Игорь Александрович Мельчук
Université de Montréal

Phraseology in the Language, in the Dictionary, and in the Computer

Резюме

Фразема определяется как 'несвободное словосочетание', т.е. как словосочетание, которое не может быть построено для данного смысла только по правилам данного языка. Вводятся понятия *композиционность сложного языкового знака* и *семантическая доминанта языкового выражения*. С их помощью определяются три основные класса фразем: 1) некомпозиционные, т.е. *идиомы*, и композиционные – 2) полуустойчивые, т.е. *коллокации*, и 3) устойчивые, т.е. *клише*; среди последних выделяется подкласс *прагматем* – клише, ограниченные ситуацией употребления. Предлагается классификация идиом на основе степени их семантической мотивированности; дается общая типология фразем. Описывается представление фразем в семантическом *Толково-комбинаторном словаре*. Рассматриваются несколько проблематичных случаев перевода с языка на язык, в которых решение проблем обеспечивается предложенным описание фразем.

1. Introduction

There is no need to insist on the importance of phraseology for linguistic studies: on this point the linguistic community is in agreement. But, curiously and unfortunately, there is no agreement on either the exact content of the notion 'phraseology', on the way phraseological expressions should be described, nor, finally, on how they should be treated in linguistic applications, among others,

in lexicography and Natural Language Processing [= NLP]. In this paper, I will try to deal with these three points: Section 2 proposes a rigorous definition of 'phraseme', a characterization of the major classes of phrasemes and an exhaustive phraseme typology, thus establishing the boundaries of phraseology; Section 3 sketches the fundamentals of the lexicographic description of phrasemes in an Explanatory Combinatorial Dictionary; in Section 4, I consider three examples of difficult cases of automatic translation where the solutions come from the dictionary and are based on the proposed description of one of phraseme classes (namely, collocations). Finally, Section 5 summarizes the most important points of the presentation.

The theoretical framework of the discussion 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 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, "Select Bibliography on Phraseology," http://www.euralex.org/main_h-o.html, 6.04.2010, and the collections Everaert et al. (eds.) 1995; Cowie (ed.) 1998; Burger et al. (eds.) 2007; and Anscombre, 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, called also a *set expression*, *set phrase*, *idiomatic phrase*, *multi-word expression*, sometimes simply *idiom*, etc., is a multiword utterance – that is, a linguistic expression formed by several (at least two) lexemes syntactically linked in a regular way.[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

---

[1] To simplify my task, in this paper I leave aside the phrasemes of the morphological level – that is, the phraseologized combinations of morphs inside a wordform. For this family of phrasemes, or morphophrasemes, see, for instance, Beck, Mel'čuk 2011.

"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 'die of natural causes [said flippantly]'. A phraseological expression, or phraseme, is thus a multiword utterance featuring some unpredictable properties, i.e., a constrained utterance, or a multiword utterance that is not free. Therefore, I have to begin with a definition of free utterance.[2]

Since in what follows only multiword utterances are considered, the modifier multiword will be omitted

DEFINITION 1: FREE UTTERANCE

> An utterance is free if and only if [= iff] each of its lexical components L_i is selected by the speaker in a non-constrained way, which means that L_i is selected strictly for its meaning and in conformity with its linguistic properties but independently of the lexical identity of other components.

This means that, while selecting L_i, the Speaker need not take into account what other particular lexemes constitute the utterance in question.

Corollary: Each lexical component of a free utterance can be replaced by any of its (quasi)synonyms without affecting its meaning and grammaticality. In the phrase *select the word freely*, you can replace any component with its synonym and the meaning is preserved: *choose the lexeme without constraint*.

DEFINITION 2: NON-FREE UTTERANCE = PHRASEME

> An utterance is non-free ⟨= phraseologized⟩ iff at least one of its lexical components L_i is selected by the speaker in a constrained way – that is, as a function of the lexical identity of other component(s).

In a non-free utterance, L_i is selected depending on other particular lexemes building up this utterance.

A non-free utterance is called, as I just said, phraseme.

Corollary: It is not true that every lexical component of a non-free utterance can be replaced by any of its (quasi-)synonyms without affecting its 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. This violation happens on the paradigmatic axis of speech production, as the speaker is looking (in his mental storage) for appropriate lexical units. The selection activity proceeds in two stages:

– First, the speaker has to construct 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].

