Ju. D. Apresjan, I. M. Boguslavsky, L. L. Iomdin, L. L. Tsinman

Lexical Functions in NLP: Possible Uses [1]

Abstract

The paper describes the use of lexical functions, an instrument proposed in Igor Melčuk's Meaning ⇔ Text linguistic model, in advanced NLP applications, including parsers, high quality machine translation, a system of paraphrasing and computer-aided learning of lexica. In parsing, collocate LFs are used to resolve or reduce syntactic and lexical ambiguity. The MT system resorts to LFs to provide idiomatic target language equivalents for source sentences in which both the argument and the value of the same LF are present. The system of paraphrasing, which automatically produces one or several synonymous transforms for a given sentence or phrase, can be used in a number of advanced NLP applications ranging from machine translation to authoring and text planning. The computer-aided system of learning lexica is also based on the concept of LFs as a tool of formal description of that part of vocabulary which is simultaneously systematic and idiomatic and is therefore most difficult for language acquisition.

1. The notion of a lexical function (LF)

An LF, in Igor Mel'čuk's "Meaning ⇔ Text" model (Mel'čuk 1974, Mel'čuk & Zholkovsky 1984, Mel'čuk *et al*. 1984, 1988, 1992), has the basic properties of a multi-value mathematical function. A prototypical LF is a triple of elements {***R,*** X, Y}, where ***R*** is a certain general semantic relation obtaining between the argument lexeme X (the keyword) and some other lexeme Y which is the value of ***R*** with regard to X (by a lexeme in this context we mean a word in one of its lexical meanings or some other lexical unit, such as a set expression). Sometimes Y is represented by a set of synonymous lexemes Y_1, Y_2, …., Y_n, all of them being the values of the given LF ***R*** with regard to X; e. g., MAGN (*desire*) = *strong / keen / intense / fervent / ardent / overwhelming*. Whenever there is a variety of LF-exponents for a given keyword they are all listed in the respective explanatory-combinatorial dictionary.

There are two types of LFs – paradigmatic (substitutes) and syntagmatic (collocates, or, in Mel'čuk's terms, parameters).

A **substitute LF** is a semantic relation ***R*** between X and Y such that Y may replace X in the given utterance without substantially changing its meaning, although some regular changes in the syntactic structure of the utterance may be required. Examples are such semantic relations as synonyms, antonyms, converse terms, various types of syntactic derivatives and the like.

A **collocate LF** is a semantic relation ***R*** between X and Y such that X and Y may form a syntactic collocation, with Y syntactically subordinating X or vice versa. ***R*** itself is a very general meaning which can be expressed by many different lexemes of the given language, the choice among them being determined not only by the nature of ***R***, but also by the keyword with regard to which this general meaning is expressed. Typical examples of collocate LFs are support verbs of the OPER/FUNC family and such adjectival LFs as MAGN = 'a high degree of what is denoted by X', BON = 'good' and the like. Examples:

(1)

Function (R)	**Argument (X)**	**Value (Y)**
MAGN	*disease*	*grave*
MAGN	*fog*	*heavy*
MAGN	*control*	*strict*
MAGN	*to sleep*	*soundly*
MAGN	*to know*	*firmly*
$OPER_1$	*control*	*exercise (control)*

[1] This research has been supported by the Russian Foundation for Fundamental Research (grant No. 99-06-80292) and the Russian Foundation for the Humanities (grant No. 99-04-00318).

$OPER_1$	*research*	*do (research*)
$OPER_1$	*invitation*	*issue (an invitation)*
$OPER_1$	*doubt*	*have (doubts)*
$OPER_1$	*defeat*	*suffer (a defeat)*
$OPER_2$	*control*	*be under (control)*
$OPER_2$	*analysis*	*undergo (an analysis)*
$OPER_2$	*invitation*	*receive (an invitation)*
$OPER_2$	*resistance*	*encounter (resistance)*
$OPER_2$	*respect*	*enjoy (respect)*

Note that one cannot say, in good English, **to sleep firmly* or **to know soundly*, **to exercise an analysis* or **to make control*.

The LFs have the following general properties:

1. They are universal in the sense that several dozen LFs describe the basic semantic relations and the basic general meanings expressible in the vocabulary of any human language.

2. The LFs are idiomatic not only intralinguistically (see the examples above) but also cross-linguistically. Cf. the following Russian equivalents of LFs MAGN, $OPER_1$ and $OPER_2$:

(2)

Function (R)	**Argument (X)**	**Value (Y)**
MAGN	*bolezn'* 'disease'	*tjazhelaja* 'heavy'
MAGN	*tuman* 'fog'	*gustoj* 'dense'
MAGN	*spat'* 'to sleep'	*krepko* 'firmly'
MAGN	*znat'* 'to know'	*tverdo* 'solidly'
$OPER_1$	*somnenie* 'doubt	*ispytyvat'* 'feel'
$OPER_1$	*issledovanie* 'research'	*provodit'* 'lead'
$OPER_1$	*porazhenie* 'defeat'	*terpet'* 'tolerate'
$OPER_2$	*analiz* 'analysis'	*podvergat'sja* 'be subject to'
$OPER_2$	*uvazhenie* 'respect'	*pol'zovat'sja* 'use'

3. The LFs of the support-verb family can form combinations with their arguments which are synonymous to the basic verb X. E. g.,

(3) *He respects* [X] *his teachers – He has* [$OPER_1$ (S_0 (X))] *respect* [S_0 (X)] *for his teachers – He treats* [$LABOR_{12}$ (S_0 (X))] *his teachers with respect – His teachers enjoy* [$OPER_2$ (S_0(X))] *his respect*.

