

# ЛЕКСИКО-ФУНКЦИОНАЛЬНАЯ РАЗМЕТКА РУССКИХ ТЕКСТОВ

**Т. И. Фролова** (tfrolova@iitp.ru)

**О. Ю. Подлесская** (olga@iitp.ru)

Институт проблем передачи информации им. А.А.  
Харкевича, РАН, Москва, Россия

**Ключевые слова:** разметка текстов, лексико-функциональная раз-  
метка текстов, СинТагРус, лексические функции.

## TAGGING LEXICAL FUNCTIONS IN RUSSIAN TEXTS OF SYNTAGRUS<sup>1</sup>

**T. I. Frolova** (tfrolova@iitp.ru)

**O. Iu. Podlesskaia** (olga@iitp.ru)

Laboratory of Computational Linguistics, Kharkevich Institute  
For Information Transmission Problems, RAS, Moscow,  
Russian Federation

The present paper describes the process and the results of tagging with Lexical Functions the texts of SynTagRus (Syntactic Russian corpus available at [www.ruscorpora.ru](http://www.ruscorpora.ru)). The present work started in 2009 and it is still in progress in the Laboratory of Computational Linguistics Kharkevich Institute For Information Transmission Problems, RAS. The lexical items which are identified as values and arguments of collocate Lexical Functions (LFs) are tagged in syntactically annotated Russian sentences. At the moment about 4300 sentences (about 5500 LF-phrases) have been processed and all of them were supplied with LF-annotation. At the end of the paper some examples of possible linguistic and educational uses for the corpus with LF tagging are offered.

**Key words:** tagging, lexical functions, SynTagRus, tags.

---

<sup>1</sup> This work has been supported in part with the program «Текст во взаимодействии с социокультурной средой: уровни историко-литературной и лингвистической интерпретации» of the Division of History and Philology of RAS.

## 1. Introduction

The paper describes tagging Russian texts with collocate Lexical Functions, especially the process of work, some results and possible uses in linguistics and education. The work is being done for the last two and a half years in the Laboratory of Computational Linguistics, Kharkevich Institute For Information Transmission Problems, RAS.

### 1.1. Concept of collocate lexical functions

Collocate Lexical Functions (LFs) are certain meanings which can be expressed by different lexemes of the given language, the choice among them being determined not only by the meaning itself, but also by the keyword (argument) with regard to which this general meaning is expressed. For example the value of LF MAGN ('a high degree of what is denoted by X') is HEAVY for noun FOG (*heavy fog*) and GRAVE for noun DISEASE (*grave disease*).

Besides this idiomaticity within the given language (different values for different arguments) the values of LFs are often idiomatic between two different languages, thus the English adjective HEAVY in *heavy fog* is the value of LF MAGN for the noun FOG, but it is not a translation for Russian ГУСТОЙ (literally 'dense, thick') which is the value of LF MAGN for the Russian noun ТУМАН in *густой туман* ('heavy fog').

The apparatus of LFs was proposed in [1,2] alongside with the list of presumably universal LFs meanings which are expressed in non-trivial way in combination with the keywords, such as 'high degree' (MAGN), 'good' (BON), 'right' (VER), 'opposite' (ANTI), 'existence' (FUNC), 'beginning' (INCEP), 'end' (FIN), 'causation' (CAUS), 'liquidation' (LIQU), 'normal use' (REAL), 'normal functioning' (FACT), 'manifestation' (MANIF) etc.

### 1.2. Processing of LFs in ETAP

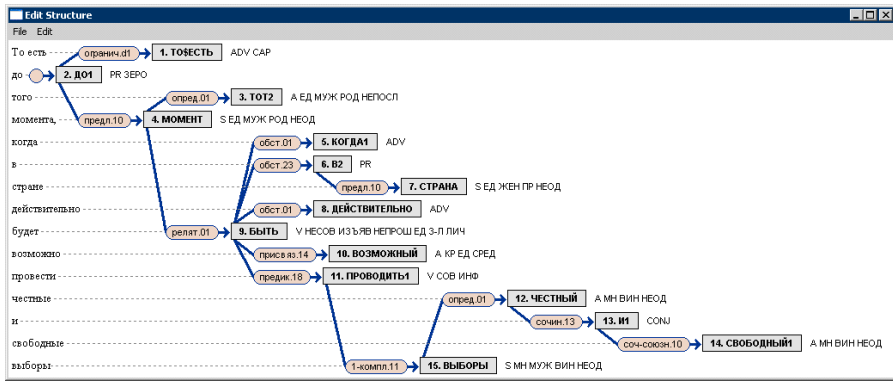
In multipurpose linguistic processor ETAP-3 developed in the Laboratory of Computational Linguistics of Kharkevich Institute For Information Transmission Problems, RAS, LF-information is recorded in special zone of combinatorial dictionary. This information is used for lexical and syntactic ambiguity resolution in analysis and for improvement of quality of translation, as well as for automatic paraphrasing. There is a special block of rules for identification LFs, which process syntactic tree structure of each sentence. For more information on ETAP-3 system and the use of LFs in this system see [3–5] and [6–7] respectively.

