

A DATABASE OF RUSSIAN VERBAL FORMS AND THEIR FRENCH TRANSLATION EQUIVALENTS¹

Kruzhkov M. G. (magnit75@yandex.ru)
IPI RAS, Moscow, Russia

Buntman N. V. (nabunt@hotmail.com)
MSU, Moscow, Russia

Loshchilova E. Ju. (lena0911@mail.ru)
IPI RAS, Moscow, Russia

Sitchinava D. V. (mitrius@gmail.com)
IRL RAS, Moscow, Russia

Zalisniak A. A. (anna.zalizniak@gmail.com)
IL RAS, IPI RAS, Moscow, Russia

Zatsman I. M. (izatsman@yandex.ru)
IPI RAS, Moscow, Russia

The paper presents the results of a project² aimed at the development of methodology and information technology for the creation of a corpus-based linguistic database of verbal forms with their translation equivalents (with bilingual grammatical search functions). Within the scope of the project the following results have been achieved:

1. *Methodology and information technology* for the creation of linguistic databases based on bilingual parallel corpora have been developed (including corpora with multiple translation variants).
2. The *polyvariant parallel subcorpus* which includes Russian literary works with French translations has been created within the Russian National Corpus (RNC). Some of the parallel texts in the subcorpus include multiple translation variants.
3. On the basis of the polyvariant Russian-French corpus a *database of Russian verbal lexico-grammatical forms* (LGFs) and their French translation equivalents has been created. Equipped with bilingual grammatical search functions, the database is a unique resource that can be used for investigating a wide range of various cross-linguistic problems.
4. A number of concepts in the areas of Russian verbal categories and Russian-French contrastive grammar have been refined.

Key words: polyvariant parallel corpus, grammatical semantics, contrastive grammar, database of verbal forms, corpus linguistics, bilingual grammatical search, Russian, French

¹ The work was carried out at the Institute of Informatics Problems of the RAS.

² The project "Information technology for creation of corpus-based linguistic database of verbal forms with their functional equivalents" was supported by "Dinastija" foundation, grant NG13-036.

1. Introduction

The emergence of electronic corpora has marked the beginning of a new era in contrastive linguistics. Stig Johansson's work with English-Norwegian parallel corpus in 1990s was ground-breaking in this field [1]. Combining the methodological advantages of computer corpus linguistics with the possibility of contrasting parallel texts in two or more languages allowed to compare the actual use of the languages involved at all levels of descriptions—with greater accuracy and detail than had been possible before. During the past two decades considerable advances have been made in this field, both in development of analytical methods and in creation of unique lexicographic descriptions. Further avenues of research in contrastive grammatical studies are discussed in the following works: [1, 2, 3, 4, 5, 6, 7].

The technology of compilation of the Russian-French polyvariant parallel subcorpus of the Russian National Corpus (RNC) is presented in [8]. The subcorpus contains Russian literary works aligned with French translations with current total volume of 2 million words. Some of the parallel texts are offered in *the polyvariant format*, i.e. a single text in Russian is aligned with more than one translation of the same text into French. Total volume of the polyvariant texts is currently 700 thousand words—more than a third of the total volume of the Russian-French subcorpus of the RNC.

Parallel subcorpora were first introduced into the RNC in 2005 [2, 4, 9, 10]. Currently the RNC includes 8 bilingual parallel subcorpora with Russian as source or target language: English, German, French, Spanish, Italian, Polish, Ukrainian and Byelorussian ones. There is also one multilingual parallel subcorpus. The Russian-French parallel subcorpus was added to the RNC in December 2012. The technology used for the compilation of this subcorpus (presented in [8]) hereafter is referred to as Parallel Corpus technology or PC-technology.

