selected
papers
of

the 10th
International
Conference of
Greek
Linguistics

Edited by

Zoe Gavriilidou
Angeliki Efthymiou
Evangelia Thomadaki
Penelope Kambakis-Vougiouklis

Komotini 2012

PHRASEOLOGY:
ITS PLACE IN THE LANGUAGE, IN THE DICTIONARY, AND IN NATURAL LANGUAGE PROCESSING

Igor Mel'čuk
Observatoire de linguistique Sens-Texte
University of Montreal, Canada
igor.melcuk@umontreal.ca

ABSTRACT

The paper introduces the concept of phraseme—that is, non-free, or selectionally constrained, phrase or clause—and characterizes the three major classes of phrasemes: idioms, collocations and clichés (the latter including an important subclass of pragmatemes).A finer classification of phrasemes is elaborated, based on the notions of compositionality of complex linguistic signs and that of the semantic pivot of a meaning. After this, are described the techniques used to represent phrasemes of all types in an Explanatory Combinatorial Dictionary *and present three examples illustrating the use of this dictionary for difficult cases of translation involving phrasemes.*

> You can't BUY TIME or SAVE TIME common idioms notwithstanding. You can only SPEND TIME.
>
> Eric Zorn, a columnist for the *Chicago Tribune*

1. Introduction

The expressions printed in small caps in the motto are examples of what is known as phrasemes; phrasemes of a language constitute its phraseology. Linguists agree that phraseology is extremely important for linguistic studies, but, quite unfortunately, they do not agree either on the exact content of the notion 'phraseology' or on the way phraseological expressions should be described and treated in linguistic applications, among others, in lexicography and Natural Language Processing [= NLP]. I will address these two points: Section **2** proposes a rigorous definition of phraseme, a characterization of the major classes of phrasemes and an exhaustive phraseme typology; Section **3** sketches the fundamentals of a lexicographic description of phrasemes, while in Section **4** I consider three examples of difficult cases of automatic translation where the solutions come from the proposed lexicographic description of one of phraseme classes (namely, collocations). Section **5** summarizes the most important points of the paper.

My theoretical framework is Meaning-Text Theory [= MTT]. I have to use certain of its notions and formalisms without explanation. For more on MTT, please consult Mel'čuk 1974, 1981, 1988a: 43-91, 1997 and Kahane 2003a.

Technical terms appear, on their first mention, in Helvetica.

2. Phraseology in the Language

The literature on phraseology is too huge to be reviewed here even cursorily; see, for instance, Cowie and Howarth 1996, and the collections Everaert *et al*. (eds) 1995, Cowie (ed.) 1998, Burger *et al.* (eds) 2007, and Anscombre and Mejri (eds) 2011. I will limit myself to Mel'čuk 1995 (a sketch of a theory of phraseology within the Meaning-Text framework) and the classics Bally 1909 and Weinreich 1969, which have most profoundly influenced my approach to phrasemes.

2.1 Two Main Families of Phrasemes: Lexical and Semantic-Lexical Phrasemes

A phraseological expression is a multiword utterance—that is, a linguistic expression formed by several lexemes syntactically linked in a regular way (in what follows, only multiword utterances are considered, which allows me to omit the modifier *multiword*).[1] The notorious expression *X kicks the bucket* ≈ 'person X dies of natural causes, I being flippant about X' is syntactically and morphologically structured exactly the same way as all similar phrases of the form "Transitive Verb—DirO:"*kick the ball, hit John,squeeze her hand*, etc. Even the expression *kick the bucket* itself can mean 'kick the bucket [full of dirty water]'! This expression is special, i.e. phraseological, only because of its "unpredictable" meaning. A phraseological expression, or phraseme, is thus an utterance featuring some unpredictable properties, i.e., a constrained utterance, or an utterance that is not free. Therefore, I begin with a definition of free utterance.

DEFINITION 1: FREE UTTERANCE

> An utterance **U** is free, or non-constrained, if and only if [= iff] each of its lexical components Lis selected by the speaker strictly for its meaning and in conformity with its linguistic properties but independently of the lexical identity of other lexical components of **U**.

Corollary: Every lexical component of a free utterance **U** can be replaced by any of its (quasi-)synonyms without taking into account any other lexical component of **U**, and this replacement does not affect **U**'s meaning and grammaticality. In the phrase *select the word freely*, you can replace any component with no matter which of its synonyms and the meaning as well grammaticality is preserved: *choose the lexeme without constraint*.

DEFINITION 2: NON-FREE UTTERANCE = PHRASEME

> An utterance**U** is non free, or constrained⟨= phraseologized⟩, iff at least one of its lexical components Lis selected by the speaker as a function of the lexical identity of other component(s) of **U**.

In a non-free utterance **U**, at least one Lis selected depending on **U**'s other particular lexemes; such an utterance is called, as I just said, phraseme.

Corollary: In a phraseme, not every lexical component can unrestrictedly be replaced by any of its (quasi-)synonyms without affecting the phraseme's meaning and grammaticality. In *kick the bucket* ≈ 'die' you cannot replace any of the components: *boot the bucket* or *kick the pail* do not mean 'die'.

A phraseme violates the freedom of selection of its lexical components on the paradigmatic axis of speech production, as the speaker is looking (in his mental lexicon) for appropriate lexical units. The selection activity proceeds in two stages:

• First, the speaker constructs his starting meaning; he selects the necessary simpler meanings and unites them into the meaning of his eventual utterance—that is, into its starting Semantic Representation [= SemR].

• Second, the speaker selects lexical units needed to express his starting SemR and unites them into the Deep-Syntactic Representation [= DSyntR] of the utterance.

Accordingly, two cases of violation of lexical selection freedom can be distinguished in the process of constructing an utterance.

The first case. The construction of the starting meaning 'σ' [= the SemR] of the utterance L('σ') is free. To obtain 'σ', the speaker selects simpler meanings 'σ$_1$', 'σ$_2$', …, 'σ$_n$' and puts them together in conformity with his needs and general rules of his language: the language does not specifically constrain the speaker's semantic choices. But the lexical components of the utterance L('σ') cannot be chosen freely: some or all of them are selected as a function of the others. The violation of the selection freedom takes place in the transition {SemR} ⇔ {DSyntR} and manifests itself in lexical constraints. Therefore, resulting phrasemes are called lexical (boldface in examples indicates the lexical components selected restrictedly):***kick the bucket***, ***pull*** [N$_Y$'*s*] ***leg*** 'lie [to N$_Y$] in order to have fun' or Rus. ***na golubom glazu*** lit. 'on a blue eye' = 'pretending to act honestly and sincerely';*The rain is falling in torrents, It rains **cats and dogs*** or Rus. *Dožd´ **l´ët kak iz vedra***lit. 'Rain is.pouring as from bucket'and ***prolivnoj*** *dožd´*'torrential rain'are typical lexical phrasemes.

---

[1] The term utterance is used here to refer to the set of linguistic expressions including (beside words) phrases, clauses and sentences, since a phraseme can be any of the above. I leave aside the phrasemes of the morphological level—that is, the phraseologized combinations of morphs within a wordform: *fight+er* 'military plane...' or *for+get*. For this family of phrasemes, or morphophrasemes, see, for instance, Beck and Mel'čuk 2011.

DEFINITION 3: LEXICAL PHRASEME

‖ A phraseme is lexical iff its meaning is constructed by the speaker freely, but its lexical ‖ components L(all or some) are selected in a constrained way.