## 2. Appearance and techniques of LF-tagging

Analysis of LF-collocations is carried out for phrases already tagged morphologically and syntactically. Syntactic tagging of Russian texts is being done in the Laboratory of Computational Linguistics already for several years. These syntactically tagged texts

form an integral part of Russian National Corpus. The lexico-grammatical search in these texts is available at <http://www.ruscorpora.ru/search-syntax.html>. Syntactically annotated corpus contains about 45 000 sentences from fiction and journalistic texts.

Each sentence of the corpus is represented with its syntactic tree structure, where each word forming its node is provided with full set of morphological features, and the links are marked with the names of syntactic relations. Such a view of syntactic structure of the sentence goes back to "Meaning ⇔ Text" theory by Melchuk and Zholkovsky [see 1 and 2]. The procedure of syntactic tagging is semi-automatic: at first morphological analyzer and syntactic parser of ETAP-3 make up the syntactic tree structure of the sentence [more on tagging scheme and tools for corpus creation see 8 and 9]. Then a trained linguist checks and if necessary manually corrects the result of machine analysis. See below the morphologically and syntactically annotated sentence — the result of semi-automatic procedure described above:



**Fig. 1.** Syntactic structure of Russian sentence: «То есть до того момента, когда в стране действительно будет возможно провести честные и свободные выборы» ('That is till the moment when it'll be really possible to hold honest and free elections in the country')

The LF-analysis of sentences is also done semi-automatically: syntactically and morphologically tagged sentence (already checked by a linguist) is put into LF-analyzer. This analyzer uses the information assigned to lexemes in LF-zones of Russian combinatorial dictionary in ETAP, and also the rules of recognition of LF-phrases (i.e. phrase which consists of an argument of certain LF and its value) in the text.

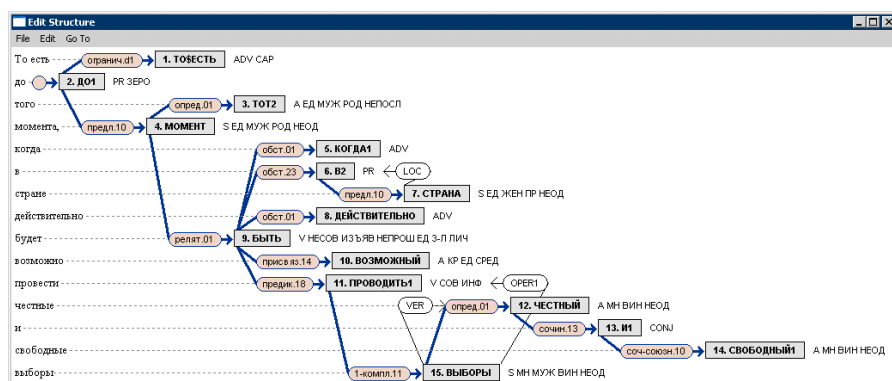
The rules of LF-recognition were created for machine translation and automatic paraphrasing. These rules define syntactic conditions for establishing LF-link between the argument and the value of certain LF.

For example, despite the presence of the record for LF MAGN:КРЕПКИЙ in LF-zone of article ЗДОРОВЬЕ in Russian combinatorial dictionary, the LF-link is not set between adjective КРЕПКИЙ ('firm') and noun ЗДОРОВЬЕ ('health') in coordinate phrase *крепкий сон и отличное здоровье* ('sound sleep and perfect health'). Adjective КРЕПКИЙ is used here as value of LF MAGN but of another noun, namely

COH ('sleep'). Correct automatic recognition is possible here due to the absence of syntactic links, specific for adjectival LFs, between adjective КРЕПКИЙ and noun ЗДОРОВЬЕ (see examples of such syntactic contexts below in part 3).