In 2013 PC-technology was extended with the introduction of new operations designed for creation of *the Database of Russian Verbal Forms and their French Translation Equivalents* (hereafter—the DB)³ and for the compilation of the polyvariant Russian-French subcorpus (hereafter—the subcorpus). The extended technology allows to simultaneously compile the subcorpus and to fill the DB. In addition the extended technology implements bilingual grammatical search functions for the verbal forms and their translations in the DB. For instance, in the DB a user can browse all Russian verbal forms in Russian present tense which are translated into French by *passé composé*. The new extended technology hereafter is referred to as Database Parallel Corpus technology or DBPC-technology.

The goal of this paper is to describe the purpose of the DB, which is based on the parallel texts from the subcorpus, and to illustrate the bilingual grammatical search function implemented in the DB.

³ The list of the database contributors is as follows: N. Buntman, B. Loktev, V. Nuriev, O. Pe-trushkina, N. Popkova, E. Roganova, E. Spiridonova, V. Stepanov, A. Shchurova.

2. Purpose and methodology

The methodology of the database construction is primarily defined by its purpose. The DB was created as a tool that should allow to describe Russian grammatical semantics “as mirrored in French” and to clarify certain concepts in Russian-French contrastive grammar. In the methodology development we relied on works by V. G. Gak, I. N. Kouznetsova, M. Guiraud-Weber [11, 12, 13, 14] and other authors. But these works were created in the pre-corpus era; today when Russian-French corpora are regularly compiled and updated, we can rely on parallel texts analysis to update descriptions of Russian-French contrastive grammar.

While developing the DB we kept in mind that the main object of analysis for linguistic experts working with the DB was the correspondences between Russian and French verbal categories in parallel texts. In order to properly describe the analyzed correspondences, a number of terms have been defined capturing the essence of the developed methodology [15, 16, 17]

Among the key notions are *lexico-grammatical form (LGF)* and *basic LGF type*. Basic LGF type is understood as a certain *combination of grammatical features* along with certain elements in the sentence structure which define a certain “construction”⁴; consider, for example, basic LGF type “PastPF + *если бы*” (= past tense, perfective aspect + *если бы*); cf. also 3rd and 5th column on Fig.1. Accordingly, LGF is understood as a *combination of elements of a sentence* which realizes a given set of features, for example (elements of the LGF are marked in bold): *если бы он пришел вовремя*; cf. also 2nd and 4th column on Fig. 1.

For Russian 15 basic LGF types were specified (see Table 1)⁵. This is the so-called Source Set that restricts the initial search of Russian LGFs in the DB on the first stage of DBPC-technology. The scope of French basic LGF types is not limited: it is continually expanded as the new types of translation variants are identified in the DB. At present, the experimental version of the DB includes 25 French basic LGF types (see Table 2).

Table 1. The Source Set: basic verbal LGF types (Russian)

1.	Present	Pres-IPF
2.	Past Imperfective	Past-IPF
3.	Past Perfective	Past-PF
4.	Simple Future	Fut-PF
5.	Compound Future	Fut-IPF
6.	Imperative Perfective	Imperat-PF
7.	Imperative Imperfective	Imperat-IPF
8.	Form with <i>бы</i> PF	Past-PF+ <i>бы</i>

⁴ As this term is understood in the Construction Grammar [15–17].

⁵ The actual DB includes only LGFs with finite verbal forms (i.e. impersonal verbs, participles, periphrases with the verb *быть* are not included). In the future the range of examined types of verbal forms will be expanded.