---

[2] The term *utterance* is used to refer to the set of linguistic expressions including (beside words) phrases, clauses and sentences, since a phraseme can be any of the above.

– Second, the speaker has to select the lexical units to express his starting SemR and unite them into the deep-syntactic representation [= DsyntR] of the utterance.

Accordingly, two cases of violation of selection freedom must be distinguished.

The first case. The construction of the starting meaning 'σ' [= the SemR] of the utterance L('σ') that describes the situation P is free. To obtain 'σ', the speaker selects simpler meanings 'σ_1', 'σ_2', …, 'σ_n' and puts them together in conformity with his own needs and general rules of his language: the language does not specifically constrain his 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 other. The violation of the selection freedom takes place in the transition {SemR} $\Leftrightarrow$ {DsyntR} and manifests itself in lexical constraints. Therefore, resulting phrasemes are called lexical: *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 xplic vedra* lit. 'Rain is.pouring as from bucket'and *prolivnoj dožd′* 'torrential rain' are typical lexical phrasemes.

DEFINITION 3: LEXICAL PHRASEME

> A phraseme is lexical iff its meaning is constructed by the speaker freely, but its lexical components L_i (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 the situation P, the Speaker is forced by the language to select the starting meaning 'σ', and he cannot take an equivalent meaning 'σ'' or 'σ''' ('σ' $\equiv$ 'σ'' $\equiv$ 'σ'''). Thus, the utterance L('σ') is constrained semantically **and** lexically. This type of phraseme is thus "doubly" constrained: in the transition {ConceptR} $\Leftrightarrow$ {SemR} (semantic constraints) and then in the transition {SemR} $\Leftrightarrow$ {DsyntR} (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, vykrašeno* (with a different aspect prefix); and in English it would be inappropriate to write on a sign #*Caution, painted*, although this is a perfectly grammatical and semantically correct utterance. Here the language prescribes the meaning to express and its specific lexical expression.

DEFINITION 4: SEMANTIC-LEXICAL PHRASEME

> A phraseme is semantic-lexical, iff not only the components of its lexical expression, but also the components of its meaning are selected by the Speaker in a constrained way.

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.'.

Thus, a major partition splits phrasemes into two subsets: lexical phrasemes and semanic-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 that a linguistic sign **s** is a triplet

$$\mathbf{s} = \langle\ '\sigma'\ ;\ /s/\ ;\ \Sigma\ \rangle,\ \text{where:}$$

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

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

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

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

$\langle$ 'vehicle designed to fly'-SG/PL;[3] /έˀrkræft/; Σ = noun, countable, Lexical Functions: *land*$_\text{v}$, *crew*, ...$\rangle$

Simple signs combine into complex signs by the operation of linguistic union $\oplus$. For a particular language, this operation is represented by a set of linguistic rules that tell us how, in this language, signs must be united:

– the signifieds are united by putting the SemR of an argument into the corresponding argumental position of the corresponding predicate;

– the signifiers are united by juxtaposing the strings of phonemes and applying all necessary morphological operations;

– the syntactics are united by retaining the combinatorial data valid for the resulting complex sign.

DEFINITION 5: COMPOSITIONAL COMPLEX LINGUISTIC SIGN

A complex linguistic sign **AB** is compositional iff $\mathbf{AB} = \mathbf{A} \oplus \mathbf{B}$.

This means that, for the sign $\mathbf{AB} = \langle\ '\text{AB}'\ ;\ /\text{AB}/\ ;\ \Sigma_\mathbf{AB}\ \rangle$, its signified 'AB' = 'A' $\oplus$ 'B', its signifier /AB/ = /A/ $\oplus$ /B/ and its syntactics $\Sigma_\mathbf{AB} = \Sigma_\mathbf{A} \oplus \Sigma_\mathbf{B}$.

From this definition, it 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 what follows I will consider only the compositionality of signifieds, i.e., the semantic compositionality.

A free utterance is necessarily compositional: it is thanks to this property that linguistic communication is possible. To master language **L** means to have in the brain 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 implies the syntagmatic axis. Taking into account the two axes of speech production guarantees that our characterization of phrasemes is exhaustive.