For the purpose of LF-tagging, these rules underwent some technical changes. In addition, the format of presenting LF-information in tagged texts was elaborated. This format in the form present on Fig. 2 was elaborated by members of the Laboratory V. G. Sizov and V. V. Petrochenkov.

The result of automatic LF-analysis of each sentence was checked and if necessary corrected and supplemented by linguists. See below on Fig. 2 one example of LF-structure marked for one Russian sentence. One can see LF-collocations: LOC — *в стране* ('in the country'), OPER1 — *провести выборы* ('hold elections') и VER — *честные выборы* ('honest elections'):



**Fig. 2.** Syntactic and LF-structure of Russian sentence: «То есть до того момента, когда в стране действительно будет возможно провести честные и свободные выборы» ('That is till the moment when it'll be really possible to hold honest and free elections in the country')

Results of LF-tagging are now available in the Laboratory, but they are also expected to be available on-line.

### 3. Principles of LF-tagging

All phrases which can be described in terms of standard LFs are marked in the process of tagging. For the list of standard LFs see for example work [10].

As already mentioned above, the information about LF-collocations is recorded in LF-zone of the combinatorial dictionary in ETAP-3 system. Syntactic conditions for identifying LF-phrases are described in ETAP-3 in the rules of LF-recognition. These rules (and these conditions) are composed in terms of syntactic links. For example, for LF MAGN (and other adjectival LFs) for nominal arguments possible contexts are the following:

- attributive phrase: *крепкое здоровье* ('sound health');
- nominal clause: *здоровье крепко* ('the health is sound');
- phrases with copulative verbs: *здоровье было (казалось, могло быть) крепким* ('the health was (seemed, could be) sound');
- and so forth.

All phrases found in these contexts are marked in the tagged sentences.

The cases are not marked and not recognized automatically when it is necessary for establishing correct LF-link to set correct anaphoric relations between sentences or within one sentence. For example, in sentence *здоровье не доставит проблем, потому что оно будет крепким* ('The health won't cause any problems because it will be sound') LF-link between noun ЗДОРОВЬЕ ('health') and adjective КРЕПКИЙ ('sound') won't be marked.

Cases are not marked when incorrect or nonstandard values of LFs are used in text. For example, the phrase *оплачивают налоги* ('(they) pay for taxes') is incorrect. The verb ОПЛАЧИВАТЬ ('pay for') is used in the sense of LF REAL1-M ('To do with regard to X that which is normally expected of P1') instead of the correct value of LF REAL1-M for НАЛОГ ('tax') — ПЛАТИТЬ ('pay'). The correct phrase is *платят налоги* ('(they) pay taxes'). In this case the phrase *оплачивают налоги* is not tagged as LF-phrase. Also the phrase *производить влияние* ('produce influence') is not tagged as LF-phrase, because in the sense of LF OPER1 ('To do X, to have X or to be in the state X') for the noun ВЛИЯНИЕ ('influence') the verb ПРОИЗВОДИТЬ ('produce') is used incorrectly instead of the correct value ОКАЗЫВАТЬ ('exert'). The phrase *плотно подсели*, nonstandard phrase used in the text in the sense of 'become strongly dependent on' is also not marked as LF-phrase, though adverb ПЛОТНО here could be interpreted here as value of LF MAGN ('high degree'). Decision not to include this phrase into the results of tagging was made due to the fact that the verb ПОДСАЖИВАТЬСЯ in the sense 'become dependent' is not the lexeme of the literary language.

Another cases excluded from LF-tagging are LF-phrases with arguments that are phrases and not single lexemes. For example the phrases *придерживаться точки зрения* or *иметь точку зрения* ('hold to or have the point of view') are not tagged as LF-phrases, though the verbs ПРИДЕРЖИВАТЬСЯ ('hold to') and ИМЕТЬ ('have') are values of LF OPER1 ('To do X, to have X or to be in the state X') for an argument ТОЧКА ЗРЕНИЯ ('point of view').

The question of including values of LFs determined by pragmatic component of arguments' meaning in LF-tagging is under discussion. For example, in the phrase *затяжная война* ('protracted war') the adjective ЗАТЯЖНОЙ ('protracted') could be interpreted as the value of LF MAGN ('high degree') for the component DURATION which is not a part of the meaning of the noun ВОЙНА ('war').