9.	Form with <i>бы</i> IPF	Past-IPF+ <i>бы</i>
10.	Form with <i>если бы</i> PF	Past-PF+ <i>если бы</i>
11.	Form with <i>если бы</i> IPF	Past-IPF+ <i>если бы</i>
12.	Form with <i>чтобы</i> PF	Past-PF+ <i>чтобы</i>
13.	Form with <i>чтобы</i> IPF	Past-IPF+ <i>чтобы</i>
14.	Form with <i>было</i> PF	Past-PF+ <i>было</i>
15.	Form with <i>было</i> IPF	Past-IPF+ <i>было</i>

Table 2. The Target Set: basic LGF types (French)

1.	présent	Pr
2.	passé composé	PasCom
3.	passé simple	PasSim
4.	imparfait	Imparf
5.	plus-que-parfait	PqParf
6.	passé antérieur	PasAnt
7.	passé immédiat	PasIm
8.	futur simple	Fut
9.	futur antérieur	FutAnt
10.	futur immédiat	FutIm
11.	impératif	Imperat
12.	subjonctif présent	SubjPres
13.	subjonctif passé	SubjPas
14.	subjonctif imparfait	SubjImparf
15.	subjonctif plus-que-parfait	SubjPqParf
16.	conditionnel présent	CondPr
17.	conditionnel passé	CondPas
18.	participe présent	PartPr
19.	participe passé	PartPas
20.	participe passe compose	PartPasComp
21.	gérondif	en PartPr
22.	infinitif	Inf
23.	préposition+infinitif	Prep+Inf
24.	préposition+infinitif passé	Prep+InfPas
25.	substantif	Subst

Apart from the features that define the basic LGF types, lists of additional features has been compiled for each of the two languages. Additional features allow to make a further specification of the type of construction. They define either the composition of the verbal group (e.g. presence of a subordinate infinitive, a modality marker, a negation marker), or the type of the sentence in which the LGF is used (e.g. subordinate clause, interrogative sentence, direct speech), see Tables 3 and 4. Each additional feature can apply to one or more basic LGF types. On Figures 1, 2, 5, 6

additional features are specified in square brackets after the basic LGF type. *LGF type* is defined as combination of a basic LGF type and a set of relevant additional features.

Table 3. Additional features for basic LGF types (Russian)

Subordinate infinitive PF	[SubInf-PF]
Subordinate infinitive IPF	[SubInf-IPF]
Modality marker	[ModDet]
Negation	[Neg]
Interrogative sentence	[Interrog]
Exclamatory sentence	[Exclam]
Verb introducing direct speech	[VerbDirSp]
Verb inside direct speech	[DialRepl]
Verb in complement clause	[SubCompl]
Verb in attributive clause	[SubAttr]
Verb in subordinate clause	[Sub]

Table 4. Additional features for basic LGF types (French)

Subordinate infinitive	[SubInf]
Subordinate past infinitive	[SubInfPas]
Subordinating predicate added	[+SuperPred]
Modality marker	[ModDet]
Negation	[Neg]
Interrogative sentence	[Interrog]
Exclamatory sentence	[Exclam]
Verb introducing direct speech	[VerbDirSp]
Verb inside direct speech	[DialRepl]
Verb in complement clause	[SubCompl]
Verb in attributive clause	[SubAttr]
Verb in subordinate conditional clause	[SubCond]
Verb in subordinate clause	[Sub]
Accusativus cum infinitivo	[Acc.c.Inf]

The process of establishing correspondences between Russian and French LGFs in aligned parallel texts is carried out as follows. First, an expert marks in the Russian text a fragment corresponding to one of the 15 *basic LGF types* from the Source Set, i.e. one Russian *LGF*. Then, the expert looks for its “functionally equivalent fragment” (FEF)⁶ in the aligned translated fragment, marks it and matches it to an appropriate *basic LGF type* in the Target Set. If the needed unit is not listed in the Table 2, the Target Set can be expanded. If a FEF cannot be located in the translation for a certain

⁶ The term “functionally equivalent fragment” was introduced in [2, 9].

Russian LGF, the ME is marked as “Nondetermined” in the DB and is not taken into account when processing the data⁷.

A pair of fragments obtained in this way is referred to as a *monoequivalence* (ME, see definition below), see Fig. 1. Extraction of LGFs and FEFs and matching them to appropriate LGF types and to each other is the initial task that is supported by the DB. Before we can pass on to further tasks, we have to define a few new terms in the area of contrastive analysis: *monoequivalence* (ME), *type of ME*, *polyequivalence* (PE), *type of PE* and *hyperequivalence* (HE).