The second case. Not only the lexical composition of the phraseme is constrained, but also its meaning. To describe a real-world situation, the Speaker is forced by the language to select the starting meaning 'σ', and he cannot take an equivalent meaning 'σ'' or 'σ''' ('σ' ≡ 'σ'' ≡ 'σ'''). As a result, the utterance L('σ') is constrained semantically **and** lexically. This type of phraseme is thus "doubly" constrained: in the transition {ConceptR} ⇔ {SemR} (by semantic constraints) and then in the transition {SemR} ⇔ {DSyntR} (by lexical constraints). This is a semantic-lexical phraseme. A simple example is the sign *Wet paint*: Russian says in this context *Ostorožno, okrašeno* lit. 'Caution, painted' rather that *Syraja kraska* 'Wet paint' or even *Ostorožno, **vy**krašeno* (with a different aspect prefix); and in English it would be inappropriate to write on a sign *Caution, freshly painted*, although this is a perfectly grammatical and semantically correct utterance. Here the language prescribes the meaning to express in the given situation and its specific lexical expression.

DEFINITION 4: SEMANTIC-LEXICAL PHRASEME

‖ A phraseme is semantic-lexical iff not only all its lexical, but also all the componentsof its ‖ meaning are selected by the Speaker in a constrained way.

In other words, a semantic-lexical phraseme is selected as a whole (just like idioms, see **2.3.1**).
Examples: *in other terms/in other words*; *to make a long story short*; Rus. *inače govorja* lit. 'speaking differently', *koroče govorja* lit. 'speaking shorter'or *čto i trebovalos´ dokazat´* 'Q.E.D.' [= Lat. *Quod Erat Demonstrandum*].

Thus, this major partition splits phrasemes into two main families: lexical phrasemes and semantic-lexical phrasemes.

2.2 Compositionality and the Semantic Pivot

To develop a finer typology of phrasemes, the notion of compositionality of linguistic signs is needed. Recall (for instance, Mel'čuk 1982: 40-41) that a linguistic sign **s** is a triplet**s** = ⟨ 's' ; /s/ ; Σ ⟩, where:

's' is the signified, or informational content, most often a linguistic meaning;

/s/ is the signifier, or a physical perceptible signal, most often a string of phonemes (or characters);

Σ is the syntactics, or a set of data specifying the cooccurrence of **s** with other signs.

For instance, the noun AIRCRAFT is represented as a linguistic sign like this:

⟨ 'vehicle designed to fly…'-SG/PL;[2]/έ°rkræft/; Σ= noun, countable, Lexical Functions: *land*$_{(V)}$,*take off*,

crew, ...⟩

Simple signs are combined into complex signs by the operation of linguistic union ⊕. For a particular language **L**, this operation is represented as a set of linguistic rules that tell us how signs must be united:

—The signifieds are united by putting the SemR of an argument into the corresponding argumental position of the corresponding predicate; for instance, 'quite(...)' ⊕'baby'⊕ 'sleep(...)' = 'quiet(sleep(baby))'.

—The signifiers are united by juxtaposing the strings of phonemes and applying all necessary morphological operations; for instance, *deploy*⊕*re-*⊕*-ed = redeployed*.

—The syntactics are united by retaining the combinatorial data valid for the resulting complex sign; for instance, L$_{(V)}$⊕ -er$_{(N)}$ = L-er$_{(N)}$, as in *sing+er*.

These are rather general rules; but the operation ⊕covers all rules necessary for appropriate combination of signs of **L**, including all syntactic and morphological rules, etc.

DEFINITION 5: COMPOSITIONAL COMPLEX LINGUISTIC SIGN

‖ A complex linguistic sign **AB** consisting of the signs **A** and **B**is compositional iff **AB** = **A**⊕**B**.

This means that, for the complex sign **AB** = ⟨ 'AB' ; /AB/ ; Σ$_{AB}$⟩, its signified 'AB' must be equal to 'A' ⊕ 'B', its signifier /AB/, to /A/ ⊕ /B/ and its syntactics Σ$_{AB}$, to Σ$_A$⊕Σ$_B$.

---

[2] This notation indicates that the meanings of the grammemes SINGULAR and PLURAL belong to the signified of the stem of the lexeme AIRCRAFT: *This aircraft is... vs. These aircraft are...*

From Definition 5it follows that compositionality is an absolute notion, which does not admit degrees: a complex sign is compositional or not. Compositionality concerns the three components of the sign independently; in this paper I will consider only the compositionality of signifieds, i.e., the semantic compositionality.

A free utterance is necessarily compositional: it is only thanks to this property that linguistic communication is possible. To master language **L** means to have in the brain a sufficient number of simple signs of **L** and the rules of the operation $\oplus$ for **L**.

The selection of lexical units happens on the paradigmatic axis of language while their combination involves the syntagmatic axis. Taking into account the two axes of speech production guarantees that our characterization of phrasemes is exhaustive.

DEFINITION 6: SEMANTIC PIVOT (OF A MEANING)

Let there be meaning 'σ' that is divided into two parts, 'σ_1' and 'σ_2' ('σ' = 'σ_1' $\oplus$ 'σ_2').

The part 'σ_1' of meaning 'σ' is called its semantic pivot iff the other part 'σ_2' is a predicate of which 'σ_1' is the argument: 'σ' = 'σ_2'('σ_1').

The semantic pivot of meaning 'σ' is logically different from the communicatively dominant component of 'σ', which is the minimal paraphrase of 'σ' (Mel'čuk 2001: 29–31). Thus, in the meaning of the phraseme *take a shower* 'wash oneself under a shower' the semantic pivot is 'shower', while the communicatively dominant component is 'wash'. (The semantic pivot will be identified in the examples by shading.) Note that:

1) The semantic pivot of a multi-word expression **E** does not have to coincide with the lexical meaning of one of **E**'s components. Thus, in the phrase *private eye* 'private detective' the semantic pivot 'detective' is not lexicalized as such.

2) In many cases, the semantic pivot of an expression coincides with its communicatively dominant component, but this is not a reason to confound them.

The notion of semantic pivot will be used to sharpen the typology of phrasemes.

2.3 Major Classes of Phrasemes

Crossing the two dimensions—lexical *vs.* semantic-lexical constraints and being compositional *vs.* non-compositional—gives four logically possible major classes of phrasemes:

Nature of constraints \ Compositionality of phrasemes	non-compositional	compositional
lexical	IDIOMS	COLLOCATIONS
semantic-lexical	~~impossible case~~	CLICHÉS

Figure 1 The Three Major Classes of Phrasemes

However, one of these classes—semantic-lexical non-compositional phrasemes—cannot exist: if a non-free (= phraseologized) utterance **U**phr is non-compositional, then it has, by definition, a "holistic" meaning that is associated with **U**phr as a whole; therefore, this meaning cannot be constructed by the speaker for the occasion; therefore, it does not make sense to talk about constrained/non-constrained character of its construction.

As a result, a natural language has just three major classes of phrasemes: idioms, collocations and clichés.

2.3.1 Idioms

DEFINITION 7: IDIOM

A lexical phraseme is an idiom iff it is non-compositional.

An idiom is indicated in print by elevated half-brackets: ⌐ … ⌐.

Examples: ⌐*cheek by jowl*⌐ 'in close association', ⌐*The game is up*⌐ 'your deceit is exposed', ⌐*come to* [N_X's] *senses*⌐ 'become conscious again', ⌐*put* [N_Y] *on the map*⌐ 'make the place Y well-known', ⌐*bull*

session˥ 'long informal talk on a subject by a group of people', ˥*game of chicken*˥ 'showdown between two opponents where none is disposed to yield and both lose if they push the conflict to the end' Rus. ˥*ostat´sja s nosom*˥ lit. 'remain with nose' ≈ 'X gets nothing in a situation where X is supposed to obtain something he wanted', ˥*sinij čulok*˥ 'bluestocking', etc.