Most paradigmatic LFs (synonyms, antonyms, converse terms, various types of syntactic derivatives and the like) can also substitute for the keyword to form synonymous sentences:

(4) *He bought a coat for 500 dollars from a retail dealer – A retail dealer sold him a coat for 500 dollars – He paid 500 dollars to the retail dealer for a coat – The retail dealer received 500 dollars from him for a coat*

etc. Note the role of converse terms in both sets of paraphrases (3) and (4).

4. All the LFs of the support-verb family have stable syntactic properties in the sense that their valency structure reflects the valency structure of the verb which is the value of the given LF, on the one hand, and the syntactic structure of the underlying predicate word, on the other. Consider, for example, the verb *to analyze*. It is a two-valency verb, with the first actant usually fulfilling the syntactic function of the grammatical subject of the respective sentence and the second actant – that of the principal object of the verb, e.g. *He analyzed the text*. It has a syntactic derivative of the S_0 type *analysis* which combines with the verb *to make* as the value of LF $OPER_1$: *He made an analysis of the text. To make* in this use is also a two-valency verb, although in its freer meanings it may be a polyvalent verb, e. g., *The boy made a small hill out of snow with a spade.*

The above-mentioned properties of LFs make possible their practical use in a variety of NLP applications. Below we shall discuss the following four uses of LF in these applications: (a) syntactic and lexical ambiguity resolution in parsers; (b) idiomatic translation of a large class of set expressions in MT; (c) sentence paraphrasing; (d) computer-aided learning of lexica.

So far, to the best of our knowledge, there have been only two isolated attempts to apply the apparatus of LFs in NLP. The first was reported in Arsentyeva *et al*. 1969, who developed a tiny system of Russian paraphrasing using LFs, with the input and output formulated in terms of formal syntactic structures, so that no real texts were involved. The second was a series of experiments by the developers of the CAT2 MT system in Saarbrücken on finding translation equivalents with the help of LFs. The number of LFs was limited, and they were coded for a limited number of lexical entries (Streiter 1996). In the experimental applications reported below we have tried to amend the situation by resorting to representative linguistic and textual material. In particular, a) we make use of the whole range of about 100 standard LFs; b) LFs are systematically coded for over 3000 keywords; c) paraphrasing experiments were carried out on real NL sentences.

2. LFs in Computerized Dictionaries and NLP Grammar Rules

The most natural place for the information on LFs is the **dictionary**. In our applications, we use computerized dictionaries of Russian and English in which LFs are systematically given in all lexical entries for which the information on LFs is relevant. As an example we shall adduce a fragment of the lexical entry for *decision*:

DECISION
……
MAGN: FINAL / FIRM1 / IRREVERSIBLE / IRREVOCABLE
BON: SOUND2
ANTIBON: HASTY/RASH/SNAP3
VER: RIGHT1
ANTIVER: WRONG
OPER1: MAKE1/TAKE
INCEPOPER1: REACH/ARRIVE

FUNC2: CONCERN1 (*He made a decision on the steps to be taken – His decision concerned the steps to be taken*)

In all our applications, based on the multifunctional ETAP-3 linguistic processor (Apresjan *et al.* 1992) the lexicographer only adds the LFs, defined for the given keyword, and their appropriate values. He or she needs not think of any specific rules which underlie the implementation of the material in the NLP system: these rules, described in some detail previously (Iomdin & Tsinman 1997), are made rather general and belong to the grammar, in the broad sense of the word, and not to the dictionary.

3. LFs as a Disambiguation Tool in Parsing

The basic LF tool of disambiguation are collocate LFs. They are used to resolve syntactic and lexical ambiguity.

Among the collocate LFs the LFs of the support verb family are of greater importance because they serve to resolve both types of ambiguity – syntactic and lexical. Non-verbal types of collocate LFs (MAGN, ANTIMAGN, BON, ANTIBON, MULT, CAP, EQUIP, FIGUR etc) are typically used to resolve only lexical ambiguity.

3.1. Syntactic ambiguity resolution

Support verb classes of LFs (cf. property 4 in Section 1 above) impose strong limitations on the syntactic behaviour of their keywords in texts.