ME#	Russian LGF	Russian LGF type	French LGF	French LGF type
4711	потом [...] плотно запер все двери	Past-PF [ModDet]	après avoir bien fermé toutes les portes.	InfPas [Prep+InfPas] [Sub]

Fig. 1. A sample ME from the DB

Monoequivalence is a pair $\langle Rn(i); Fm(j) \rangle$, where $Rn(i)$ is a specific LGF of Russian basic LGF type Rn (see Table 1) in the original text, and $Fm(j)$ is a specific LGF of French basic LGF type Fm (see Table 2) in one of the translations. All LGFs in the DB have unique identifiers, so in this case specific LGFs are uniquely identified by indexes i and j .

Type of ME $\langle Rn(i); Fm(j) \rangle$ is the pair of the corresponding basic LGF types $\langle Rn; Fm \rangle$, e.g. for the ME represented on the Fig.1 it is: $\langle \text{Past-PF}; \text{InfPas} \rangle$.

Polyequivalence $\langle Rn(i); \{Fm(j), Fk(r), \dots\} \rangle$ is a combination of monoequivalences $\langle Rn(i); Fm(j) \rangle$, $\langle Rn(i); Fk(r) \rangle$ etc. with identical Russian LGF in the first position. A PE reflects different variants of translation of the same original LGF (see Fig. 2).

Type of PE $\langle Rn(i); \{Fm(j), Fk(r), \dots\} \rangle$ is the pair $\langle Rn; \{Fm, Fk, \dots\} \rangle$ e.g. $\langle \text{Present}; \{\text{Présent}, \text{Présent}\} \rangle$ (see 2nd and 5th columns of Fig. 2).

Hyperequivalence is a pair $\langle Rn; \{F\} \rangle$, which represent aggregation of all possible *types of ME* in the DB with the same value at the first position. It comprises of one Russian basic LGF type Rn and a multitude of French basic LGF types $\{F\}$ that enter into MEs with Russian LGFs of basic LGF type Rn .

Based on the terms defined above we enumerate the tasks that the developed DB is meant to accomplish:

- building of MEs, PEs and HEs;
- bilingual grammatical search of MEs and PEs;
- calculating frequencies for each type of ME and PE in the DB.

⁷ Here we refer to such cases when the lexical items used to translate the semantic content enclosed in the original LGF are substantially different from the original, to the extent that it is impossible to establish a correspondence between the LGFs using the existing apparatus. E.g.: ты [...] так теребишь за носы, что еле держатся—tu tirais tellement sur leur nez [...] que tu as failli le leur arracher.

Russian LGF	Russian LGF type	LGFs in French translations		
		ME#	French LGFs	French LGF Types
Я иногда в театр хожу	Pres-IPF [ModDet] [DialRepl]	596	Il m'arrive d'aller au théâtre,	Pr [SubInf] [+SuperPred] [DialRepl]
		5927	Non, je vais parfois au théâtre, et en visite.	Pr [ModDet] [DialRepl]

Fig. 2. A sample PE combined from two MEs⁸

In order to facilitate accomplishment of these tasks we have developed a web interface that allows users (linguistic experts) to interact with the DB using the common web browsers (Internet Explorer, Mozilla Firefox, Google Chrome).

The DB functions can be divided into two major groups: the first group supports building and editing of MEs (see Fig. 3), and the second group supports viewing and searching of MEs and PEs (for PE search functions see Fig. 4).

Building and editing functions allow filtering of the aligned fragments of Russian and French texts by book title, translator and basic LGF types that the user looks for in the original text. The users then browse the selected pairs of aligned fragments in order to find LGFs and to build MEs.

By the beginning of 2014, 10527 MEs have been built, and 4128 PEs have been automatically generated by matching MEs with the same Russian LGF at the first position.