#### 4. Results of LF-tagging

The main result of the work is the corpus of 4300. The corpus which was examined contains journalistic texts having from twenty to more than two hundred

sentences each. Sentences, containing LF-collocations, make up approximately one third of all analyzed sentences (about 4300 sentences). These sentences contain about 5500 LF-collocations. Some additional information about quantities of LFs in corpus is given below. It should be noticed, however, that due to relatively small amount of current LF-corpus, it is now unreasonable to jump to almost any substantial linguistic conclusions on basis of this information: for example, the fact that the phrase *в понедельник* has 24 occurrences in corpus, whereas *в пятницу* has 11 occurrences, doesn't mean that the first phrase is twice more common in Russian. The data below is given only to provide more detailed view of the corpus.

The most frequent LF in the corpus is LOC ('A preposition denoting the normal spatial or temporal localization of something with regard to X', more than 1800 occurrences). Here are the most frequent collocations with this LF in corpus (see in brackets the number of occurrences for each of these collocations in corpus): *в году* (250), *в России* (172), *в стране* (121), *во время* (109), *в Москве* (48), *в мире* (47), *в школе* (41), *на рынке* (31), *в городе* (28), *в сфере* (28), *на территории* (27).

The most frequent verbal LF is LF OPER1 ('To do X, to have X or to be in the state X', about 950 occurrences). The most frequent arguments of this LF are the following (see in brackets values of this LF for each argument and the number of occurrences): РЕШЕНИЕ (ПРИНИМАТЬ/ВЫНОСИТЬ, 34), ВЫВОД (ДЕЛАТЬ/ПРИХОДИТЬ К, 29), РОЛЬ (ИГРАТЬ, 27), ИССЛЕДОВАНИЕ (ПРОВОДИТЬ/ВЕСТИ, 27), ВНИМАНИЕ (ОБРАЩАТЬ, 25).

There are some other verbal LFs present in corpus relatively frequently: FUNC0 ('X exists or is taking place', 191 occurrence), CAUSFUNC0 ('To cause X to happen or to exist', 188 occurrences), INCEROPER1 ('To start to do X, to have X or to be in the state X', 154 occurrences). See below the most frequent arguments of these LFs in corpus:

FUNC0: РЕЧЬ (ИДТИ, 36), ПРОЦЕСС (ИДТИ/ПРОХОДИТЬ, 15), ВОЗМОЖНОСТЬ (БЫТЬ/СУЩЕСТВОВАТЬ, 13), ПРОБЛЕМА (ИМЕТЬСЯ/СУЩЕСТВОВАТЬ, 8).

CAUSFUNC0: ЗАДАЧА (СТАВИТЬ, 13), РЕЗУЛЬТАТ (ДОСТИГАТЬ/ПОЛУЧАТЬ, 13), ИТОГ (ПОДВОДИТЬ, 7), ПАМЯТНИК (УСТАНАВЛИВАТЬ/ВОЗДВИГАТЬ, 7), ФИЛЬМ (СНИМАТЬ, 6).

INCEROPER1: ЗНАНИЕ (ПРИОБРЕТАТЬ/ПОЛУЧАТЬ, 8), ИНФОРМАЦИЯ (СОБИРАТЬ/ПОЛУЧАТЬ, 7), ВЛАСТЬ (ПОЛУЧАТЬ/ПРИХОДИТЬ К, 6), ДЕНЬГИ (ЗАРАБАТЫВАТЬ, 6), ДИПЛОМ (ПОЛУЧАТЬ, 6), РАБОТА (НАЧИНАТЬ, 6).

The most frequent adjectival LF is MAGN ('a large degree or a high intensity of X', 522 occurrences). The most frequent arguments with this LF are: УРОВЕНЬ (ВЫСОКИЙ, 14), РОСТ (БЫСТРЫЙ/СТРЕМИТЕЛЬНЫЙ, 12), РАСТИ (БЫСТРО/СТРЕМИТЕЛЬНО, 11), БИЗНЕС (БОЛЬШОЙ/КРУПНЫЙ, 10), ЗАРПЛАТА (ВЫСОКИЙ/БОЛЬШОЙ, 9), ЗНАТЬ (ХОРОШО/ТВЕРДО, 8), ИЗВЕСТНЫЙ (ХОРОШО, 6), КОЛИЧЕСТВО (БОЛЬШОЙ, 6), ПОНИМАТЬ (ХОРОШО/ЯСНО, 6).

For the purpose of additional illustration some sentences are given below from one of the texts in corpus, namely «Расслабьтесь и наслаждайтесь (Relax and Enjoy)», by Yevgeny Grigoryevich Yasin, from the newspaper "Trud" of the 22nd of September 2008. This text contains 48 sentences, of which only 19 with LF-phrases. For each of the sentences the picture of syntactic and LF-structure (or its fragment) is given.