To give a few simple examples, many types of abstract predicate nouns, in English and other European languages alike, typically instantiate their different valencies in the same way (i.e. requiring the same prepositions or case marking; consider the well-known *amor patris* type of syntactic ambiguity). Phrases like *support of the president* or *the president's support* are syntactically ambiguous as they may mean both 'support from/by the president' and 'support (given) to the president'. A similar ambiguity is represented in *love of a beautiful woman, conviction of the judge,* etc. This type of ambivalence is often extremely difficult to resolve, even within a broad context.

LF support verbs can be successfully used to disambiguate such phrases. In particular, verbs of the $OPER_1$ type (Y) may form such collocations with their keywords (X) only if the **subject** valency of X is **not** instantiated. E.g.

(5) *The president spoke* [Y=$OPER_1$ (X)] *in support* [X] *of the parliament*

can only be interpreted as describing the support given to the parliament (by the president). Conversely, verbs of the $OPER_2$ type form collocations with their keywords only on condition that the latter's **object** valency is not instantiated, so that

(6) *The president had* [Y=$OPER_2$ (X)] *the support* [X] *of the parliament*

implies the support by the parliament (of the president). In much the same way, verbs of the $FUNC_1$ type require that the keyword's subject valency be not instantiated, while verbs of the $FUNC_2$ type combine with the keywords having uninstantiated object valencies. Accordingly,

(7) *The fear* [X] *of his wife possessed* [Y = $FUNC_1$ (X)] *Peter*

only means that Peter was afraid of his wife but, reasonably enough, not that his wife's fear of somebody or something did something to Peter. Consider a similar Russian example:

(8) *Obvinenija* [X] *prokurora otnosilis'* [Y=$FUNC_2$ (X)] *ko vsem uchastnikam sdelki* lit. 'The accusations of the prosecutor bore upon all partners of the deal'

(8) can only imply that the prosecutor accused somebody and not that he himself was accused.

Such facts are well known in theoretical syntax and may be explained quite easily: the support verb takes over a valency of the keyword, which is promoted to the rank of the respective actant.

However, in order to use the facts for disambiguation in parsing sentences, the parser should be able to **identify** the keyword and the support verb in the sentence. The LFs are an ideal instrument that helps to do this. When the parser has identified both the argument and the value of an LF in sentences like those described above, it is instructed to delete the impermissible syntactic hypotheses in favour of the correct ones. Importantly, our parser is able to identify the keyword and its LFs in a number of quite complicated contexts, e.g. not only in sentences like (6) but also in rather cumbersome utterances like *The support* (Y) *of the parliament that the president may have had* (X).

To conclude this section, we shall consider one less trivial example of syntactic ambiguity[2].

The Russian noun *muzhestvo* and its English equivalent *courage* have the valency of the subject – the courageous person – and the valency of the action in which his or her courage is manifested. In both languages the latter valency can be expressed by an infinitive, but only on condition that the noun is governed by its LF-verb with the core meaning of $OPER_1$ or $FUNC_1$.

The list of such verbs contains the following items: *imet' (muzhestvo sdelat' chto-libo)* 'to have the courage to do something' = $OPER_1$; *davat' (komu-libo muzhestvo sdelat' chto-libo)* 'to give somebody the courage to do something' = $CAUSOPER_1$; *xvatat' (muzhestva sdelat' chto-libo)* lit. 'His courage sufficed for him to do something' = $FUNC_1$; *nedostavat' (muzhestva sdelat' chto-libo)* lit. 'His courage did not suffice for him to do something' = not$FUNC_1$; (cf. *He lacked the courage to do something*).

Note that neither *muzhestvo* nor *courage* can govern the infinitive outside the context of such LF verbs: sentences (9a) and (9b) are ungrammatical:

[2] It is due to a discussion of the pattern of government of the nouns *courage* and *muzhestvo* 'courage' that one of the authors had with A. V. Lazursky.

(9a) **Muzhestvo pereprygnut' cherez zabor bylo narisovano na ego lice,*

(9b) **The courage to jump the fence was manifest in face.*

In Russian the same valency can also be instantiated by a *chtoby*-sentence ('in order that' sentence):

(10) *U tebja xvatit muzhestva, chtoby pereprygnut' cherez zabor* 'Do you have the courage to jump the fence'?

In (10), the *chtoby*-sentence is the first complement of the noun *muzhestvo*. On the other hand, the same kind of *chtoby*-sentence can be an adverbial modifier of purpose if the verb governing the noun *muzhestvo* falls outside the OPER-FUNC family. Consider sentence

(11) *Emu prishlos' sobrat' vse svoe muzhestvo, chtoby pereprygnut' cherez zabor* 'He had to muster up all his courage in order to jump the fence'.

Notwithstanding a great outer similarity, the underlying syntactic structures of (10) and (11) are widely different, with the chtoby-sentence filling in the valency of *muzhestvo* in (10) and serving as an adverbial modifier of the verb *sobrat'* in (11). As our glosses show, this difference is crucial for the right translation of (10) and (11) into English.