3. Bilingual grammatical search functions

The PE search page allows users to view collections of PEs (Fig. 4) that are generated according to the specified search queries. Users can filter all the PEs in the DB using such search features as original book title, Russian and French LGF types, specific lexemes or text fragments in the original text or in the translation, etc. The search features can be specified separately or in any combinations. After the query is executed, the user can see the number of selected PEs and browse the found PEs.

⁸ The French LGFs in this PE belong to the same basic LGF type but they have different additional features.

4255. **чтоб он пошел выше.** - **qu'il pût même s'élever plus haut.**

Номер пары	Оригинальный текст	Перевод
883	Ни ему самому [и] [никому] [и] [никому] [и] [и] [ктоб] [он] [пошел] [выше]	Ni lui-même, ni les autres n'avaient su imaginer qu'il pût même s'élever plus haut.

Редактирование ЛФ, составляющих МЭ

Изменить ЛФ ориг.

Выборать главные слова

[чтоб] [и] [он] [пошел] [выше]

Ок

ЛФ языка оригинала

Текст ЛФ	Код ЛФ	Вид ЛФ	Дополн. признаки ЛФ
чтоб он пошел выше.	Past-PF-чтобы	Форма с «чтобы» СВ	<input type="checkbox"/> SubInf.PF <input type="checkbox"/> SubInf.PF <input type="checkbox"/> ModDet <input type="checkbox"/> Neg <input type="checkbox"/> VerbDirSp <input type="checkbox"/> SubAttr <input type="checkbox"/> Sub-Compl <input checked="" type="checkbox"/> Sub <input type="checkbox"/> Interrog <input type="checkbox"/> Exclam <input type="checkbox"/> DualRepl

Изменить ЛФ перев.

Выборать главные слова

[qu] [pût] [même] [s] [élever] [plus] [haut]

Ок

ЛФ языка перевода

Текст ЛФ	Код ЛФ	Вид ЛФ	Дополн. признаки ЛФ
qu'il pût même s'élever plus haut.	SubImparf	subjonctif imparfait	<input checked="" type="checkbox"/> SubInf <input type="checkbox"/> SubInfPas <input type="checkbox"/> Acc.c. Inf <input type="checkbox"/> +SuperVerb <input type="checkbox"/> ModDet <input type="checkbox"/> Neg <input type="checkbox"/> SubAttr <input type="checkbox"/> Sub-Compl <input type="checkbox"/> Si-Cond <input type="checkbox"/> Sub

Fig. 3. Web interface for editing MEs

просмотр полнотекстовых записей - Google Chrome
 a179.ipi.as.ru/corpora/PolyEquivalence.aspx?userid=58&projectid=1

Поиск полнотекстовых записей по параметрам

Книги	Виды ЛПФ оригинала <input type="checkbox"/> Исключить	Виды ЛПФ перевода 1 <input type="checkbox"/> Исключить	Виды ЛПФ перевода 2 <input type="checkbox"/> Исключить	Лексема в оригинале
Признак наличия различных базовых видов ЛПФ в разных переводах	Доп. признаки ЛПФ оригинала <input type="checkbox"/> Исключить	Доп. признаки ЛПФ перевода 1 <input type="checkbox"/> Исключить	Доп. признаки ЛПФ перевода 2 <input type="checkbox"/> Исключить	Лексема в переводе 1 Лексема в переводе 2
	Текст из ЛПФ оригинала	Текст из ЛПФ перевода 1	Текст из ЛПФ перевода 2	<input type="checkbox"/> в точной форме <input type="checkbox"/> в "главных" словах

[Показать ЛПФ с заданными параметрами](#) [Сброс](#)