---

[3] This notation indicates that the meanings of the grammemes SINGULAR and PLURAL are part of the signified of the lexeme AIRCRAFT.

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 the **semantic pivot** of 'σ' 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'. 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.

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 idioms, see below, **2.3.1.**

2.3. Major Classes of Phrasemes

Crossing the two dimensions–paradigmatic, i.e., lexical *vs.* semantic-lexical constraints and being compositional *vs.* non-compositional–gives four major classes of phrasemes:

Compositionality of phrasemes — Nature of constraints	non-compositional	compositional
lexical	IDIOMS	COLLOCATIONS
semantic-lexical	**i m p o s s i b l e**	CLICHÉS

Figure 1: Three Major Classes of Phrasemes

One of these logically possible classes–semantic-lexical non-compositional phrasemes–cannot exist in reality: if a non-free utterance U_{phr} is non-compositional, 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' ≈ 'get nothing', ⌐*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: 'AB' ⊅ 'A' **and** 'AB' ⊅ '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*¬ 'be too often or for too long' before Y's eyes' (lit. 'make corns on 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:
> > 'AB' ⊃ 'A', **and** 'AB' ⊃ 'B', **and** 'AB' ⊃ 'C'.

Examples:

Fr. ˝*donner le sein à Y*˝ 'breast-feed Y' = 'feed the baby Y by putting the teat
of one breast into the mouth of Y'

˝*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]'

2.3.2. Collocations

<small>DEFINITION 11: COLLOCATION</small>

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'' PRIVJAZNOJ REMEN'* lit. 'button up seatbelt'; *leap* YEAR, Rus.
Visokosnyj GOD (the adjective VISOKOSNYJ cannot be used without 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 as a function of the base. A col-
location 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 Britain, you can also *take* it. For the same
thing, French says *prendre* [= 'take'] *une la 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* 'Jean has taken the decision
to stay' ≅ *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.

DEFINITION 12: STANDARD COLLOCATION

> A collocation "Base–**r**–Collocate" is standard iff the semantic relation **r** is applicable, in language **L**, to many different bases and defines many different collocates.

In other words, **L** has many collocations where the relation between the base and the collocate is **r**. "Many" means here at least several dozen or, better, hundreds.

Examples: STORM *howls*, WHIP *cracks*, INSECT *buzzes*, DOG *barks* ⟨*yaps*, *whines*⟩, COW *moos*, CRICKET *chirps*, MACHINE-GUN *chatters*, TRAIN *rambles*, (DRY) TREE *squeaks*, etc.; **r** = 'produce the typical sound.

DEFINITION 13: NON-STANDARD COLLOCATION

> A collocation "Base–**r**–Collocate" is non-standard iff the semantic-syntactic relation **r** is applicable, in language **L**, only to a few different bases (in the minimal case, to one base) and defines a few different collocates (again, minimally, just one).

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

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 thus 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 xpl 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 a lexical anchor (or anchors), which is the lexeme(s) whose meaning controls 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.

2.4. An Important Subclass of Clichés: Pragmatemes

Along with lexical and semantic-lexical constraints that violate the freedom of lexical selection on the paradigmatic axis, natural language features a third type of constraint–situational, or pragmatic, constraints. Such constraints stipulate that a particular lexical expression may be required by a particular situation of its use. Thus, as a warning on a container of perishable food, 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 and compositional–that is, they are clichés. But this is a particular type of clichés, since they are used in a very 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 particular 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!* [**to a performer who is going on stage**]
collocations   : *Wet paint* [**on a sign**]
clichés   : *No parking* [**on a sign**]
lexemes   : *Roger!* 'I understood' [**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 phrasemes 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**], etc. (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.)

An interesting subclass of pragmatemes constitute complex proper names: *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].