3.2. Lexical ambiguity resolution

LFs are also useful in resolving lexical ambiguity. For the sake of brevity, we will only give one illustrative example. The Russian expression *provodit' razlichie* and its direct English equivalent *to draw a distinction* both contain keyword nouns and support verbs that play the role of the $OPER_1$ type of LF[3]. Both the Russian and the English verbs are extremely ambiguous. *Provodit'*, for example, has half a dozen of meanings ranging from 'spend' via 'perform' to 'see off', while *draw* is a polysemic verb for which dictionaries list 50 meanings or more. However, in both expressions the mutual lexical attraction between the argument of the LF and its value is so strong that, once the fact of their co-occurrence is established by the parser, we can safely ignore all other meanings and keep for further processing only the one relevant here.

4. Finding Idiomatic Equivalents in MT with the help of LF

Two important properties of LFs mentioned in Section 1, i.e. their semantic universality and cross-linguistic idiomaticity, make them an ideal tool for ensuring idiomatic translations of collocations on a very large scale in an MT system. As in the previous section, we will confine ourselves to a few clear examples of how this can be done.

As is well known, locative and temporal prepositions used to form prepositional phrases denoting places, sites, directions, time points, periods, intervals etc. reveal great versatility within one language and incredibly fanciful matching across languages. If we were to account properly for the discrepancies existing between the uses of these prepositions, say, in English and Russian, we would have to write very intricate translation rules involving complicated semantic and pragmatic data. Incidentally, a large share of the task may be achieved with the help of LFs. Consider the following correspondences that may be easily found with the help of two LFs: LOC (preposition denoting a typical location) and TEMP (preposition denoting a typical time point):

	English:		Russian:	
Function	**Argument**	**Value**	**Argument**	**Value**
LOC	*institute*	*In*	*institut*	*v2*[4]
LOC	*school*	*At*	*shkola*	*v2*
LOC	*work*	*At*	*rabota*	*na2*
LOC	*menu*	*On*	*menju*	*v2*
LOC	*list*	*On*	*spisok*	*v2*
LOC	*tree*	*In*	*derevo*	*na2*

[3] To be more exact, they are instantiations of a compound LF CausManif, which approximately means "render evident".

[4] Numbers refer to the cases that are subcategorized by the Russian prepositions. Namely, *v1* and *na1* require the accusative, whereas *v2* and *na2* the locative.

LOC	*North*	*In*	*sever*	*na2*
LOC	*Florida*	*In*	*Florida*	*v2*
LOC	*Cuba*	*In*	*Kuba*	*na2*
TEMP	*March*	*In*	*mart*	*v2*
TEMP	*Tuesday*	*On*	*vtornik*	*v1*
TEMP	*dawn*	*At*	*rassvet*	*na2*
TEMP	*moment*	*At*	*moment*	*v1*
TEMP	*noon*	*At*	*polden'*	*v1*
TEMP	*o'clock*	*At*	*chas*	*v1*

In order to ensure the production of these equivalents in MT, we must only identify the arguments and the value of the LF during parsing and substitute the correct value from the target language dictionary during generation.

One of the assets of this approach is that information on the LFs for a given language is independent of the respective information for some other language. This is as much as to say that idiomatic translation of LF material among any number of languages can be effected without requiring any revision of the already existing rules. In an MT system not resorting to the apparatus of LFs, idiomatic translations among a number of languages can also be achieved, but only at the expense of introducing new blocks of transfer rules for every pair of languages.

Another practical advantage of this approach is the fact that the lexicographers who code the data need only know one language (their native tongue) to be able to provide the values of the LFs.

5. An LF-Based Computer System of Paraphrasing Utterances

5.1. Theory

We presume some familiarity with the model of paraphrasing proposed by I. Mel'čuk (1974) and therefore will not expound it here. What is important for us in the present context is that we attempted to formalize and computerize it within a theoretical and application environment somewhat different from the classical variety of the "Meaning ⇔ Text" framework. This is the environment of the operative multifunctional NLP processor known as ETAP-3.

The latest version of ETAP-3 is a highly modulated multifunctional and multi-language NLP system using complete formal grammars of the working languages (morphology and syntax) and combinatorial Russian and English dictionaries of approximately 45 000 items each, storing multifarious information on the syntactic, semantic and combinatorial properties of lexical items. In the framework of ETAP-3 paraphrasing Russian sentences turns out to be just one more mode of operation – the mode of Russian-to-Russian machine translation.

Below we give a list of those peculiarities of ETAP-3 which make it distinct from its prototype – the classical version of the "Meaning ⇔ Text" model.

1. The input to ETAP-3 in any of its options is an NL text which is processed sentence by sentence. In other words, the starting point for all types of processing is text analysis, not text synthesis.

2. The input to a specific option within ETAP-3 is the (surface)-syntactic structure of the processed sentence yielded by ETAP's multifunctional parser. A (surface)-syntactic structure is a linearly ordered dependency tree whose nodes represent the word-forms of the sentence, pairs of nodes being linked by one of several dozen (55 for Russian) syntactic dependency relations.

3. ETAP-3 makes use of a single system of tree transformation rules which effect all the necessary changes in the syntactic structure. Therefore the rules of paraphrasing are no longer divided into lexical and syntactic.