Список полнотекстовых записей

(Всего записей: 4136)
 Страница 1 из 42 из 45 из 20 из 100

Номер ЛПФ ориг.	Текст ЛПФ оригинала ПЭ	Вид ЛПФ оригинала ПЭ	Доп. признаки ЛПФ оригинала	ЛПФ переводов ПЭ															
328	Наконец достал он свое исподнее платье	Прошедшее СВ		<table border="1"> <tr> <th>Номер МЭ</th> <th>Текст ЛПФ перевода</th> <th>Вид ЛПФ перевода</th> <th>Дополн. признаки ЛПФ перевода</th> <th>Дополн. Комментарий к МЭ</th> </tr> <tr> <td>181</td> <td>Finalment, il prit son linge de dessous</td> <td>passé simple</td> <td></td> <td></td> </tr> <tr> <td>182</td> <td>il [...] enfila toutes ces hardes, simple</td> <td>passé simple</td> <td></td> <td></td> </tr> </table>	Номер МЭ	Текст ЛПФ перевода	Вид ЛПФ перевода	Дополн. признаки ЛПФ перевода	Дополн. Комментарий к МЭ	181	Finalment, il prit son linge de dessous	passé simple			182	il [...] enfila toutes ces hardes, simple	passé simple		
Номер МЭ	Текст ЛПФ перевода	Вид ЛПФ перевода	Дополн. признаки ЛПФ перевода	Дополн. Комментарий к МЭ															
181	Finalment, il prit son linge de dessous	passé simple																	
182	il [...] enfila toutes ces hardes, simple	passé simple																	
1194	тепло не чувствует его на себе;	Настоящее	Отрицание	<table border="1"> <tr> <th>Номер МЭ</th> <th>Текст ЛПФ перевода</th> <th>Вид ЛПФ перевода</th> <th>Дополн. признаки ЛПФ перевода</th> <th>Дополн. Комментарий к МЭ</th> </tr> <tr> <td>564</td> <td>cette robe de chambre [...] ne pesait pas sur le</td> <td>imparfait</td> <td>Отрицание</td> <td>Смена подлежащего</td> </tr> </table>	Номер МЭ	Текст ЛПФ перевода	Вид ЛПФ перевода	Дополн. признаки ЛПФ перевода	Дополн. Комментарий к МЭ	564	cette robe de chambre [...] ne pesait pas sur le	imparfait	Отрицание	Смена подлежащего					
Номер МЭ	Текст ЛПФ перевода	Вид ЛПФ перевода	Дополн. признаки ЛПФ перевода	Дополн. Комментарий к МЭ															
564	cette robe de chambre [...] ne pesait pas sur le	imparfait	Отрицание	Смена подлежащего															

Fig. 4. Web interface for viewing and searching PEs

The bilingual grammatical search which can be applied to one or more translations (a polyvariant bilingual query) is a fundamentally new research tool. For instance, we can specify a Russian basic LGF type (*Past-PF*) and two different French

basic LGF types for two translation variants (*CondPr* and *PasSim*). Such a query will result in two PEs being found (see Fig. 5).

он решил оставить [...] липовые и дубовые деревья	Past-PF [SubInf]	ME#	Translation LGF	Translation LGF type
		2931	alors qu'il garderait les [...] tilleuls et chênes,	CondPr
		8011	Il décida de laisser tels quels les [...] tilleuls et les chênes,	PasSim [SubInf]
он решил [...] яблони и груши уничтожить	Past-PF [SubInf]	ME#	Translation LGF	Translation LGF type
		2932	il se débarrasserait des pommiers et des poiriers	CondPr
		8013	Il décida [...] de supprimer les pommiers	PasSim [SubInf]

Fig. 5. 2 PEs found in the DB by specifying the basic LGF types and two translation variants

Apart from the basic LGF type, the user can specify additional features of Russian and French LGFs (see Tables 3 and 4). For instance, we can specify Russian LGF type *Pres [SubInf-PF]* and French LGF Type *CondPr [SubInf]* in at least one of the translation variants. Such a query will result in three PEs being found (see Fig 6).