2.5. Typology of Phrasemes

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

```
                        p h r a s e m e s

        lexical phrasemes              semantic-lexical phrasemes

n o n - c o m p o s i t i o n a l      c o m p o s i t i o n a l

        idioms                 collocations              clichés

  full    semi-   quasi-   standard   non-standard  pragmatically   pragmatemes
 idioms  idioms   idioms  collocations collocations non-constrained
                                                       clichés
```
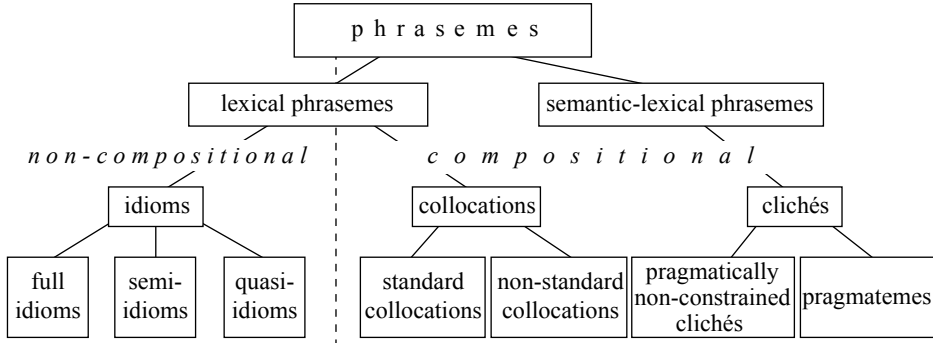
Figure 2. Phraseme Typology

3. Phraseology in the Dictionary

The dictionary considered here is the *Explanatory Combinatorial Dictionary* [= ECD]; its principles, structure and basic notions are taken for granted [see Mel'čuk, Zholkovsky 1984; Mel'čuk 1988b; Mel'čuk et al. 1984–1999; Mel'čuk et al. 1995; Mel'čuk 2006; Mel'čuk, 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 article, featuring the same structure as a lexeme article, with one important difference: since an idiom is a multiword utterance, it is supplied with its Ssynt-tree; cf.:

ˊPULL THE WOOL OVER [N_Y's] EYESˋ, 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'.

The surface-syntactic structure

┌── **oblique.obj** ────────────────────────┐
                                            ↓
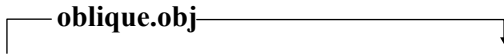
SsyntS: PULL–**dir.obj**→WOOL$_{SG}$–**determin**→THE OVER–**prepositional**→EYE$_{PL}$
   Government Pattern

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, Makin, McCaig 1993 contains about 7,000 idiomatic expressions, and the French idiom dictionary Rey, 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 about 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, for facilitating 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, Mel'čuk 1967; Mel'čuk 1974: 78–109; 1981; 1996; 2003a, b; Kahane 2003b; Kahane, Polguère 2001; Wanner (ed.) 1996]. I will limit myself to a few examples of standard and non-standard LFs, 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 ~(®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	~ 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 $\mathbf{f_{non\text{-}stand}}$ 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, $\mathbf{f_{nonstand}}$ must be described in the same metalanguage as that used for lexicographic definitions:

with no diary products added(COFFEE)   : *black* [~]
draining 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 (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	1. *with* N 2. *against* N	1. *for* N 2. *over* N
3. *between* N *and* N		3. *to* V_{inf}

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*
Mult : *//hostilities; war*
Loc_{in} : *in* [~]
Ver : *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_Y]
$Real_1$: *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 font 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.

Clichés (including pragmatemes) and pragmatically constrained lexemes are presented in the articles of their anchor(s) in a way similar to 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 late, 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 the 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. 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, Wawrzyńczyk 1993 and Bogusławski, 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* $N_{(period)Y}$" 'X who/which became well known during 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. I hope that the ECD is 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. 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 the 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 (1a) and its translations in Russian and French (1b-c):

  (1) a. She competes internationally.
       b. Rus.   Ona učastvuet v meždunarodnyx sorevnovanijax
                  'She participates in international competitions'.
       c. Fr.   Elle xplicativ à des xplicative xplicativels [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 $S_\emptyset(V)$ and one of its support verbs: $V \Leftrightarrow S_\emptyset(V) \leftarrow$**II**$-$Oper$_1(S_\emptyset(V))$. This formula describes all equivalences of the type *compete* $\Leftrightarrow$ *participate in competition(s)*; the noun *competition* has direct equivalents in Russian and French.

For a universal Dsynt-paraphrasing system, see Žolkovskij, Mel'čuk 1967; Mel´čuk 1974: 141–176, 1988c, 1992 and 2004, and Milićević 2007; Mel'čuk, 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, Wanner 2007. The paraphrasing system for Russian has been

implemented and tested in a series of computer experiments: Apresjan, 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 (2a):

(2) a. *The book thief struck again*.