An idiom can be characterized by the degree of its transparence/opacity: the degree to which its meaning includes the meanings of its components. Three types of idioms can be distinguished in such a way: full idioms, semi-idioms and quasi-idioms. All of them are non-compositional, but the degree of their transparence varies.

DEFINITION 8: FULL IDIOM

> An idiom **AB** is a full idiom iff its meaning does not include the meaning of any of its lexical components:
>
> $$\text{'AB'} \not\supset \text{'A'} \text{ and } \text{'AB'} \not\supset \text{'B'}.$$

Examples: ˥*put* [N_Y] *through its paces*˥ 'to test Y thoroughly', ˥*go ballistic*˥ 'suddenly become very angry', ˥*by heart*˥ 'remembering verbatim', ˥*bone of contention*˥ 'reason for quarrels or fights', Rus. ˥*jabloko razdora*˥ lit. 'apple of discord' = 'bone of contention', ˥*delat´ nogi*˥ lit. 'do legs' = 'flee', ˥*polezt´ v butylku*˥ lit. 'try.to.get into bottle' = 'stubbornly insist on something in a dangerous situation', etc.

DEFINITION 9: SEMI-IDIOM

> An idiom **AB** is a semi-idiom iff its meaning 1) includes the meaning of one of its lexical components, but not as its semantic pivot, 2) does not include the meaning of the other component and 3) includes an additional meaning 'C' as its semantic pivot: 'AB' ⊃ 'A', **and** 'AB' ⊅ 'B', **and** 'AB' ⊃ 'C'.

Thus, a semi-idiom is semi-transparent (or semi-opaque, depending on whether you are an optimist or a pessimist).

Examples: ˥*private eye*˥ 'private detective', ˥*sea anemone*˥ 'predatory polyp dwelling in the sea', Rus. ˥*mozolit´ glaza*˥ lit. 'make.corns.on Y's eyes' = 'be too often or for too long before Y's eyes'.

DEFINITION 10: QUASI-IDIOM (= WEAK IDIOM)

> An idiom **AB** is a quasi-idiom, or weak idiom iff its meaning 1) includes the meaning of both of its lexical components, neither as the semantic pivot, and 2) includes an additional meaning 'C' as its semantic pivot:
>
> $$\text{'AB'} \supset \text{'A', } \textbf{and } \text{'AB'} \supset \text{'B', } \textbf{and } \text{'AB'} \supset \text{'C'}.$$

Examples: ˥*start a family*˥ '[conceive the first child with one's spouse, [thereby] starting a full-fledged family';

˥*barbed wire*˥ '[artifact designed to make obstacles with and constituted by] wire with barbs [fixed on it in small regular intervals]'; ˥*lightning rod*˥ '[artifact designed to protect buildings against lightning—a [metallic] rod [fixed at the upper part of the building and grounded]'.

2.3.2 Collocations

DEFINITION 11: COLLOCATION

> A lexical phraseme is a collocation iff it is compositional.

Examples: *heavy* ACCENT, Rus. *sil´nyj* AKCENT lit. 'strong accent', Fr. ACCENT *à couper au couteau* lit. 'accent to cut with. the knife'; *soundly* ASLEEP, Rus. SPAT´ *glubokim snom* lit. 'asleep with deep sleep'; ARMED *to the teeth*; *fasten* ⟨= *buckle up*⟩ *the* SEATBELT, Rus. *zastegnut´* ˥PRIVJAZNOJREMEN˥ lit. 'button.up seatbelt'; *leap* YEAR, Rus. *visokosnyj* GOD (the adjective VISOKOSNYJ is used only with GOD 'year').

A collocation consists of a base, lexical unit chosen freely by the speaker (shown in SMALL CAPS), and of a collocate, lexical unit chosen to express the intended meaning as a function of the base. A collocation is semantically compositional, since its meaning is divisible into two parts such that the first one corresponds to the base and the second to the collocate. The meaning of the base is always the semantic pivot of the collocation. For more on collocations in the Meaning-Text framework, see Mel'čuk 2003a, 2003b and 2004.

This should not be understood as implying that a collocate—taken as such, outside the collocation—necessarily has the meaning it expresses within the collocation. Thus, in the collocation

sit for an exam 'undergo an exam', the verb SIT expresses the meaning 'undergo'; but in an English dictionary, the verb SIT does not have to carry this meaning: 'undergo' is not its inherent, but context-imposed signified.

In English, you **make** *a decision*, and in British English, you can also **take** it. For the same thing, French says **prendre** [= 'take'] *une décision*, German—*eine Entscheidung***treffen/fällen**[= 'meet/fell'], Russian—**prinjat´**[= 'accept'] *rešenie*, Turkish—*karar* **vermek** [= 'give'], Polish—**podjąć** [= 'take up'] *decyzję*, Serbian—**doneti** [= 'bring'] *odluku*, and Korean—*gyeoljeongeul* **haerida**⟨**naerida**⟩ [= 'do ⟨take/put down⟩']. This clearly shows that boldfaced verbs are selected as a function of the noun meaning 'decision'. If instead of DÉCISION a French speaker uses CHOIX 'choice' (*Jean a pris la décision de rester* lit. 'Jean has taken the decision to stay' E*Jean a ... le choix de rester* 'Jean has ... the choice to stay'), he has to say FAIRE 'make' rather than PRENDRE 'take': *Jean a fait ⟨*a pris⟩ le choix de rester* 'Jean has made the choice to stay'.

Collocations are extremely variegated and very numerous in any particular language (in the millions). Two major types are distinguished: standard and non-standard collocations.

Let there be a collocation "Base–**f**–Collocate," where the collocate expresses the meaning '**f**' as a function of the base: roughly, '**f**'(Base) = Collocate.

DEFINITION 12: STANDARD COLLOCATION

A collocation "Base–**f**–Collocate" is standard in language **L** iff the meaning'**f**' meets simultaneously the following two conditions:
1. '**f**' is applicable to many different bases and specifies many different collocates.
2. L_i('**f**') —that is, the lexical units that express '**f**'—participate in Deep-Syntactic paraphrasing.

In other words, **L** has many collocations where the relation between the base and the collocate is **f** (here,"many" means at least several dozen or, better, hundreds) and which are actively exploited in DSynt-paraphrases.

Examples: so-called support, or light, verbs, such as those used with $ATTACK_{(N)}$in *On Friday, the rebels* ***carried out*** *an ATTACK on the port*. E*On Friday, the port* **came** *under the rebels' ATTACK*.E*The (last) rebels' ATTACK on the port* **came** *on Friday*. E*On Friday, the rebels attacked the port*. Or else intensifiers, as in *The rebels ATTACKED the port* ***from all sides***. E*The rebels launched a* ***large-scale*** *ATTACK on the port*.

DEFINITION 13: NON-STANDARD COLLOCATION

A collocation "Base–**f**–Collocate" is non-standard, in language **L** iff the meaning'**f**' is not applicable to many different bases and does not specify many different collocates (in the minimal case, it can apply just to one base and produce just one collocate) and does not participate in the DSynt-paraphrasing.

Examples: *leap YEAR*, where '**f**' = 'having 366 days'; *black COFFEE*, '**f**' = 'with no diary product added'; *LAUGHin* [N_X's] *sleeve*, '**f**' = 'trying to hide the fact of…';*spiked HEELS,*'**f**'='long and thin'; etc.