4. On the other hand, we have introduced a division of all rules into canonization rules and paraphrasing rules proper. This allows to avoid in a natural way the process of paraphrasing *ad infinitum* which was a real danger because paraphrasing rules of the classical "Meaning ⇔ Text" theory were reversible.

Canonization rules reduce the input sentence to its syntactically and lexically simplest paraphrase comprising just those lexemes which are prototypical exponents of the respective concepts.

Paraphrasing rules proper are used right after canonization rules and yield clusters of paraphrases.

5. Various paraphrasing rules of the "Meaning ⇔ Text" model have incommensurate scopes. Cf. the

rule X ⇔ $OPER_1(S_0(X))$ + $S_0(X)$, with a very broad, if not universal, scope, and the rule X ⇔ $A_0(X)$ + GENER(X), with a very narrow scope (cf. *linguistics – linguistic science,* but not *botany – *botanical science*).

This situation can be quite adequately handled within the ideology of ETAP-3 in which two basic types of rules are distinguished – general rules (for more or less universal phenomena) and dictionary rules (to handle word-specific phenomena).

6. Accordingly the bulk of paraphrasing rules were formulated in terms of those LFs which are likely to have broad scopes in paraphrasing. They include SYN, CONV, ANTI, S_0, A_0, S_1, S_2, A_1, A_2 and the principal parametric LFs of the OPER-FUNC family – about 40 LFs all in all.

7. For a number of reasons, the majority of the original LFs were revised and redefined. However, considerations of space preclude a detailed discussion of these changes.

The resulting model of paraphrasing makes use of all the information from the combinatorial dictionary and most of the rules of ETAP-3.

In the system of paraphrasing the following sets of rules are used: (a) morphological analysis; (b) parsing; (c) LF-interpretation of the syntactic structure (deep-syntactic analysis); (d) normalization of the interpreted structure; (e) canonization rules; (f) paraphrasing rules proper including a block of the simplest word-order rules which preserve the well-formedness of the processed syntactic structure when the introduction of certain LFs requires its profound recasting; (g) reinterpretation and correction rules which prepare the syntactic structure for the input to lexical transfer rules of ETAP-3; (h) introduction of lexical material such as specific prepositions required by the respective LFs; (i) expansion rules; (j) syntactic synthesis; (k) morphological synthesis.

Below we list some of the more important rules of paraphrasing in the proper sense, only in terms of lexical substitution, illustrated, for brevity's sake, with English examples which however have very close Russian correlates (for a detailed description of the system of paraphrasing see Apresjan & Tsinman 1998).

(12) X SYN (X) *(to throw – to cast)*;

(13) X $CONV_{12}$ (X) *(The group consists of 20 persons – Twenty persons comprise this group);*

(14) X $CONV_{132}$ (X) *(He bought a coat for 500 dollars – He paid 500 dollars for a coat);*

(15) X $CONV_{321}$ (X) *(He paid 500 dollars for the coat – The coat cost him 500 dollars);*

(16) X Neg + $ANTI_2$ (X) *(He was absent at the meeting – He was not present at the meeting);*

(17) X + Y $ANTI_1(X)$ + $ANTI_2(Y)$ *(He began to observe the rules – He stopped violating the rules);*

(18) X X, pass *(The monopolies control the prices – The prices are controlled by the monopolies);*

(19) X COPUL + $A_1(X)$ (*This man interested me – This man was interesting to me*);

(20) X COPUL + $A_2(X)$ *(I hate this man – This man is hateful to me*);

(21) X COPUL + $S_1(X)$ (*He taught me at school – He was my teacher at school*);

(22) X $OPER_1$ + $S_0(X)$ (*They resisted the enemy – They offered resistance to the enemy)*;

(23) X $LABOR_{12}$ + $S_0(X)$ (*He respects his parents – He treats his parents with respect*);

(24) X $OPER_2$ + $S_0(X)$ (*They control the prices – The prices are under their control*);

(25) X $FUNC_0$ + $S_0(X)$ (*They are arguing heatedly – A heated argument between them is on*);

(26) X $FUNC_1$ + $S_0(X)$ (*He is afraid – Fear possesses him*);

(27) $INCEPOPER_1$ + $S_0(X)$ $INCEPOPER_2$ + $S_0(X)$ (*He conceived a dislike for her – She caused his dislike*);

(28) $FINOPER_1$ + $S_0(X)$ $FINOPER_2$ + $S_0(X)$ (*England lost control of this territory – This territory went out of England's control*);

(29) $LIQUOPER_1$ + S_0 (X) $LIQUOPER_2$ + S_0 (X) *(The government deprived the monopolies of control over the prices – The government took the prices out of the monopolies' control);*

(30) $LIQUOPER_1$ + S_0 (X) $LIQUFUNC_1$ + S_0 (X) *(We freed him of this burden – We lifted this burden from him).*

5.2. Experimental data

Below by way of illustration we shall adduce three clusters of paraphrases (out of several hundred) which we received in the course of our Russian computer experiments with the system described in 5.1. The paraphrases are listed in the order in which they actually appear on the screen.