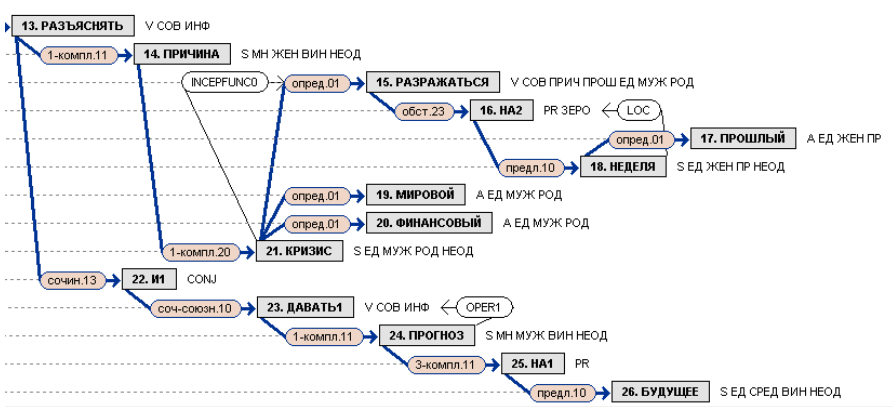
**Sentence 1** (see Fig. 3)

The sentence contains three LF-collocations:

НЕДЕЛЯ, LOC: НА,

КРИЗИС, INCEPFUNCO ('X starts to exist or to be taking place'); РАЗРАЖАТЬСЯ,

ПРОГНОЗ, OPER1: ДАВАТЬ.



**Fig. 3.** Syntactic and LF-structure of the sentence: «С сегодняшнего дня "Труд" начинает публикацию статей известных экспертов, которых мы попросили разъяснить причины **разразившегося на прошлой неделе** мирового финансового **кризиса** и **дать прогнозы** на будущее» (fragment)

**Sentence 4** (see Fig.4)

The sentence contains two LF-collocations:

РЫНОК, LOC: НА,

СДВИГ, FUNC0: ПРОИСХОДИТЬ.

**Sentence 5** (see Fig. 5 and 6)

The sentence contains three LF-collocations:

ПОДЪЕМ, FUNC0: НАЧИНАТЬСЯ,

РЫНОК, LOC: НА,

НАПРЯЖЕНИЕ, CAUSFUNC0: ВЫЗЫВАТЬ.

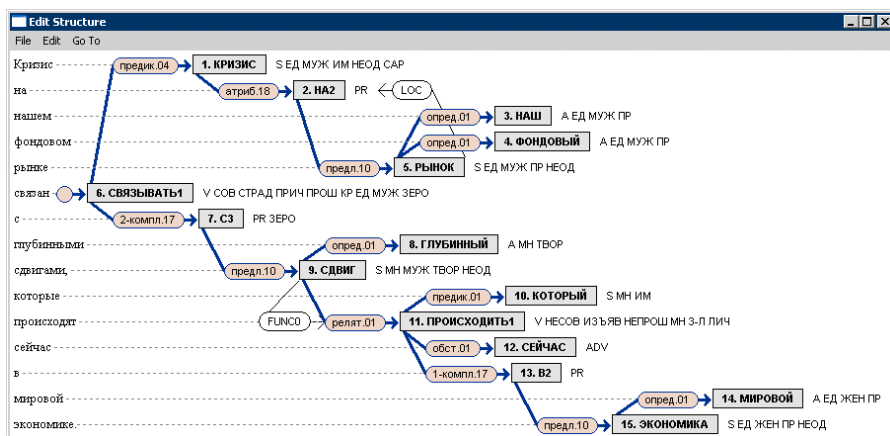
**Sentence 43** (see Fig. 7)

The sentence contains one LF-collocation:

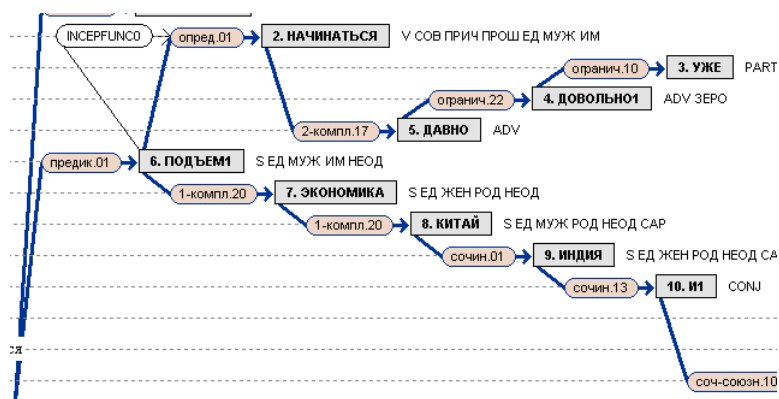
ДОХОДНОСТЬ, ANTIMAGN ('a small degree of X'): НИЗКИЙ.