The meaning '**f**' corresponds to a Lexical Function, see below, **3.2.1**.

2.3.3 Clichés

DEFINITION 14: CLICHÉ

A semantic-lexical phraseme is a cliché.

Examples: *If you've seen one, you've seen them all*; *Happy birthday to you!*; *no matter what*; *We all make mistakes*; *Will you marry me?*; etc.

A cliché is a compositional expression used for a complex meaning 'σ' that the language prescribes to use instead of an equivalent one 'σ´'. Thus, in English we ask *What is your name?* and answer *My name is* [N] or *I am* [N]; Russians say *Kak vas zovut?* lit. 'How do they call you?' and *Menja zovut* [N] 'They call me [N]'. The sentences *Kak vaše imja?* and *Ja* [N], the literal renderings of the English expressions, are fully understandable and grammatical, but not quite standard.

A cliché is characterized by itslexical anchor(or anchors), which is the lexeme(s) whose meaning(s) control(s) the use of the cliché: *What is your name?* and *Kak vas zovut?* 'What do the call you?' have NAME/IMJA as their anchor. (As we see in *Kak vas zovut?*, a cliché's lexical anchor does not have to be explicitly present in the cliché.) In a dictionary, clichés are described under their lexical anchors.

The clichés of a language fall into two major subclasses depending on whether the meaning of the cliché under consideration is constrained or not by the extralinguistic situation in which the cliché is felicitously used. If it is the case, we have pragmatically constrained clichés, or pragmatemes; otherwise, the cliché is pragmatically non-constrained.

Pragmatically non-constrained clichés include several subclasses, of which I will mention two:

• Complex proper names, which have just one referent each : *The Old Testament* [**name of the first part of the Bible**], *Farewell to Arms* [**name of a novel by E. Hemmingway**], *The Moonlight Sonata* [**name of a piano sonata by Beethoven**], *City of Lights* [**nickname of Paris**], *Eternal City* [**nickname of Rome**] (Bosredon 2011).

• Proverbs, which state «eternal» truths: *A friend in need is a friend indeed*; *Fortune helps the brave; The end justifies the means.* (**NB**: Not every proverb is a cliché; some of them are not compositional—that is, idioms, such as *Every Jack has his Jill* or *When in Rome, do as the Romans do*.)

Pragmatically constrained clichés can be illustrated by the case of warnings printed on containers of perishable foods. English says *Best before…*, while in Russian, this will be *Srok godnosti…* lit. 'Term of.validity…', in Polish, *Najlepiej spożyć...*lit. 'The.best [is] to.consume…', in French, *À consommer avant...*lit. 'To consume before…', and in German, *Mindestens haltbar bis...*lit. 'At.least keepable till...'. All these expressions are fully constrained but compositional—that is, they are clichés. However, this is a particular type of cliché, since they are used only in a particular situation: as an official inscription [**on a container of perishable food fabricated for sale**]. The boldfaced indication in brackets is a pragmatic constraint on this cliché.

Pragmatic constraints are in principle applicable to any type of lexical expression—not only to phrasemes but to lexemes as well; here are examples:

pragmatically constrained

idioms	: *Break a leg!*'Good luck!' [**to a performer who is going on stage**]
	(*I*) *copy that!* 'I understood you' [**in a radio communication**]
collocations	: *Wet paint* [**on a sign**]
clichés	: *No parking* [**on a sign**]
	Enjoy your meal! [**to people starting a meal**]
lexemes	: *Roger!* 'I understood you' [**in a radio communication**]
	Pol. *Smacznego!* lit. 'Of.tasty!' ≈ 'May your food be tasty!' [**to people starting a meal**]

However, among pragmatically constrained lexical expressions, clichés occupy a special place: a crushing majority of pragmatically constrained expressions are clichés. Therefore, it is convenient to give pragmatically constrained clichés a special name: pragmatemes.

DEFINITION 13: PRAGMATEME

║ A pragmatically constrained cliché is a pragmateme.

Examples: *Hold the line!*[**in a telephone conversation**], *Watch your step!* [**on a sign**], *X—all you can eat* [**on a sign in a restaurant**], *Emphasis mine* [**after a quotation in a written text**], *Return to sender* [**on a postal sending**], *Who's there?* [**answering a knock on the door**], *Enjoy your meal!* vs. Fr. *Bon appetit !*lit. 'Good appetite!' [**to people starting a meal**], etc.

Let it be emphasized: such a cliché as *What's your name?* is not a pragmateme: it can be used in any situation; likewise, *on top* [*of*Y], *Sorry to keep you waiting*, *I am in the mood* [*for* Y], *Would you mind* [Y-*ing?*], *It's a proven fact*, etc.

Here are a few Turkish pragmatemes (from Öztürk Kasar 2009: 168-175; they are given in a literal translation):

MARRIAGE
«I wishyou both to age together happily» : *May God allow you to age on the same bolster!* [**a wedding guest to the newlyweds at a marriage ceremony**]

WORK
«I wish you that your work be successful» : *May this come easily!* [**a passerby to somebody who is working**]

MEDICINE [drug]
«I wish you that this cures you» : *May this be beneficial!*[**to somebody who is taking a medicine**]

SNEEZE
«`I wish you everything good`»   :*May you live long!* **[to a person who sneezes]**

SHOWER, BATH, SHAVING
«`I wish you everything good`»   : *May there be plenty of healths!* **[to somebody who has just**
showered,   **bathedor shaved]**
SLEEP
«`I wish you to sleep well`»   : *May God give much comfort!* **[to somebody who is going**
to bed]
Note that in Spanish, in this situation you wish *¡Buenas noches !*lit. 'Good nights!', in Russian—
Spokojnoj noči! 'Quiet night!', and in Serbian—*Laku noć!* 'Easy night!'

However, the Turkish clichés *May God give strength to your arms and your legs!* and *May your
eyesburst!* are not pragmatemes, since they can be used in any circumstances to express the following
conceptual content:
THANK
«`I thank you for physical efforts`
`with which you have just helped me`» : *May God give strength to your arms and your legs!*
DAMN
«`I damn you`»   :*May your eyes burst!*

As noted by A. Polguère(orally), many—perhaps even the majority of—pragmatemes are
independent utterances, i.e., sentences. Moreover, their meaning includes most often the meaning 'I' (=
'the Speaker'); therefore, they are non-descriptive lexical units, more precisely signalatives (Mel'čuk
2001: 353-356).

2.5 General Typology of Phrasemes

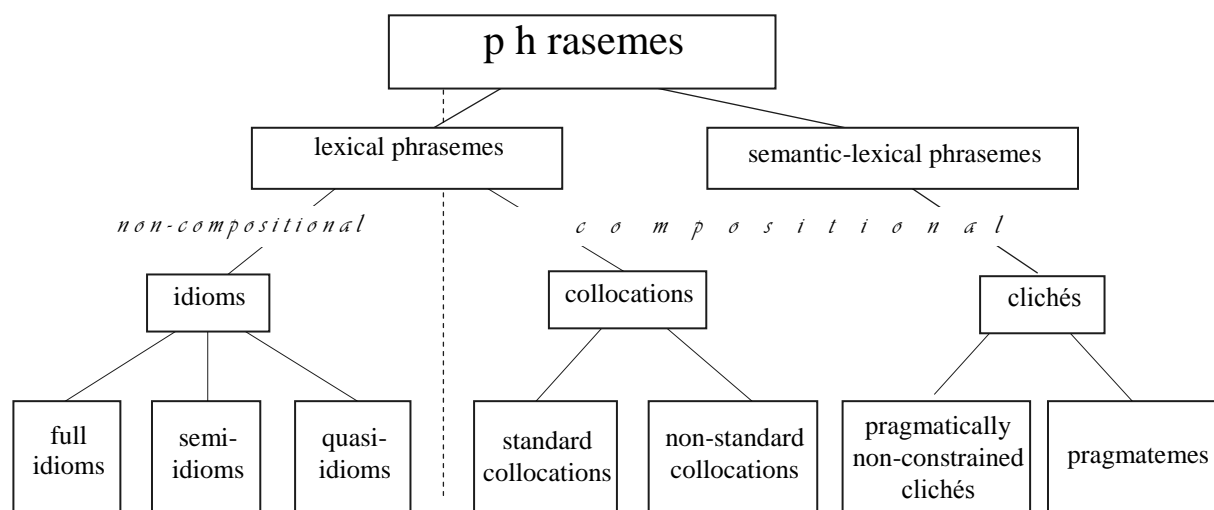I can now present all the major subclasses of phrasemes and their taxonomy:

Figure 2 General Typology of Phrasemes

3. Phraseology in the Dictionary

The dictionary considered here is the *Explanatory Combinatorial Dictionary* [= ECD]; its
principles, its structure and basic underlying notions are taken for granted (see Mel'čuk1974, Mel'čuk
and Zholkovsky 1984, Mel'čuk1988b, Mel'čuk *et al.*1984–1999, Mel'čuk *et al.* 1995, Mel'čuk 2006,
Mel'čuk and Polguère 2007). I will discuss only the lexicographic presentation of phrasemes.

3.1 Lexicographic presentation of non-compositional phrasemes (idioms)

An idiom is a lexical unit, just as a lexeme is one. Idioms are, then, described in an ECD the same way as lexemes: each has its proper dictionary entry, featuring the same structure as a lexeme entry, with one important difference: since an idiom is a multiword utterance, it is supplied with its Surface-Syntactic structure.Cf.:

ˊPULL THE WOOL OVER [N_Y's] EYESˋ,full verbal idiom

Definition

X pulls the wool over Y's eyes : 'X tries to deceive Y in order to hide from Y what X is really doing'.

Surface-Syntactic Structure

```
        ┌── oblique-objectival ──────────────────────┐
        │                                            ↓
PULL–direct-objectival→WOOL_SG–determinative→THE  OVER–prepositional→EYE_PL
```

Government Pattern

X ⇔ I	Y ⇔ II	
1. N	1. *of*N	| THE←**determ**–EYES–**attrib**→OF N
	2. N's	| EYES–**possessive**→N's
	3. A_(poss)(N)	| EYES–**determ**→A_(poss)(N)

Don't pull the wool over foreigners' eyes! | He tried to pull the wool over my⟨John's⟩eyes.

The number of idioms in a particular language is probably around 10 000; thus, the English idiom dictionary Cowie*et al*. 1993 contains about 7 000 idiomatic expressions, and the French idiom dictionary Rey and Chantreau 1993, about 9 000; an excellent Russian-English dictionary of idioms (Lubensky 1995) presents some 13 000 idiomatic units.

3.2 Lexicographic presentation of compositional phrasemes (collocations and clichés)

Compositional phrasemes—collocations and clichés—are not lexical units; they do not have their own dictionary articles and are described in the articles of their bases/anchors. For instance, the collocation *ARMED to the teeth* does not have a separate entry, but appears under ARMED; Rus. *vypisat´ ček*lit. 'write out a check' = 'draw a check' is given under ČEK; and so forth. The same is true of clichés.

3.2.1 Lexicographic presentation of collocations

The number of collocations in languages of *Standard Average European* type is very high: no less than ten times the number of lexemes, which means millions. Therefore, the lexicographic description of collocations requires a special formal apparatus that would allow for their elegant systematic presentation in the dictionary and, at the same time, facilitate automatic processing. Such an apparatus is the system of lexical functions [= LFs]. It is of course impossible to introduce here the notion of LF or offer the reader a detailed review thereof (see Žolkovskij and Mel'čuk 1967, Mel'čuk 1974: 78-109, 1982, 1996, 2003a, b, Kahane 2003b, Kahane and Polguère 2001, Wanner (ed.) 1996). I will limit myself to a few examples of standard and non-standardLFs, used for the description of, respectively, standard and non-standard collocations—in order to show afterwards how LFs can be exploited in NLP.

Standard collocations described by standard LFs

A standard LF $\mathbf{f_{stand}}$describes a family of standard collocations where the semantic relation between the base and the collocate is institutionalized in the language; $\mathbf{f_{stand}}$ specifies this relation simply by its name.

• **Verbal Standard LFs**

 – Support verbs

	RESPONSIBILITY X's ~ concerning Y	CARE X's ~ concerning Y	ACCUSATION X's ~ of Y of Z	AID X's ~ to Y
$Oper_1$	*carry* [ART ~]	*give* [~ *to* N_Y]	*level* [ART ~ *at* N_Y]	*come* [*to the* ~($\rightarrow$*of* N_Y)]
$Func_2$	~ *includes* [N_Y]	~ *is aimed* [*at* N_Y]	~ *weighs* [*on* N_Y]	~ *comes* [*to* N_Y*from* N_X]
$Labor_{12}$	———	*surround* [N_Y*with* ~]	*bring* [N_Y*under* ~]	*support* [N_Y*with* N_X*'s* ~]

 – Realization verbs

	PRIZE X's ~ to Y for Z	DOCTOR ~ X of Y	TRAP X's ~ for Y	ASPHALT ~ used by X on Y
$Real_2$	*win* [ART ~]	*see* [ART ~]	*fall* [*into* ART ~]	———
$Fact_2$	~ *goes* [*to* N_Y]	~ *sees* [N_Y]	~ *catches* [N_Y]	~ *covers* [N_Y]
$Labreal_{12}$	[N_Y *of* ART ~]	———	*catch* [N_Y *with* ART ~]	*cover* [N_Y*with* ~]

• **Adjectival Standard LFs (intensifiers/mitigators)**

	WET	DRUNK	BREATHE	ROLE	LAUGHTER
Magn	~ *to the bone*	*dead, stone~, ~ as a skunk//smashed*[3]	~ *heavily*	*important < crucial < critical*	*hysterical, side-splitting ~; uncontrollable ~*

	WOUND	DRUNK	BREATHE	ROLE	LAUGHTER
AntiMagn	*light ~ //scratch*	*slightly ~ //tipsy*	~ *lightly*	*small, secondary ~*	*muffled ~*

Non-standard collocations described by non-standard LFs

A non-standard LF **f_non-stand** describes a non-standard collocation where the semantic relation between the base and the collocate is not institutionalized in the language; to specify this relation, **f_non-stand** must be described in the same metalanguage as that used for lexicographic definitions:

```
with no diary products added(COFFEE)      : black [~]
drinking up the contents of the glass
               at one go(DRINK)            : [~] bottoms up
used too much(EXAMPLE)                     : hackneyed [~]
```

To illustrate the lexicographic description of collocations, here is a lexical entry for the noun BATTLE**1**(as in *Fierce battles are raging within 25 miles of Tangkin*):

BATTLE**1**, noun, countable

Definition

 Battle between X and Y for Z: 'Armed confrontation between group X and group Y for Z'.