4. Conclusion

The created DB allowed us to clarify some concepts in Russian-French contrastive grammar. In particular, the list of correspondences described in [11, 12] and summarized in [13] has been:

- inverted (in works by Gak and Kouznetsova language material was examined from the viewpoint of translating French texts into Russian because their goal was interpretation of meanings and functions of French forms);
- expanded, i.e. new types of translation correspondences have been established.

Of particular interest are the results of the frequency analysis of translational correspondences. For example, correlation between oppositions “perfective vs. imperfective aspect” in Russian and “passé composé/passé simple vs. imparfait” in French can be refined based on quantitative indicators. Russian basic LGF type Past-PF only in 49,4% of cases corresponds to French basic LGF type Imparf and in 21% of cases to PasCom/PasSim. These figures highlight the width of the semantic range of the Russian imperfective aspect.

Не может постараться для барина!	Pres [SubInf- PF] [Neg]	ME#	Translation LGF	Translation LGF type
		661	Tu pourrais tout de même faire un effort pour ton maître!	CondPr [SubInf] [Exclam]
		5897	Il ne peut même pas faire un petit effort pour son maître!	Présent [SubInf] [Exclam]
теперь можете отдать	Pres [SubInf- PF]	ME#	Translation LGF	Translation LGF type
		945	maintenant vous pouvez me rembourser.	Présent [SubInf]
		7584	alors vous pourriez peut-être me rembourser?	CondPr [SubInf] [ModDet] [Interrog]
Разве я могу все это [...] перенести?	Pres [SubInf- PF] [Interrog]	ME#	Translation LGF	Translation LGF type
		8939	Est-ce que je puis [...] le supporter?	Présent [SubInf] [Interrog]
		8940	Je pourrais [...] supporter tout ça?	CondPr [SubInf] [Interrog]

Fig. 6. 3 PEs found in the DB by specifying the basic LGF types with additional features

Furthermore, the DB based on the polyvariant subcorpus allows to clarify the semantics of Russian verbal forms: French translation variants with more detailed network of grammatical positions in the domain of tense/mood features make it possible to detect specific semantic components in Russian LGFs.

The DB creation has confirmed the efficiency of “construction” as a tool of linguistic analysis: the contrastive approach based on the term “LGF” allows to bring to light various relationships between actional, temporal, aspectual and modal components in meanings of Russian verbal forms.

Finally, the developed DBPC-technology is readily adoptable to be used in other cross-linguistic projects based on parallel aligned texts (i.e. projects dedicated to investigation of other categories of LGFs, not necessary verbal ones). The customization can be accomplished without significant changes to the structure of the DB. To customize the DBPC-technology one should basically supply the list of languages used and the lists of basic LGF types and additional features for these languages according to goals and objectives of a specific project.

Currently the DBPC-technology is being adopted for investigation of Russian language-specific units.