**Sentence 44** (see Fig. 8)

The sentence contains one LF-collocation:  
ДОЛГ, INCEROPER1: ВЛЕЗАТЬ В.



**Fig. 4.** Syntactic and LF-structure of the sentence: «Кризис на нашем фондовом **рынке** связан с глубинными **сдвигами**, которые **происходят** сейчас в мировой экономике»



**Fig. 5.** Syntactic and LF-structure of the sentence: «Во-первых, **начавшийся** уже довольно давно **подъем** экономики Китая, Индии и других развивающихся стран усилил конкуренцию **на рынках** развитых стран и **вызвал** там определенное **напряжение** и с работой, и с доходами» (fragment 1)



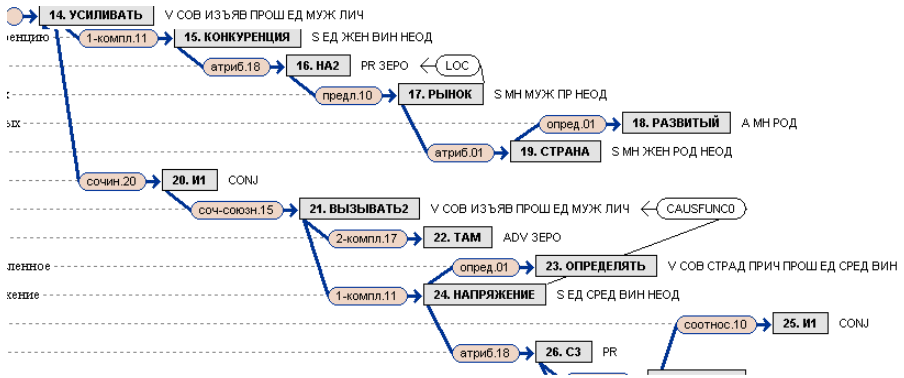


Fig. 6. Syntactic and LF-structure of the sentence: «Во-первых, **начавшийся** уже довольно давно **подъем** экономики Китая, Индии и других развивающихся стран усилил конкуренцию **на рынках** развитых стран и **вызвал** там определенное **напряжение** и с работой, и с доходами» (fragment 2)

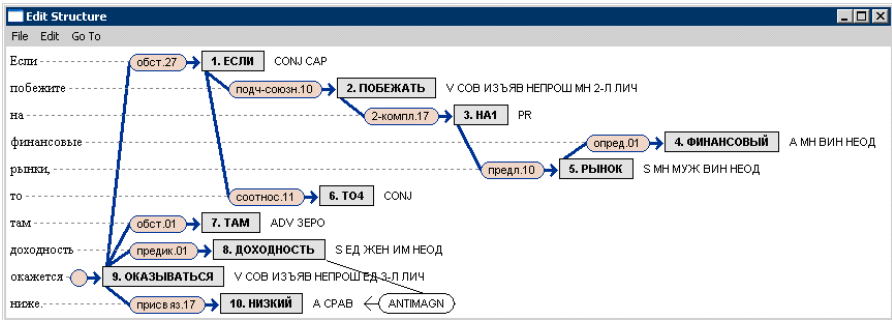


Fig. 7. Syntactic and LF-structure of the sentence: «Если победите на финансовые рынки, то там **доходность** окажется **ниже**»

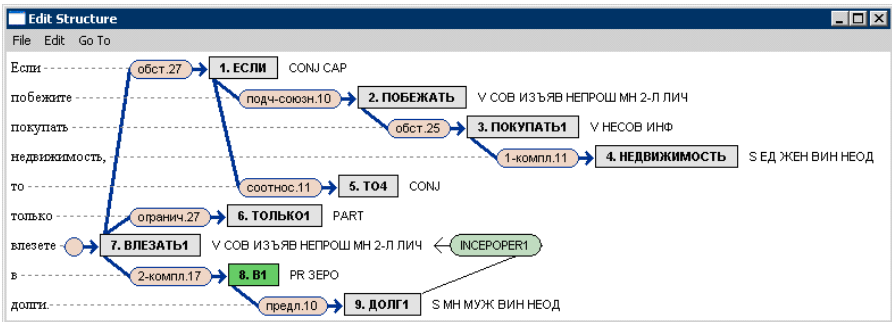


Fig. 8. Syntactic and LF-structure of the sentence: «Если победите покупать недвижимость, то только **влезете в долги**»