Government Pattern

X ⟺I	Y ⟺II	Z ⟺III
1. *of* N 2. N's 3. A_poss(N)	1. *with* N 2. *against* N	1. *for* N 2. *over* N 3. *to* V_{INF}
4. *between* N *and* N		

a battle of Philippino guerrillas/their battle with ⟨= *against*⟩ *the Japanese*;
battles between Palestinian factions for ⟨= *over*⟩ *the border control* ⟨= *to control the border*⟩

Lexical Functions

$Syn_{\cap}$	: *engagement< combat; action; fight; firefight*
$V_{0\cap}$	: *battle*$_V$
$S_{1/2}$	: *combatant; adversary, enemy*
S_{loc}	: *battlefield, battleground*

---

[3] The symbol "//" indicates that all LF values that follow it are fusioned—that is, each one expresses together rhe meaning of the base and its own; thus smahed means 'very drunk'. Another example: Magn(*fog*) = *dense, thick//pea-soup*, where *pea-soup* means 'dense fog'.

Mult	: //*hostilities; war*
Loc_{in}	: *in* [~]
Ver	: *winning* [*a winning ~*]
AntiVer	: *losing*
Magn	: *pitched*; *ferocious, fierce, grueling, intense, rude, violent*; *bloody<murderous* [*huge losses*] *<mortal*; *royal* / postposed
AntiMagn	: //*skirmish*
$Oper_1$	: *fight* [ART ~]; *be, be locked, be engaged* [*in* ART ~ →*against* N_Y]
$IncepOper_1$	: *engage* [ART ~]
$ContOper_1$	: *continue* [ART ~]
$FinOper_1$	: *stop* [ART ~]
$CausOper_1$	: *send* [N_X *in* ~]
[$Magn+Func_0$]	: *rages*
$Func_{1+2}$	: *opposes* [N_X *to* N_Y; N_X *and* N_Y], *pits* [N_X *against* N_Y]
$nonFunc_0$	: //*guns are silent*
$IncepFunc_0$	: *breaks out*
$IncepLabor_{12}$	: //*engage* [N]
$Real_1$	: *win* [ART ~]
$AntiReal_1$	: *lose* [ART ~]
Son	: *rumbles*
X and Y being individuals in physical contact	: *close, hand-to-hand* [~]
X and Y being ships	: *naval*
X and Y being planes	: *aerial, air* [~] //*dogfight*
X and Y being of quite unequal forces	: *unequal*; *see-saw* [~]
X and Y being of rather equal forces	: *tight*
more difficult for X	: *up-hill* [~]
X's first B.	: //*baptism of fire*
X begins to participate in B.	: *joins* [the ~]

3.2.2 Clichés

Being compositional, the meaning of a cliché need not be indicated in the dictionary. (Although in a pedagogical dictionary it might be.) What has to be indicated is the conceptual content to which a given cliché corresponds. Thus, for the content «I want you to tell me your name» (a conceptual representation is printed in «Monaco» in double quotes and is underlined), English says *What's your name?*, while in Russian the corresponding expression is *Kak vas zovut?* lit. 'How do they call you?', which has a different meaning—that is, a different semantic representation.

Clichés (including pragmatemes) and pragmatically constrained lexemes are presented in the articles of their anchor(s) in a way similar to the presentation of non-standard LFs, except that instead of the description of their meaning, the dictionary gives a description of their conceptual content. For instance:

PAY(V)

«without having to pay»	: *free of charge*

LATE

«Even if this is happening later than needed, it is OK»	: *Better late than never.*

PUBLISH

«[the text in question] is supposed to be published shortly»	: *forthcoming* [**in a bibliographic reference**]

DOG

«There is an aggressive dog on premises»	: *Beware of dog* [**on a sign**]

3.3 New Type of General Dictionary

The proposed lexicographic description of phrasemes entails a new concept of general dictionary. Traditionally, a dictionary is a huge list of words supplied with all types of necessary or useful information. But if the dictionary also has to store and systematically describe all set phrases, which outnumber words at least 10 to 1, it ceases to be a dictionary of words: it becomes a dictionary of phrases or, more precisely, of minimal utterances—that is, utterances that cannot be fully represented in the lexicon in terms of other, smaller utterances and rules for their combinations. The idea that what is actually needed is a dictionary of multiword expressions was put forth in a concise article Becker 1975; coming form a different direction (language teaching), Nattinger 1980 also underscored the necessity of a "phrasal" dictionary. Bogusławski and Wawrzyńczyk 1993 and Bogusławski and Danielewiczowa 2005 constitute an excellent illustration of what such a dictionary should look like: their dictionary includes idioms, collocations and clichés, but also syntactic constructions (for instance, «$N_X of$ $N_{(period)Y}$»: *book of the year* or *cover girl of the month*). More recently, many dictionaries of idioms and collocations have been published for different languages, but what I am aiming at here is a general dictionary where words and multiword expressions are stored and described together and in parallel. The ECD is intended to be such a dictionary.

4. Phraseology in Natural Language Processing

Idioms and clichés have to be listed in the dictionary, and I have shown how this could be done. From the viewpoint of NLP, they are similar to single words and can be dealt with in the same way as single words are. But the collocations pose a serious problem for automatic processing, in particular for automatic translation, given their number and variety. Lexical Functions offer a reasonable solution.

LFs can be used in NLP—in particular, in automatic translation and text generation—in two ways. On the one hand, LFs ensure correct lexical selection when translating the collocations of the type (English-Russian) **grave** *illness ~ **tjažëlaja** bolezn´* lit. 'heavy illness', **put** [N_Y] *in danger ~ **podvergat´*** [N_Y] *opasnosti*$_{DAT}$lit. 'submit [N_Y] to danger' or **take**flight ~ **obratit´sja** *v begstvo* lit. 'turn.oneself in flight'. All such "exotic" equivalences are covered by pairs of ECD-type dictionaries; LFs, being linguistically universal, play the role of an interlingua.

On the other hand, LFs underlie paraphrasing at the deep-syntactic level. This paraphrasing is necessary to resolve syntactic mismatches between the input and output sentences S_{source} and S_{target}, such mismatches being extremely frequent in parallel texts. Only paraphrasing can allow a translation system to construct an acceptable deep-syntactic structure for the output sentence S_{target} in the case of a serious mismatch between the vocabulary of S_{target} and its DSyntS, "inherited" from S_{source}. Thus, consider the sentence 0a) and its translations in Russian and French 0b-c):

(1) a. *She competes internationally.*

b. Rus.*Ona učastvuet v meždunarodnyx sorevnovanijax*
'She participates in international competitions'.

c. Fr.*Elle participe à des compétitions internationales* [idem].

The verb meaning '[*to*] compete' (in the needed sense) does not exist in Russian or French. However, a verb V can always be paraphrased by the deverbal noun $S_0(V)$ and one of its support verbs:$V \Leftrightarrow S_0(V) \leftarrow$II$\rightarrow$Oper$_1(S_0(V))$. This formula describes all equivalences of the type *compete⇔participate in competition(s), trip⇔take a trip, look⇔have a look*, etc.; the noun *competition* has direct equivalents in Russian and French.

For a universal DSynt-paraphrasing system, see Žolkovskij and Mel'čuk 1967, Mel'čuk 1974: 141-176, 1988c, 1992 and 2004, and Milićević 2007; Mel'čuk and Wanner 2006 deals specifically with the problem of syntactic mismatches in machine translation; the use of LFs in text generation is described in Iordanskaja *et al*. 1996 and Lareau and Wanner 2007. The paraphrasing system for Russian has been implemented and tested in a series of computer experiments: Apresjan and Cinman 1998 and 2002.

I will now present three examples of translation that are difficult because of the collocations involved, in order to show how the use of LFs ensures good results.

Example 1: The verb STRIKE
Take the sentence in 0a) and its closest (= most literal) Russian translation in 0b):

(2) a. *The book thief struck again*.
b. *Knižnyj vor snova soveršil kražu* lit. 'Book thief again committed theft'.