References

1. *Aijmer K., Altenberg B.* (2013), *Advances in Corpus-based Contrastive Linguistics. Studies in honour of Stig Johansson, John Benjamins, Amsterdam.*
2. *Dobrovol'skij D. O., Kretov A. A., Sharov S. A.* (2005), *Parallel text corpus [Korpus paralel'nyh tekstov], Scientific and technical information. Ser.2. Information processes and systems [Nauchnaja i tehničeskaja informacija. Ser. 2. Informacionnye processy i sistemy], Vol. 6, pp. 16–27.*
3. *Kiseleva K. L., Plungjan V. A., Rahlina E. V., Tatevosov S. G.* (Eds.) (2009), *Corpus-based research on Russian grammar [Korpusnye issledovanija po russkoj grammatike], Probel-2000, Moscow.*
4. *Dobrovol'skij D. O.* (2009), *Parallel text corpus in studying of culture-specific words [Korpus paralel'nyh tekstov v issledovanii kul'turno-spezifichnoj leksiki], Russian National Corpus: 2006–2008. New results and prospects [Nacional'nyj korpus russkogo jazyka: 2006–2008. Novye rezul'taty i perspektivy], Nestor-Istorija, St. Petersburg, pp. 383–401.*
5. *Sichinava D. V.* (2011), *Complex study of monolingual and parallel text corpora in grammatical studies [Kompleksnoe issledovanie odnojazychnogo i paralel'nogo korpusov v grammatičeskikh issledovanijah], Proceedings of the international conference “Corpus Linguistics—2011” [Trudy mezhdunarodnoj konferencii «Korpusnaja lingvistika—2011»], St. Petersburg, pp. 316–322.*
6. *Sichinava D. V., Shvedova M. A.* (2010), *Parallel corpora within Russian National Corpus: technologies and the addressed issues [Paralel'nye korpusa v sostave Nacional'nogo korpusa russkogo jazyka: tehnologii i reshaemye zadachi], Computational linguistics: areas of study and research [Komp'juternaja lingvistika: nauchnoe napravlenie i uchebnaja disciplina], Gomel', GGU imeni F. Skoriny, pp. 30–34.*
7. *Sichinava D. V., Arhangel'skij T. A.* (2012), *Belarusian-Russian and Russian-Belarusian parallel corpora: Russian National Corpus collaboration [Paralel'nye belorussko-russkij i russko-belorusskij korpusa: sovmestnyj proekt Nacional'nogo korpusa russkogo jazyka], TEL-2012 workshop proceedings [Trudy shkoly-seminara TEL-2012], Kazanskij (Privolzhskij) federal'nyj universitet, Kazan'.*
8. *Zalizniak Anna A., Sitchinava D. V., Loiseau S., Kruzhkov M., Zatsman I. M.* (2013), *Database of Equivalent Verbal Forms in a Russian-French Multivariant Parallel Corpus, 2013 International Conference on Artificial Intelligence (ICAI'13) Vol. I, Las Vegas, pp. 101–107.*
9. *Dobrovol'skij D. O., Kretov A. A., Sharov S. A.* (2005), *Parallel text corpus: architecture and possible application [Korpus paralel'nyh tekstov: arhitektura i vozmožnosti ispol'zovanija], Russian National Corpus: 2003–2005 [Nacional'nyj korpus russkogo jazyka: 2003–2005], Indrik, Moscow, pp. 263–296.*
10. *Andreeva E. G., Kasevich V. B.* (2005), *Grammar and Lexis (evidence from English-Russian parallel text corpus) [Grammatika i leksika (na materiale anglo-russkogo korpusa paralel'nyh tekstov)], Russian National Corpus: 2003–2005 [Nacional'nyj korpus russkogo jazyka: 2003–2005], Indrik, Moscow, pp. 297–307.*
11. *Gak V. G.* (2006), *The Russian language in comparison with the French language [Russkij jazyk v sopostavlenii s francuzskim], URSS, Moscow.*

12. *Gak V. G.* (2009), *Comparative typology of the French language and the Russian language* [Sravnitel'naja tipologija francuzskogo i ruskogo jazykov], URSS, Moscow.
13. *Kouznetsova I. N.* (2009), *Contrastive Grammar of French and Russian* [Grammaire contrastive du francais et du russe], Nestor Academic Publishers, Moscow.
14. *Guiraud-Weber M.* (2011), *Essays on Russian and contrastive syntax* [Essais de syntaxe russe et contrastive], Université de Provence, Aix-en-Provence.
15. *Rahilina E. V.* (red.) (2010), *Construction linguistics* [Lingvistika konstrukcij], Azbukovnik, Moscow.
16. *Goldberg A.* (1995), *Constructions: A Construction Grammar Approach to Argument structure*, Univ. of Chicago Press, Chicago.
17. *Goldberg A.* (2006), *Constructions at Work. The nature of generalization in grammar*, Oxford Univ. Press, Oxford.