#### 4.1. Additional results. Improvements in ETAP-3 system

In the course of work some inaccuracies in description of not widespread syntactic constructions in rules of LF-recognition in ETAP-3 were revealed and corrected. Due to mass work with the texts, the LF-zones of both Russian and English combinatorial dictionaries of ETAP-3 were widened considerably. For more detailed description of changes in ETAP-3 see [11].

### 5. Perspectives and possible uses

The LF-tagging is still in progress. At the same time a new program for LF-search in texts is being elaborated. There is a program in Laboratory enabling LF-search in texts, but operating this program requires the knowledge of ways to record linguistic data in ETAP.

The new program is expected to let every user concerned to set and to solve different linguistic problems related to LFs. It is worth mentioning that search in LF-corpus will differ from search of collocations in corpora without syntactic and LF-tagging (see, for example, two dictionaries [12] and [13] published in the Internet in 2008). Search of Lf-collocations gives an opportunity to search not only adjacent words in certain types of phrases (such as NV, i.e. noun + verb, or VN, i.e. verb + noun, or VAN, i.e. verb + adjective + noun), but also phrases with free word order and long distances between words.

Collected corpus can be also used in education. See below several examples of possible uses of corpus in each of the two ways.

#### 5.1. Questions about use of LF-phrases in texts

1) It is obvious that the same LF often has several values for the same argument. For example, the noun МНЕНИЕ ('opinion') has the values of LF OPER1: ИМЕТЬ ('have') and ПРИДЕРЖИВАТЬСЯ ('hold to'). It is now possible to investigate the influence of lexical and syntactic environment on the choice of one of the two aforementioned variants. Here are some other examples for phrase with different values for the same LF and the same argument: *проводить исследование vs. вести исследование* ('conduct research'); *достигать результата vs. получить результат* ('obtain result'); *проводить эксперимент vs. ставить эксперимент vs. производить эксперимент vs. вести эксперимент vs. прodelьвать эксперимент* ('to carry out or conduct or perform or run experiment').

1.1) The same task can be set also for a class of arguments. For example, the same question seems to be sensible for values of LF FUNC2: ДОСТИГАТЬ ('reach'), РАВНЯТЬСЯ ('equal'), СОСТАВЛЯТЬ ('make up') for nouns denoting parameters: СКОРОСТЬ, ВЫСОТА, РАЗМЕР and so on.

2) It is well known that LFs can be used in paraphrasing. Special block of paraphrasing rules works in ETAP-3 system [see 6–7]. If the corpus with tagged LF-phrases is available the use of paraphrases can be investigated. For example, what are the differences that determine the choice between phrases *начал испытывать воздействие* ('started to be influenced') vs. *попал под воздействие* ('became influenced'), that is *начал* ('began') + *OPER1 X* vs. *INCEPOPER1 X*. Another example of such paraphrases is *испытывать давление* ('experience pressure') vs. *быть под давлением* ('be under pressure'), i. e. *OPER2 X* vs. *copula + ADV2 X*. One more example: *не*+LF ('not') vs. ANTILF (where LF is the value of any adjectival or some verbal LFs), cf. *медленное движение* ('slow motion') vs. *небыстрое движение* ('not quick motion'), that is *ANTIMAGN X* vs. *неMAGN X* ('not MAGN X').

3) There appears an opportunity to observe stylistic (or some other) peculiarities of the texts with more or less LF-phrases.

## 5.2. Uses of texts with LF-tagging for educational purpose

1) For students of Russian language the corpus may be used as a source of exercises for studying LF-collocations. For example, it is quite easy to pick out of corpus some sentences for the tasks like the following: — “fill the gap, marked with asterisks in the following sentence

*Евросоюз на минувшей неделе прописал евро радикальное лечение: надо \*\*\* ошибку, допущенную в 1999 году при введении единой валюты.*

with one of the verbs: УЛУЧШИТЬ, ПОЧИНИТЬ, ИСПРАВИТЬ”.

(The correct answer is *исправить*).

The translation of the sentence is: ‘Last week the European Union prescribed euro curative treatment: it is necessary to correct the mistake made in 1999 in the time of creating the single currency’.