Its closest (= most literal) Russian translation is (2b):

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 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 $Fact_0$: $Fact_0(L) \approx$ '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 himself'

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 $Fact_\emptyset$ are obtained automatically.

Example 2: The Polish verb OBOWIĄZYWAĆ 'oblige'

Here is a sign seen in the hall of a Warsaw building:

(4) a. *Mieszkańców budynku obowiązuje cisza nocna*
tenant-PL.ACC building-SG. oblige-IND. silence-SG. nocturnal
GÉN PRES.3SG NOM
lit. 'Nocturnal silence obliges tenants of [the] building'.

The meaning is immediately clear: the tenants are asked not to make noise at night. The closest Russian translation is as follows:

b. *Žiteli   doma   objazany   noč´ju   sobljudat´   tišinu*
   tenant-PL.   building-   are.obliged at.night observe   silence-SG.
   NOM   SG.GÉN   ACC

lit. 'Tenants of [the] building are.obliged at.night to.observe silence'.

Sentence (4b) can be obtained from (4a) by using dictionary information on lexemes and phrasemes. More specifically, (4a) presents two difficulties: the translation of OBOWIĄZYWAĆ 'oblige' and that of the expression *cisza nocna* 'nocturnal silence'.

The first difficulty is the specific government of the Polish verb, because of which it does not have a direct Russian–or English, for that matter–equivalent: **tišina objazyvaet…* *'the silence obliges…'. A Polish ECD must then contain, for OBOWIĄZYWAĆ 'oblige' the following indication:

$$X \text{ obowiązuje } Y\text{-}a \Leftrightarrow Y \text{ is obliged to Real}_I(X)\text{-}\mathbf{II}{\rightarrow}X$$

A Russian ECD has, under TIŠINA, Real$_I$(TIŠINA 'silence') = *sobljudat´* 'observe', which allows for the correct construction of the sentence's initial part: *Žiteli doma objazany sobljudat´ tišinu*.[4] (The English ECD has, under SILENCE, Real$_I$(SILENCE) = //*be quiet*.)

The second difficulty concerns a particularity of the Russian lexicon: the noun TIŠINA 'silence' has the collocate *nočnaja* 'nocturnal', which corresponds exactly to the Polish adjective *nocna* 'nocturnal'. Nonetheless, you cannot say **sobljudat´ nočnuju tišinu*–this is not a proper way of saying it; you have to put it as follows: *noč´ju sobljudat´ tiding.* To solve this difficulty it is sufficient to indicate, under TIŠINA 'silence', in the subarticle of Real$_I$(TIŠINA) = SOBLJUDAT´, that the temporal or locative modifier of the noun TIŠINA must be syntactically transferred to the verb SOBLJUDAT´:

$$\text{*SOBLJUDAT´-}\mathbf{II}{\rightarrow}\text{TIŠINA-}\mathbf{ATTR}{\rightarrow}\Xi_{(\text{«temp», « loc»})} \Rightarrow \Xi_{(\text{«temp», « loc»})}{\leftarrow}\mathbf{ATTR}\text{-SOBLJUDAT´-}\mathbf{II}{\rightarrow}\text{TIŠINA}$$

(This particular rule corresponds to a general rule of deep syntactic paraphrasing: Mel'čuk 1992: 50, Rule n° 19; for instance: *They launch* **regular**←**ATTR**-*attacks.* ≡ *They* **regularly**←**ATTR**-*launch attacks*.) In our case, the result is *soubljudat´-***ATTR**→*noč´ju tišinu*.

---

4 Three more examples of the same type for OBOWIĄZYWAĆ 'oblige' (thanks to G. Sekunda for them):

(i) *Czy sądu w tej sprawie nie* **obowiązuje** *tajemnica lekarska?*
lit. 'Is.this.that tribunal$_{\text{DirO}}$ in this affair doesn't oblige secret$_{\text{Subj}}$ medical?' =
'Isn't the tribunal obliged, in this affair, to **keep** the medical secret?'

(ii) *Kodeks* **obowiązuje** *także osoby zatrudnione na podstawie umów prawa prywatnego*
lit. 'Code obliges also persons employed on basis of.contracts of.law private'. =
'The people employed according to contracts of private law are also obliged to **respect** the code'.