(31) *Pravitel'stvo kontroliruet ceny – Kontrol' cen osushchestvliaetsia pravitel'stvom – Pravitel'stvo osushchestvliaet kontrol' cen – Ceny naxodjatsja pod kontrolem pravitel'stva – Ceny derzhit pravitel'stvo pod kontrolem – Ceny kontrolirujutsja pravitel'stvom* [The government controls prices – Control over prices is exercised by the government – The government exercises control over prices – The prices are under the government's control – The government keeps prices under control – The prices are controlled by the government].

(32) *Pravitel'stvo ustanovilo kontrol' nad cenami – Ceny byli postavleny pravitel'stvom pod kontrol' – Pravitel'stvo postavilo ceny pod kontro ' – Kontrol' byl ustanovlen pravitel'stvom nad cenami* [The government established control over prices – The prices were put under the control of the government – The government put the prices under its control – The government's control over prices was established].

(33) *Pravitel'stvo otmenilo kontrol' nad cenami – Ceny byli osvobozhdeny pravitel'stvom ot kontrolja – Pravitel'stvo osvobodilo ceny ot kontroljia – Ceny byli vyvedeny pravitel'stvom iz-pod kontrolja – Kontrol' byl otmenen pravitel'stvom nad cenami* [The government cancelled control of the prices – The prices were freed of the government's control – The government freed the prices of its control – Control over prices was cancelled by the government].

There are apparent mistakes in the quoted texts. They can in principle be eliminated, but the cost of elimination will be different depending on the type of mistake. Some mistakes can be improved quite easily. Unfortunately, there is quite a number of mistakes which can only be eliminated at the cost of introducing so many constraints into the rules of paraphrasing that their practical value will be called in question at least with regard to certain tasks. In such cases the choice of optimal strategy should be determined by the nature of the task at hand.

6. Computer-aided learning of lexica

The idea of using lexical functions for a system of computer-aided learning of lexica was conceived at the very beginning of the 90-ies. In 1993 we developed a prototype of the system using up to a hundred lexical functions and based on the vocabularies of Russian and English counting around five hundred lexical items each. We designed five basic linguistic games supplied with a system of automatic numerical assessment of the user's performance. In the spring of 1993, during the stay of Ju. D. Apresjan as a Humboldt award winner at the Lehrstuhl of Computerlinguistik of the Heidelberg University, this prototype was demonstrated to Prof. P. Hellwig. That laid a foundation for a fruitful two-year joint research of three teams (from Moscow, Heidelberg, and Klagenfurt) in the framework of an INTAS project coordinated by P. Hellwig. The result of this cooperation was a system named CALLex (Computer-Aided Learning of LEXical functions) including German material as well. Below we give a brief description of the system.

6.1. Dictionaries

The most important part of the system are dictionaries of the working languages storing, among others, the following types of information on the lexemes (a lexeme is a word taken in one of its meanings): (a) the translation or translations of the lexeme into the other working languages; (b) its analytical definition formulated in a special semantic metalanguage – a reduced, simplified and standardized version of the object language; (c) the set of LFs defined for it, supplied with the values of every LF. We shall consider below only the last two types of information.

6.1.1. Analytical definitions

The analytical definitions were written by Ju. D. Apresjan and are supposed to meet the following theoretical requirements:

1) Non-circularity.
2) Completeness and non-redundancy: the definition should contain all and only the semantic components making up the lexical meaning at issue.
3) Reducibility: the definitions should be decomposable, immediately or through a number of intermediate stages, into semantic primitives.

4) Systematicity: the set of definitions should be constructed in such a way as to bring out explicitly (by means of common semantic components) systematic semantic links of the given lexeme with other lexemes of the language.

To illustrate the general strategy of definitions consider the following set of temporal words: AUTUMN, DAY1, DAY2, EVENING, HOUR, JANUARY, MINUTE, MONDAY, MONTH1, MONTH2, MORNING, NIGHT, PERIOD, SEASON, SECOND, SPRING, SUMMER, SUN, SUNRISE, SUNSET, TODAY, TOMORROW, WEEK, WINTER, YEAR, YESTERDAY. Note that 'sun' is presumed to be a semantic primitive underlying, in some way or other, the meanings of most temporal (and many spatial) words.

Autumn = 'the season of the year between summer and winter when it starts to be getting colder'.

Day1 = 'the part of time between two consecutive sunrises'.

Day2 = 'the lightest part of day1 which follows morning, precedes evening and ends between 16 and 17 o'clock'.

Evening = 'the part of day1 when it starts to be getting dark, which includes sunset and ends between 23 and 0 o'clock'.

Hour = 'the 24th part of day1'.

January = 'the first month of the year'.

Minute = 'the sixtieth part of an hour'.

Monday = 'the first working day of the week'.

Month1 = 'one of the twelve parts into which a year is divided'.

Month2 = 'any period of approximately 30 days'.