It is absolutely out of the question to translate STRIKE in this sentence as UDARJAT′ 'strike':the result would be incomprehensible. The correct choice is the collocation *soveršit′ kražu* 'commit a theft'. But where and how do we establish the equivalence *strike ≡ soveršit′ kražu*? In different contexts, the verb STRIKE has lots of other equivalents in Russian:

(3) a. *The hurricane **struck** the island again.* ≡
   *Uragan snova **obrušilsja** na ostrov* lit. 'Hurricane again fell.down on island'.

   b. *The bullet **struck** him in the shoulder.* ≡
   *Pulja **popala** emu v plečo* lit. 'Bullet hit to.him in shoulder'.

   c. *A suicide bomber **struck** in the market.* ≡
   *Terrorist-smertnik **podorval sebja** na rynke* lit. 'Suicide bomber exploded himself in market'.

And so forth.

However, if we think of LFs, the answer comes immediately: all the illustrated uses of STRIKE are values of LF $Fact_0$: $Fact_0$(L) ≈ 'perform the action that (the denotation of) L is supposed to perform in conformity with its nature'. A Russian ECD must have:

$Fact_0$(VOR 'thief')	: *krast′* 'steal', *soveršat′ kražu* 'commit a theft'
$Fact_0$(URAGAN 'hurricane')	: *obrušit′sja* [*na* N] 'strike [N]'
$Fact_0$(PULJA 'bullet')	: *popast′* 'hit'
$Fact_0$(TERRORIST-SMERT-NIK 'suicide bomber')	: *podorvat′ sebja* 'explode oneself'

An English ECD gives the same indications for the above uses of STRIKE: $Fact_0$(THIEF) = *strike*, etc.

Given the regular translation equivalents THIEF ≡ VOR, HURRICANE ≡ URAGAN, etc., the equivalencies between the corresponding values of their $Fact_0$ are obtained automatically.

Example 2: The French noun APPOINT 'exact sum paid by X to Y for Z such that Y does not have to give X any change' (= *exact change*)

A sign in a French bus shown in (4a) could be translated into Russian as (4b):

(4) a. *Merci de faire l'appoint* lit. 'Thank.you for doing the exact.change'.

   b. *Platite za proezd bez sdači* lit. 'Pay for transportation without change'.

Such an equivalent is produced, using a pair of dictionaries of the ECD type, in five steps.

• *Merci de Y* is a pragmatically constrained lexeme that must be described in a French ECD as a non-standard LF under PRIER 'ask':

   «Authorities ask you to Y»   : *Merci* [*de* V('Y')$_{INF}$] [**on a sign**]

• PRIER has a regular Russian equivalent PROSIT′ 'ask'.

Under PROSIT′, the Russian ECD has the above non-standard LF:

   «Authorities ask you to Y»   : V('Y')$_{IMPÉR.2PL}$ [**on a sign**]

• *Faire* in 0a) is described in the French ECD as $Real_1$ of APPOINT:

   $Real_1$(APPOINT)   : *faire* [*l'~*]

• APPOINT is translated as PLATA BEZ SDAČI lit. 'sum paid by X to Y for Z such that Y does not have to give X any change'.

• $Real_1$(PLATA 'the sum paid')   : //*platit′* 'pay'

These five steps produce *Platite bez sdači* lit. 'Pay without change'. But Russian also requires the indication of the thing paid for: *platit′ za čto?* 'pay for what?' – *za proezd* 'for transportation'. This indication can be extracted from general knowledge about the situation in which the relevant utterance is made: if the sign is placed in a public transportation vehicle, you have to add *za proezd*; if it is hung on a ticket office, *za bilet* 'for ticket' is a must; if it is over the counter of a diner, it will read *za obed* 'for lunch'.

There is another way, perhaps even simpler, to establish the equivalence in question; namely, *faire l'appoint* can be described as a non-standard LF of PAYER:

   giving the exact sum due, so that
   Y does not have to give the change to X: //*faire l'appoint*

The corresponding non-standard LF in Russian is described under PLATIT′ 'payer':

PLATIT′ 'payer'

   giving the exact sum due, so that
   Y does not have to give the change to X   : *bez sdači* lit. 'without change'

The equivalence is then obtained in one step. Nevertheless, I wanted to present multiple paths that could lead to the same result.

Example 3: The verb STAB
The Russian sentence in 0a) must be translated into English as 0b):

(5) a. Rus. *Žertve bylo naneseno pjat´ noževyx ran* lit. 'To.victim were dealt five knife wounds'.

b. *The victim was stabbed five times*.

Here are the steps necessary to obtain this equivalence:

- Rus. *Nanesti* [*ranu*] = `CausFunc`$_2$(RANA'wound'); *noževaja* [*rana*] = `caused by a knife`(RANA)
- Eng. [`caused by a knife` + `CausFunc`$_2$(WOUND)] = //[*to*] *stab*
- A general Deep-Syntactic paraphrasing rule (i stands for a DSynt-actantial relation):

$$V_{support}(L_1)-\mathbf{i} \rightarrow L_1 - \mathbf{ATTR} \rightarrow L_2 \Leftrightarrow V_0(L_1) - \mathbf{ATTR} \rightarrow L_2,$$

that is, a modifier L_2 of the noun L_1 in a collocation "Support Verb + N"becomes a modifier of the verb that is semantically equivalent to this collocation; in our case, FIVE←[STAB WOUND$_{PL}$]⇔FIVE←TIME$_{PL}$←STAB$_{(V)}$. (*Five times* is a surface lexical realization of the meaning 'five' when applied to a verb.)

5. Conclusions

In my view, the five most important points of this presentation are as follows:

1. Phrasemes constitute a very significant part of the lexicon of any language; therefore, they have to be presented in a formal dictionary (of the ECD type) in a systematic way.

2. A dictionary of the ECD type is the key for the automatic production of high quality texts.

3. Such a dictionary must reserve a place of honor for collocations described in terms of Lexical Functions as well as for all other phrasemes (idioms and clichés).

4. Lexical Functions must be exploited in two major respects: for lexical selection and for deep-syntactic paraphrasing.

5. A paraphrasing system must be part of any reliable NLP system.

Acknowledgments

References

Anscombre, Jean-Claude and Mejri, Salah, eds. 2011. *Le figement linguistique : la parole entravée*. Paris:Honoré Champion.

Apresjan, Jurij and Cinman, Leonid. 1998."Perifrazirovanie na komp´jutere [Paraphrasing with Computer]."*Semiotika i informatika* 36: 177-202.

Apresjan, Jurij and Cinman, Leonid. 2002. "Formal´naja model´ perefrazirovanija predloženij dlja sistem pererabotki tekstov na estestvennyx jazykax [A Formal Model of Sentence Paraphrasing for Natural Language Text Processing Systems]."*Russkij jazyk v naučnom osveščenii* No. 2 [= 4]: 102-146.

Bally, Charles. 1909 [1951]. *Traité de stylistique française*. Genève: Georg – Paris: Klincksieck.

Beck, David and Mel'čuk, Igor. 2011. "Morphological Phrasemes and Totonacan Verbal Morphology." *Linguistics* 49(1): 175-228.

Becker,Joseph. 1975. "The Phrasal Lexicon." In *Proceedings of the Workshop on Theoretical Issues in NLP*, Cambridge (MA): MIT Press, p. 70-73.

Bogusławski, Andrzej and Wawrzyńczyk, Jan. 1993. *Polszczyzna, jaką znamy. Nowa sonda słownikowa* [Polish as We Know It. A New Lexicographic Probe]. Warszawa: Uniwersytet Warszawski.

