

# Database of Russian Verbal Forms and their French Translation Equivalents

Kruzhkov M. G. (magnit75@yandex.ru), IPI RAS,  
Buntman N. V. (nabunt@hotmail.com), MSU,  
Loshchilova E. Ju. (lena0911@mail.ru), IPI RAS,  
Sitchinava D. V. (mitrius@gmail.com), IRL RAS,  
Zalisniak A. A. (anna.zalizniak@gmail.com), IL RAS, IPI RAS,  
Zatsman I. M. (izatsman@yandex.ru) IPI RAS

- The goal of our project was to investigate grammatical semantics of Russian verbal forms “as mirrored in French” and to clarify certain concepts in Russian-French contrastive grammar based on data from Russian-French parallel subcorpus of Russian National Corpus (RNC).
- Current functionalities of the RNC are insufficient for completing this goal because one cannot easily identify grammatical constructions in the RNC. Thus arises the need for a methodology that allows to identify grammatical constructions in parallel texts and to establish correspondences between them.

- The methodology for creation of the Database of Russian Verbal Forms and their French Translation Equivalents was created as an extension of the methodology for compilation of the Russian-French polyvariant subcorpus within the RNC.
- The Database is based on data from the polyvariant parallel subcorpus. It allows to produce a large searchable array of fragments aligned on the level of constructions (in our case, verbal ones). Such constructions in our project are called lexicogrammatical forms, or LGFs.

- A monoequivalence (ME) is a correspondence between one LGF in the original text and one of its functionally equivalent fragments (FEF) in French translations. Currently the Database contains over 10000 MEs.
- In case of a polyvariant corpus, Polyequivalences (PEs) may be automatically generated by the Database.
  - The Database provides bilingual lexicogrammatical search functions for both MEs and PEs. It also can aggregate verifiable statistical information on types of MEs and PEs.

- The proposed methodology may be used in other researches related to cross-linguistic investigations of constructions. To do so researchers will have to redefine typologies of constructions for each language that they wish to use according to their goals.
- We are also investigating the possibility of integrating this methodology directly into the RNC, to make it publicly accessible.

The proposed methodology is novel because:

- translation correspondences are recorded on the level of construction (LGF) instead of the sentence level
- multiple translations of the same text are supported (polyvariance)
- bilingual lexicogrammatical search functions are introduced for both MEs and PEs.

Thank you! Please visit us on our stand!