Morning = 'the part of day1 when it starts to be getting light which includes sunrise and ends between 11 and 12 o'clock'.

Night = 'the darkest part of day1 which follows evening, precedes morning and ends between 4 and 5 o'clock'.

Period = 'part of time'.

Season = 'one of the four large parts into which a year is divided on the basis of the natural cycle considerations'.

Second = 'the sixtieth part of a minute'.

Spring = 'the season of the year between winter and summer when it starts to be getting warmer'.

Summer = 'the warmest season of the year with the longest days2 and shortest nights'.

Sun = 'this word is a semantic primitive'.

Sunrise = 'the process of the sun appearing on the horizon and the initial stages of its going up in the sky, or the time during which this process takes place'.

Sunset = 'the process of the sun disappearing beyond the horizon and the final stages of its going down in the sky, or the time during which this process takes place'.

Today = 'that day1 through which the speaker is living at the time of speech'.

Tomorrow = 'the day1 immediately after today'.

Week = 'a period of seven days singled out on the basis of human activity cycle considerations'.

Winter = 'the coldest season of the year with the shortest *days2* and longest nights'.

Year = 'the part of time equal to 365 days1, in which the Earth makes one full circle around the sun'.

Yesterday = 'the day1 immediately before today'.

6.1.2. Lexical functions

The set of around a hundred LFs was defined anew with a view to provide a systematic description of the algebra of LFs, on the one hand, and making the definitons transparent for the user, on the other. Below we give sample definitions of a dozen LFs used in the linguistic games. In the definitions, X stands for the keyword (the argument of the respective LF), P1 for its first actant and P2 for its second actant.

SYN [a synonym of X]
ANGRY: *mad / furious*
DATA: *information*
LOOK: *glance*
THANKFUL: *grateful*
WISH (a verb): *want / desire*

MAGN [a large degree or intensity of X]
FAITH: *deep / profound / unshakeable*
IMAGINATION: *rich / lively*
LOOK (a noun): *sharp*
POWER: *firm / stable / unshakeable*
SEE: *clearly / distinctly*

OPER1 [to do X, to have X or to be in the state of X (a support verb taking P1 as its grammatical subject and X as its principal complement)]
CHOICE: *make (one's choice)*
DESPAIR: *be in (despair)*
INTERVIEW: *take (an interview)*
LOOK: *give / take / throw (a look)*
MAJORITY: have (the majority) / be in (the majority)

OPER2 [to undergo the action of X or to be in the scope of X (a support verb taking P2 as its grammatical subject and X as its principal complement)]
BLESSING: *receive (a blessing)*
INFLUENCE: *be under (the influence)*
INTERVIEW: *give (an interview)*
RESPECT: *enjoy (smb's respect)*
TEST: *undergo (a test)*

FUNC1 [to be done by or to characterize smb (a support verb taking X as its grammatical subject and P1 as its principal complement)]
ACCIDENT: *happens to (smb)*

COMPLAINT: *comes (from smb)*
FEAR: *possesses (smb)*
FEVER: *tortures (smb)*
RESPONSIBILITY: *rests with (smb)*

LABOR1-2 [To subject P2 to the action of X (a support verb taking P1 as its grammatical subject, P2 as its principal complement and X as its secondary complement)]
ARREST: *keep (smb) under (arrest)*
CAUTION: *treat (smth) with (caution)*
CUSTODY: *keep (smb) in (custody)*
PUNISHMENT: *subject (smb) to (punishment)*
RESPECT: *hold (smb) in (respect)*

INCEPOPER1 [to start to do X, to have X or to be in the state of X (a support verb taking P1 as its grammatical subject and X as its principal complement)]
GRIPPE: *catch (the grippe)*
IMPORTANCE: *acquire (importance)*
LOVE: *fall in (love)*
MAJORITY: *get / receive / win (the majority)*
POWER: *come to (power) / win (power)*

FINOPER1 [to cease to do X, to have X or to be in the state of X (a support verb taking P1 as its grammatical subject and X as its principal complement)]
ACQUAINTANCE: *break off (the acquaintance)*
BUSINESS: *go out of (business)*
CONTROL: *lose (control)*
LIFE: *part with (life)*
STRUGGLE: *cease (the struggle)*

REAL1 [to use X in accordance with its destination (a verb taking P1 as its grammatical subject and X as its principal complement)]
APRICOT: *eat (an apricot)*
BICYCLE: *ride (a bicycle)*
HAT: *wear (a hat)*
MEDICINE: *take (medicine)*
NEWSPAPER: *read (a newspaper)*

INCEPREAL1 [to start to use X in accordance with its destination (a verb taking P1 as its grammatical subject and X as its principal complement)]
BUS: *board (a bus) / get on (a bus)*
DRESS: *put on (a dress)*
HOSPITAL: *go to (hospital)*
HOTEL: *check in at (the hotel)*
LIGHT: *turn on (the light)*

FINREAL1 [to cease to use X in accordance with its destination (a verb taking P1 as its grammatical subject and X as its principal complement)]
BUS: deboard (the bus) / *get off (the bus)*
DRESS: *take off (the dress)*
HOSPITAL: *leave (hospital)*
HOTEL: *check out of (the hotel)*
LIGHT: *turn off (the light)*

6.2. Linguistic games

CALLex offers 4 entirely innovative linguistic games based on the analytical definitions of lexemes and the LFs assigned to them: 1) supplying translations for a lexeme offered by the computer; 2) supplying the values of all the LF's (offered by the computer) for a word chosen by the user; 3) supplying the values of a LF (chosen by the user) for all the lexemes offered by the computer; 4) guessing a particular lexeme from its analytical definition (decomposition) offered by the computer.

