at the SSynt-level, the DirO ('John is-combing')



- (5) Apapantilla Totonac

Tamāwá pancín 'He buys bread'. ~ *Tamāwa+nán* 'He is buying (things)'.

- 7) (ABSOLUTE SUPPRESSIVE) : suppression of both DSyntAs I and II ('There-is-combing')



- (6) Germ. *Hier wird viel gelesen*, lit. 'Here becomes much read'.

- 8) (AGENTLESS PROMOTIONAL PASSIVE) : permutation of DSyntAs, with suppression of the DSyntA II, i.e. the one which should correspond to X

('Mary's hair is-being-combed')





(7) Ar. *Al-žisr+u jubnaŝu*, lit. 'The bridge [I, Subject, NOM] is-being-built[PASS]?'.

9) {PATIENTLESS DEMOTIONAL
PASSIVE} :

permutation of DSyntAs, with suppression of the DSyntA I, i.e. the one which should correspond to Y

(["It"] is-combing by-John)



□I□II□□II□—□□(8) a. Rus. *Mne mečtaetsja*, lit. ['It"] to-me [II, IndirO, DAT] is-dreaming?.

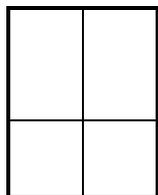
b. Germ. *Von Politikern wird hier gegessen*, lit. 'By politicians becomes here eaten?.

10) {OBJECTLESS REFLEXIVE} : □referential identification of the SemAs, with suppression of the DSyntA II ('John is-combing-himself') □□
 $X \square Y \square \Rightarrow \square X = Y \square \square \square I \square II \square \square I \square \square \square (9)$ Rus. *Otec+Ø pričěsyva+Ø+et+sja*

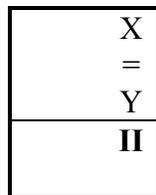
lit. 'Father [I, Subject] is-combing-himself' = 'Father is combing his hair'.

11) {SUBJECTLESS REFLEXIVE} :

referential identification of the SemAs, with suppression of the DSyntA I ('By-John is-combing-himself')



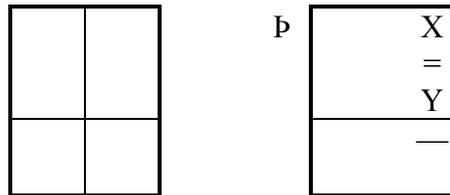
⇒



(10) Lit. *Jon+o su+si+čukuo+t+a*

lit. 'By-Jonas [II, AgCo] have-been-combed-itself' = 'Jonas has combed his hair).

12) (absolute reflexive) : identification of the SemAs, with suppression of both the DSyntAs I and II ((There-is-combing-oneself))



(11) Pol. *Uczesa+no si'*, lit. (Have-combed itself) = (Some people have combed their hair).

□ The joke quoted in the motto:

Someone knocks at a doctor's door at the dead of the night and once admitted, tells the half-asleep doctor:

—Now, Doc, castrate me as soon as possible!

—Listen, buster, are you crazy?

—Not at all, but I cannot wait. Will you castrate me right away?

—I'll call the police, that's what I'll do!

—So!—and the would-be patient produces an automatic. —Get on with it, Doc, and do what I tell you—or I'll shoot.

The doctor obeys, takes the guy to the operation table and operates on him. When everything is over and the happy patient offers to pay, the doctor asks him for the reason for such a bizarre wish and the extreme urgency.

—It is quite simple. I am terribly in love with a Jewish girl who said she would go to bed with me and eventually even marry me but under the strict condition that I get castrated before next Friday—and tonight is Thursday night.

—O gosh, perhaps she said 'circumcized'?

—Yeah, I guess this was the term she used. But, Doc, do you really believe terminology is that important?

□ Before the Second World War, J. Dieudonné and A. Weil taught at the University of Nancy, and during the war, at the University of Chicago. Because of this, it was an additional pleasure for me to give this talk at the University of Chicago.

□ The publishing house that started the publication of the book—Les Presses de l'Université de Montréal —was sold to a commercial publisher, who (for legal reasons) interrupted the production.,