(iii) *Uczestników* **obowiązuje** *noszenie na czapkach białych pokrowców*
lit. 'Participants$_{\text{DirO}}$ obliges wearing on caps white covers'. =
'The participants are obliged to **wear** on their caps white covers'.

Example 3: 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 on a French bus:

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

A possible translation into Russian could be 0b):

  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':

  « <u>Authorities ask you to Y</u> »   : *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:

  « <u>Authorities ask you to</u> 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.

5. Conclusions

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

1. Phrasemes constitute a very significant part of any language lexicon; 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.
4. LFs 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 J., Mejri S. (eds.), 2011, *Le figement linguistique: la parole entravée*, Paris: Honoré Champion.

Apresjan J., Cinman L., 1998, Perifrazirovanie na komp´jutere [Paraphrasing with Computer], *Semiotika i informatika*, No. 36, p. 177–202.

Apresjan J., Cinman L., 2002, Formal´naja model´ pererazirovanija predloženij dlja xplic pererabotki tekstov na estestvennyx jazykax [A Formal Model of Sentence Paraphrasing for Natural Language Text Processing Systems], [in:] *Russkij jazyk v naučnom osveščenii*, No. 2 [= 4], p. 102–146.

Bally Ch., 1909/1951, *Traité de stylistique française*, Genève: Georg et Cie, Paris: Klincksieck.

Beck D., Mel'čuk I., 2011, Morphological Phrasemes and Totonacan Verbal Morphology. *Linguistics*, No. 49: 1, p. 175–228.

Becker J., 1975, The Phrasal Lexicon [in:] *Proceedings of the Workshop on Theoretical Issues in NLP*, Cambridge (MA): MIT Press, p. 70–73.

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

Bogusławski A., Danielewiczowa M., 2005, *Verba polona abscondita. Sonda słownikowa III* [Concealed Polish Words. Lexicographic Probe III], Warszawa: Uniwersytet Warszawski.

Bosredon B., 2011, Dénominations monoréférentielles, figement et signalétique [in:] J. Anscombre, S. Mejri (eds.), *Le figement linguistique: la parole entravée*, Paris: Honoré Champion, p. 155–169.

Burger H., Dobrovol´skij D., Kuhn P., Norrrick N. (eds.), 2007, *Phraseology: An International Handbook of Contemporary Research*, vols. 1–2, Berlin: de Gruyter.

Cowie A. (ed.), 1998, *Phraseology. Theory, Analysis, and Applications*, Oxford: Clarendon Press.

Cowie A., Makin R., McCaig I., 1993, *Oxford Dictionary of English Idioms*, Oxford: Oxford University Press.

Everaert M., van der Linden E.J., Schenk A., Schreuder R. (eds.), 1995, *Idioms. Structural and Psychological Perspectives*, Hillsdale, N.J. Hove, UK: Lawrence Erlbaum Associates.

Grossmann F., Tutin A. (eds.), 2003, *Les collocations: analyse and traitement.* [= *Travaux et recherches en linguistique appliquée, Série E: Lexicologie et lexicographie*, nº 1], Amsterdam: De Werelt.

Iordanskaja L., Kim M., Polguère A., 1996, Some Procedural Problems in the Implementation of Lexical Factions for Text Generation [in:] L. Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam–Philadelphia: Benjamins, p. 279–297.

Kahane S., 2003a, The Meaning-Text Theory [in:] V. Agel, L. Eichinger, H.-W. Eroms, P. Hellwig, H.J. Heringer, H. Lobin (eds.), *Dependency and Valency. An International Handbook of Contemporary Research*, vol. 1, Berlin–New York: W. of Gruyter.

Kahane S., 2003b, Sur le lien entre la xplicativ lexicographique et les fonctions lexicales: une blessure profonde dans le DEC [in:] F. Grossman, A. Tutin (eds.), *Les collocations: analyse and traitement.* [= *Travaux et recherches en linguistique appliquée, Série E: Lexicologie et lexicographie*, nº 1], Amsterdam: De Werelt, p. 61–73.

Kahane S., Polguère A., 2001, Formal Foundations of Lexical Functions [in:] *Proceedings of "COLLOCATION: Computational Extraction, Analysis and Exploitation", 39th Annual Meeting and 10th Conference of the European Chapter of the Association for Computational Linguistic*s, Toulouse, p. 8–15.