In the basic mode the user is supposed to play those games unilingually. However it is possible to duplicate the second game in the bilingual mode whereby the search for the right answer is guided by the user's knowledge of his/her mother tongue, irrespective of what the mother tongue in question is.

That forms the fifth game called "Word with a tip".

Depending on the semantic and syntactic properties of a particular LF the games are arranged into three levels of difficulty.

The first level of difficulty is composed of LF's whose meaning is transparent and whose syntactic (or actant) structure is simple. Cf. such LFs as SYN or MAGN.

The second level of difficulty is composed of the LFs with a straightforward meaning but a complicated syntactic (or actant) structure; cf. the LFs OPER-FUNC family.

The third level of difficulty is composed of the LF's with a complicated meaning and an involved syntactic structure. Cf. the REAL-FACT family.

Each level of difficulty is broken into two sublevels, depending on the foregroundedness of the value of a particular LF. The prototypical values of LFs which are presumed to be in the foreground of the user's language competence, if properly guessed, score less than more peripheral values.

CALLex is supplied with a system of quantitative assessment of the extent of lexical knowledge the user displays in a particular linguistic game. The first level of difficulty scores the user one point for every correct answer. Every next level of difficulty scores him or her one point more. The user's performance is considered to be normal if he or she supplies at least one correct answer for every question. It is considered to be excellent if the user supplies more than one correct answer, in which case he or she is rewarded with ever greater scores for every answer beyond the “standard” .

At every stage in every game the user is shown the number of points he/she has scored for the current answer and his/her total score for the game.

A computer-driven dialogue system has been devised intended for specifying a) the language the user wishes to practise in, b) the linguistic game, and c) the level of difficulty.

All the games are supplied with exhaustive menus so that there is no need for further comments. Programming facilities are envisaged providing for constant updating and expansion of the ECDs and all the other linguistic resources used in CALLex.

References

Apresjan et al. 1992: Ju. D. Apresjan, I. M. Boguslavsky, L. L. Iomdin *et al*. ETAP - 2: The Linguistics of a Machine Translation System. META, 1992, Vol. 37, No 1, pp. 97-112.

Apresjan & Tsinman 1998: Ju. D. Apresjan, L. L. Tsinman. Perifrazirovanie na komp'jutere. [Paraphrasing on a Computer]. Semiotika i informatika, 1998, No. 36, pp. 177-202.

Arsentyeva et al. 1969. N. G. Arsent'eva, N. A. Balandina, A. I. Krasovskaja. O mashinnoj realizacii sistemy perifrazirovanija [On the Computer Implementation of a Paraphrasing System]. Institute for Applied Mathematics, Academy of Sciences of the USSR. Preprints 25, 26, 27. Moscow, 1969.

Iomdin & Tsinman 1997: L. L. Iomdin, L. L. Tsinman. Lexical Functions and Machine Translation. Proceedings of Dialogue'97 Computational Linguistics and its Applications International Workshop. Yasnaya Polyana, Russia, June 10-15, 1997.

Mel’čuk 1974: I. A. Mel’čuk. Opyt teorii lingvisticheskix modelej "Smysl ⇔ Tekst" [A Theory of Meaning ⇔ Text Linguistic Models"]. Moscow, Nauka, 1974, 314 p.

Mel’čuk & Zholkovskij 1984: I. A. Mel’čuk, A. K. Zholkovskij. Tolkovo-kombinatornyj slovar' sovremennogo russkogo jazyka. [An Explanatory Combinatorial Dictionary of the Contemporary Russian Language] Wiener Slawistischer Almanach, Sonderband 14, 1984, 992 p.

Mel'čuk et al. 1984: Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Lidija Iordanskaja, Adèle Lessard. Dictionnaire explicatif et combinatoire du français contemporain, Recherches lexico-sémantiques I. Les Presses de l'Université de Montréal, 1984.

Mel'čuk et al. 1988: Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Louise Dagenais, Léo Elnitsky, Lidija Iordanskaja, Marie-Noëlle Lefebvre, Suzanne Mantha. Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques II. Les Presses de l'Université de Montréal, 1988.

Mel'čuk et al. 1992: Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Lidija Iordanskaja, Suzanne Mantha. Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques III. Les Presses de l'Université de Montréal, 1992.

Streiter 1996: Oliver Streiter. Linguistic Modeling for Multilingual Machine Translation. Shaker Verlag. Aachen.