To make such exercises it is enough to pick out of the corpus sentences with LFs and put asterisks instead of values of LF.

2) For students of linguistics, studying the theory of LFs the same sentences may be a source of another two types of exercises:

a) “identify the name of LF in boldface phrase

*Евросоюз на минувшей неделе прописал евро радикальное лечение: надо **исправить ошибку**, допущенную в 1999 году при введении единой валюты”.*

(The correct answer is REAL1-M).

b) “fill the gap, marked with asterisks in the following sentence with value of LF REAL1-M for the boldface argument

*Евросоюз на минувшей неделе прописал евро радикальное лечение: надо \*\*\* **ошибку**, допущенную в 1999 году при введении единой валюты”.*

(The correct answer is *исправить*).

## References

1. *Apresian Iu. D., Boguslavskii I. M., Iomdin L. L., Lazurskii A. V., Pertsov N. V., San'nikov V. Z., Tsinman L. L.* 1989. Linguistic Supply of the System ETAP-2 [Lingvističeskoe Obespečenie Sistemy EETAP-2].
2. *Apresian Iu. D., Boguslavskii I. M., Iomdin L. L., Tsinman L. L.* 2007. Lexical Functions in Actual NLP-Applications. Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In honour of Igor Meľcuk, 84 : 199–230.
3. *Apresian Iu. D., Boguslavskii I. M., Iomdin L. L., Tsinman L. L.* 2002. Lexical Functions in NLP: Possible Uses. Computational Linguistics for the New Millennium: Divergence or Synergy? Festschrift in Honour of Peter Hellwig on the occasion of his 60th Birthday : 55–72.
4. *Apresian Iu. D., Diachenko P. V., Lazurskii A. V., Tsinman L. L.* 2007. On Electronic Textbook for Russian Lexic Studies [O Komp'juternom Učebnike Leksiki Russkogo Iazyka]. *Russkii Iazyk v Nauchnom Osveshčenii* : 48–112.
5. *Biriuk O. L., Gusev V. Iu., Kalinina E. Iu.* Dictionary of Russian Non-Object Nouns' Verbal Compatibility [Slovar' Glagol'noi Sochetanosti Nepredmetnykh Imen Russkogo Iazyka]. Slovar' na osnove Natsional'nogo Korpusa Russkogo Iazyka dict.ruslang.ru.
6. *Boguslavskii I. M., Chardin I. S., Grigor'eva S. A., Grigor'ev N. V., Iomdin L. L., Frolova T. I., Shemanaeva O. Iu.* 2010. Lexical-Functional Texts Markup in SinTagRus [Leksiko-funktsional'naiia Razmetka Teksta v SinTagRus]. *Informatsionnye Tekhnologii i Sistemy (it is'10): Sbornik Trudov Konferentsii, (Proc. of the Conference "Information Technologies and Systems")* : 320–324.
7. *Boguslavskii I. M., Grigor'ev N. V., Grigor'eva S. A., Iomdin L. L., Kreidlin L. G., Frid N. E.* 2002. Development of Syntactically Marked Out Russian Corpus [Razrabotka Sintaksicheskii Razmečennogo Korpusa Russkogo Iazyka]. *Doklady Nauchnoi Konferentsii "Korpusnaia Lingvistika i Lingvisticheskie Bazy Danykh"*, (Proc. of Scientific Conference "Corpus Linguistics and Linguistic Databases"): 40–50.
8. *Kreydlin L. G., Frid N. E.* 2002. Development of a Dependency Treebank for Russian and its Possible Applications in NLP. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC-2002)*, III : 852–856.
9. *Kustova G. I.* Russian Idiomatic Expressions Dictionary. Words Combinations with the Meaning of a High-Scale [Sochetaniia Slova so Znacheniem Vysokoi Stepeni]. Slovar' na osnove Natsional'nogo Korpusa Russkogo Iazyka dict.ruslang.ru.
10. *Meľchuk I. A.* 1974. Experiment of the Theory of Linguistic Models "Meaning - Text" [Opyt Teorii Lingvisticheskikh Modelei "Smysl - Tekst"].
11. *Sizov V. G., Tsinman L. L.* 2003. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT. *MTT 2003, First International Conference on Meaning — Text Theory* : 279–288.
12. *Zholkovskii A. K., Meľchuk I. A.* On Semantic Synthesis [O Semanticheskom Sinteze]. *Problemy Kibernetiki*, 19 : 177–238.