Lareau F., Wanner L., 2007, Towards a Generic Multilingual Dependency Grammar for Text Generation [in:] T. Holloway King, 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 [9.10.2009].

Lubensky S., 1995, *Russian-English Dictionary of Idioms*, New York: Random House.

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

Mel'čuk I., 1981, Meaning-Text Models: A Recent Trend in Soviet Linguistics, *Annual Review of Anthropology*, No. 10, p. 27–62.

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

Mel'čuk I., 1988b, Semantic Description of Lexical Units in an *Explanatory Combinatorial Dictionary*: Basic Principles and Heuristic Criteria, *International Journal of Lexicography*, No. 1: 3, p. 165–188.

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

Mel'čuk I., 1992, Paraphrase and lexique: Vingt ans après [in:] I. Mel'čuk et al., *Dictionnaire xplicative et combinatoire du français contemporain. Recherches lexico-semantics I–IV*, Montréal: Les Presses of l'Université de Montréal, p. 9–58.

Mel'čuk I., 1995, Phrasemes in Language and Phraseology in Linguistics [in:] M. Everaert et al. (eds.), *Idioms. Structural and Psychological Perspectives*, Hillsdale, N.J. Hove, UK: Lawrence Erlbaum Associates, p. 167–232.

Mel'čuk I., 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, p. 37–102.

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

Mel'čuk I., 2003a, Collocations dans le dictionnaire [in:] Th. Szende (ed.), *Les écarts culturels dans les dictionnaires bilingues*, Paris: Honoré Champion, p.19–64.

Mel'čuk I., 2003b, Les collocations: definition, xpl et utilité [in:] F. Grossmann, A. Tutin (eds.),*Les collocations: analyse and traitement.* [= *Travaux et recherches en linguistique appliquée, Série E: Lexicologie et lexicographie*, nº 1], Amsterdam: De Werelt, p. 23–31.

Mel'čuk I., 2004, Verbes supports sans peine, *Lingvisticæ Investigationes*, No. 27: 2, p. 203–217.

Mel'čuk I., 2006, Explanatory Combinatorial Dictionary [in:] G. Sica (ed.), *Open Problems in Linguistics and Lexicography*, 2006, Monza (Italy): Polimetrica Publisher, p. 225–355. See also: http://www.polimetrica.com/?p=productsList&sWord=lexicography [11.03.2010].

Mel'čuk I., Clas A., Polguère A., 1995, *Introduction à la lexicologie explicative et combinatoire*, Bruxelles: Duculot.

Mel'čuk I., Polguère A., 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, p. 204–215.

Mel'čuk I., Polguère A., 2007, *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations semantics and collocations du français*, Bruxelles: De Boeck.

Mel'čuk I., Wanner L., 2006, Syntactic Mismatches in Machine Translation, *Machine Translation*, No. 21, p. 81–138.

Mel'čuk I., Zholkovsky A., 1984, *Explanatory Combinatorial Dictionnary of Modern Russian*, Vienna: Wiener Slawistischer Almanach.

Mel'čuk I. et al., 1984, 1988, 1992, 1999, *Dictionnaire xplicative et combinatoire du français contemporain. Recherches lexico-semantics I–IV*, Montréal: Les Presses of l'Université de Montréal.

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

Nattinger J., 1980, A Lexical Phrase Grammar for ESL, *TESOL Quarterly*, No. 14, p. 3, 337–344.

Rey A., Chantreau S., 1993, *Dictionnaire des expressions et locutions*, Paris: Dictionnaires Le ROBERT.

Wanner L. (ed.), 1996, *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam–Philadelphia: Benjamins.

Weinreich U., 1969, Problems in the Analysis of Idioms [in:] J. Puhvel (ed.), *Substance and Structure of Language*, Berkeley–Los Angeles, CA: University of California Press, p. 23–81. Reprinted in: U. Weinreich, *On Semantics*, 1980, Philadelphia: University of Pennsylvania Press, p. 208

Žolkovskij A., Mel'čuk I., 1967, O semantičeskom sinteze [On Semantic Synthesis], *Problemy kibernetiki*, vol. 19, p. 177–238.