Bogusławski, Andrzej and Danielewiczowa, Magdalena. 2005. *Verba polona abscondita. Sonda słownikowaIII* [Concealed Polish Words. Lexicographic Probe III]. Warszawa: Uniwersytet Warszawski.

Bosredon, Bernard. 2011. "Dénominations monoréférentielles, figement et signalétique." In: Anscombre and Mejri (eds.) 2011: 155-169.

Burger, Harald, Dobrovol´skij, Dmitrij, Kuhn, Peter, and Norrrick, Neal, eds. 2007. *Phraseology: An International Handbook of Contemporary Research: Vols. 1-2*, Berlin: de Gruyter.

Cowie, Antony, ed. 1998.*Phraseology. Theory, Analysis, and Applications*. Oxford: Clarendon Press.

Cowie, Anthony and Howarth, Peter. 1996. "Phraseology—a Select Bibliography."*International Journal of Lexicography* 9(1): 38–51.

Cowie, Anthony, Makin, Ronald and McCaig, Isabel. 1993. *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press.

Everaert, Martin, van der Linden, Erik-Jan, Schenk, André and Schreuder, Rob, eds. 1995. *Idioms. Structural and Psychological Perspectives*. Hillsdale, N.J.—Hove, UK: Lawrence Erlbaum Associates.

Grossmann, Francis and Tutin, Agnès, eds. 2003 *Lescollocations: analyse et traitement.* [= *Travaux et recherches en linguistique appliquée, Série E: Lexicologie et lexicographie*, nº 1]. Amsterdam: De Werelt.

Iordanskaja, Lidija, Kim, Myunghee and Polguère, Alain. 1996. "Some Procedural Problems in the Implementation of Lexical Factions for Text Generation." In: Wanner (ed.) 1996: 279-297.

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

Kahane, Sylvain. 2003b. "Sur le lien entre la définition lexicographique et les fonctions lexicales: une blessure profonde dans le DEC." In: Grossman and Tutin (eds.) 2003: 61-73.

Kahane, Sylvain and Polguère, Alain. 2001. "Formal Foundations of Lexical Functions". In: *Proceedings of "COLLOCATION: Computational Extraction, Analysis and Exploitation", 39th Annual Meeting and 10ᵗʰConference of the European Chapter of the Association for Computational Linguistic*s, Toulouse, 8-15.

Lareau, François and Wanner, Leo. 2007. "Towards a Generic Multilingual Dependency Grammar for Text Generation." In: T. Holloway King and E. Bender (eds.), *Proceedings of the Grammar Engineering Across Frameworks 2007 Workshop. CSLI Publications*, Stanford, CA: Stanford University. See also: http://csli-publications.stanford.edu.

Lubensky, Sophia. 1995. *Russian-English Dictionary ofIdioms*. New York: Random House.

Mel´čuk, Igor. 1974. *Opyt teorii lingvističeskix modelej Smysl - Tekst* [Outline of a Theory of Meaning-Text Type Linguistic Models]. Moskva: Nauka. [Reprint: 1999.]

Mel'čuk, Igor. 1981. "Meaning-Text Models: A Recent Trend in Soviet Linguistics."*Annual Review of Anthropology* 10: 27-62.

Mel'čuk, Igor. 1982. *Towards the Language of Linguistics*. München: Wilhelm Fink.

Mel'čuk, Igor. 1988a. *Dependency Syntax: Theory and Practice*. Albany, N.Y.: State University of New York Press.

Mel'čuk, Igor. 1988b. "Description of Lexical Units in an *Explanatory Combinatorial Dictionary*: Basic Principles and Heuristic Criteria."*International Journal of Lexicography* 1(3): 165-188.

Mel'čuk, Igor. 1988c. "Paraphrase et lexique dans la théorie linguistique Sens-Texte." In: G. Bès and C. Fuchs (eds), *Lexique et paraphrase* [= *Lexique 6*]. Lille: Presses Universitaires de Lille, 13-54.

Mel'čuk, Igor. 1992. "Paraphrase et lexique: Vingt ans après." In: Mel'čuk and *al.* 1992: 9-58.

Mel'čuk, Igor. 1995. "Phrasemes in Language and Phraseology in Linguistics." In: Everaert *et al.* (eds.) 1995: 167-232.

Mel'čuk, Igor. 1996. "Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon." In: L. Wanner, ed., *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins, 37-102.

Mel'čuk, Igor. 1997. *Vers une linguistique Sens-Texte. Leçon inaugurale*. Paris: Collège de France.

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

Mel'čuk, Igor. 2003a. "Collocations dans le dictionnaire." In: Th. Szende (ed.), *Lesécarts culturels dans les dictionnaires bilingues*. Paris: Honoré Champion, 19-64.

Mel'čuk, Igor. 2003b. "Les collocations: definition, rôle et utilité." In: Grossmann and Tutin (eds.) 2003: 23-31.

Mel'čuk, Igor. 2004. "Verbes supports sans peine."*Lingvisticæ Investigationes* 27: 2, 203-217.

Mel'čuk, Igor. 2006. "Explanatory Combinatorial Dictionary." In: G. Sica (ed.), *Open Problems in Linguistics and Lexicography*, 2006, Monza (Italy): Polimetrica Publisher, 225-355. See also:

http://www.polimetrica.com/?p=productsListandsWord=lexicography

Mel'čuk, Igor, Clas, André and Polguère, Alain. 1995. *Introduction à la lexicologie explicative et combinatoire*. Bruxelles: Duculot.

Mel'čuk, Igor and Polguère, Alain. 1991. "Aspects of the Implementation of the Meaning-Text Model for English Text Generation." In: N. Lancashire (ed.), *Research in Humanities 1*, Oxford: Clarendon Press, 204-215.

Mel'čuk, Igor and Polguère, Alain. 2007. *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations semantiques et collocations du français*. Bruxelles: De Boeck.

Mel'čuk, Igor and Wanner, Leo. 2006."Syntactic Mismatches in Machine Translation." *Machine Translation*21: 81-138.

Mel'čuk, Igor andZholkovsky, Alexander. 1984. *Explanatory Combinatorial Dictionary of Modern Russian*,Vienna: Wiener Slawistischer Almanach.

Mel'čuk, Igor, *etal*. 1984, 1988, 1992, 1999. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-semantiques I-IV*. Montréal: Les Presses of l'Université de Montréal.

Milićević, Jasmina. 2007. *La paraphrase. Modélisation de la paraphrase langagière*. Bern etc.: Peter Lang.

Nattinger, James. 1980. "A Lexical Phrase Grammar for ESL."*TESOL Quarterly*14(3): 337-344.

Öztürk Kasar, Sündüz. 2009. *Introduction à la linguistique pour futurs traducteurs*. Istanbul: Multilingual.

Rey, Alain and Chantreau, Sophie. 1993. *Dictionnaire des expressions et locutions*. Paris: Dictionnaires Le ROBERT.

Wanner, Leo, ed. 1996. *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins.

Weinreich, Uriel. 1969. "Problems in the Analysis of Idioms." In: J. Puhvel (ed.), *Substance and Structure of Language*, Berkeley —Los Angeles, CA: University of California Press, 23-81. Reprinted in: U. Weinreich, *On Semantics*, 1980, Philadelphia: University of Pennsylvania Press, 208-264.

Žolkovskij, Aleksandr and Mel´čuk, Igor. 1967. "O semantičeskom sinteze [On Semantic Synthesis]."*Problemy kibernetiki* 19: 177-238.