# Corpus Linguistics Tools for Loanwords and Borrowings Studies[1]

*Liliya R. Abdullina[2], Anastasia V. Ageeva[3], & Ekaterina V. Artamonova[4]*

**Abstract**

The present article deals with studies and analyzes the lexical innovations of French origin in the Russian literary language on materials of fiction, opinion journalism, as well as etymological, defining, and foreign dictionaries under well-defined criteria of foreign word extraction. Various investigations dedicated to the study of language contacts offer variety of criteria describing the functioning of foreign vocabulary in the Russian lexical system as the result of interaction of various manifestation of language contacts. The article highlights the main principles and potentials of using methods of corpus analysis for studying language contacts from a historical perspective. This study analyzes the gallicisms of the military terminology that came to the Russian language at the end of the 18th and the 19th centuries and were recorded in numerous significant examples of Russian literature. Conducting the diachronic analysis of their functional development using the methods of corpus linguistics allows us to establish and demonstrate some common and particular processes of foreign language vocabulary: the development of the form and the content of foreign vocabulary, the periodization of their maximum and minimum distribution, and the correlation of language data with extralinguistic factors.

***Keywords:*** Linguistics; Bilingual; Borrowing; French; Russian; Corpus.

## 1. Introduction

The active use of neologisms of foreign origin is a distinctive feature of the modern language situation: the vocabulary of the Russian - like of any other - language is an open system, one of the sources of which is borrowing. Being a direct consequence of the Gallophilia of the Russian society of the 18th and 19th centuries, which was one of the strongest traditions of the Russian nobility, the lexical items of French origin filled many lexical and thematic groups, actively displacing the original

---

[1] Please cite this paper as follows:

Abdullina, L.R., Ageeva, A.V., & Artamonova, E. V. (2019). Corpus linguistics tools for loanwords and borrowings studies. *Journal of Research in Applied Linguistics,10* (SP), 443-451.

[2] Kazan Federal University, Institute of International Relations; *lilioven@mail.ru*

[3] Kazan Federal University, Institute of International Relations

[4] Kazan State Power Engineering University

units and earlier borrowings from everyday life (Gabdreeva, 2011; Galdi, 1958; Kalegina et al., 2015). Today, borrowing becomes again an important source of replenishment of vocabulary - the French borrowings, along with Anglicisms, occupy a dominant position among all the neologisms, which confirms the relevance of this study.

However, despite the abundance of scientific papers devoted to the problem of borrowing (Craig, 2017; Daniels, 2017; Pennycook, 2016; Adamou et al., 2016; Ageeva et al., 2015; Cubberley, 1993), the state of modern linguistic science, that comprehends the process of borrowing, is characterized by the irregularity of the conceptual construct. In our previous works, we studied the history of the development and formation of gallicisms (Ageeva et al, 2015; Ageeva & Abdullina, 2018). The present study is a continuation of this work, the purpose of which is to demonstrate the possibilities of using corpus linguistics methods in the study of linguistic contacts and borrowings for getting both a detailed description of the functioning of a single word in Russian literature and linguistic processes typical for a particular era of its development.

## 2. Material and Methods

Among the methods of processing linguistic data, it is necessary to single out, first of all, the corpus approach, which allows analyzing foreign language vocabulary directly on real written texts and applying statistical methods to research. The concept of corpus linguistics itself has a comparatively long history: the first English dictionary, created through the analysis of authentic texts, was written in 1755. The modern model of the digital complex (Computational Analysis of Present-Day American English) was made in 1997 by a team of researchers from Brown University (USA) under the leadership of Henri Kučera and W. Nelson Francis; it contained about 500 sample texts in English. The practice of compiling national language corpora is being distributed in Europe: Survey of English Usage created by Randolph Quirk, Frantexte, which includes texts belonging to different functional styles, or the National Corpus of the Russian language, which is considered authoritative in Russian science (http://www.ruscorpora.ru). Containing more than 300 million words this corpora accorded to our study his resources.

Thus, the methods used in the present paper include linguistic description methods (examination, description, classification, comparison), as well as lexical-semantic and comparative methods and the method of diachronic description of language which are considered to be traditional for lexical-historical research. The statistical method of quantitative description was used when studying some aspects of the problem.

## 3. Results

According to the fair remark of Shmelev (1973), the lexicon of the language directly reflects the extra-linguistic reality, so that the analysis of the vocabulary allows us to identify not only intra-language factors, determined by the relationship of lexical units, but also extra-linguistic factors that themselves, to some extent or another, determine the semantics and functioning of lexemes. Combining these phenomena, each language creates its own, unique cast of reality, where these conditions are so closely intertwined that it becomes impossible to isolate the purely linguistic factors.

In the process of language contacts, we are faced with a unique phenomenon: figuratively speaking, there is a peculiar overlay of a foreign reality imprint on the system of the recipient language, which, on the one hand, leads to the formatting of the latter and, on the other, changes the outlines of the impression itself, adapting it to the structure of the target language, the mentality of its carriers, the peculiarities of their thinking, and the national picture of the world.

Let us illustrate this statement on the material of the military vocabulary of the Russian language.

## 4. Discussion

The reforms of the Russian army started by Peter I completely changed its structure in comparison to the 16th—17th centuries: they introduced not only the new types of weapons, but also the European system of ranks, tactics of warfare, a new understanding of the army's organization and role in the state, which had formed as a result of a close connection with the historical processes in Europe. All these factors contributed to the fact that the French lexical units of the military sphere still remain one of the most significant layers of Russian literary language. Moreover, in the process of historical development, most of them have lost their highly specific meaning, typical for terminological vocabulary, and replenished literary language.A clear and logical system of Russian military ranks, for example, by over 70% consists of units of French origin: *capitaine – капитан, caporal – капрал, colonel – колонель, général – генерал, lieutenant – лейтенант, major – майор, maréchal – маршал, sergent – сержант.* Analyzing some of the presented units and their etymology will give us the understanding of the historical processes in the sphere of vocabulary and the degree of their correlation with the general social evolution and particular forms of its expression.The word *сержан*т (from French sergent – from Latin servientem, accusative of serviens) – was first fixed in the French language in the middle of the 11th century (ca. 1050) in the sense of serviteur, homme de confiance employé par un seigneur 'servant, a person trusted by the master', then relatively quickly acquires the meaning of officier domanial de justice 'judge, servant of the law' and eventually evolves into homme de guerre de condition inférieure 'lower rank

military man' (ca. 1130). In the 16th century, the word acquires modern meaning: officier militaire (qui fait ranger les soldats) ‘a military officer (who is in charge of tactical formations of the soldiers)’ and is included in the Russian language, being fixed in the literature of the 17th century. This word frequently appears in literature of the first half of the 19th century, then the frequency of occurrences drops to near-zero marks, and only in the middle of the 20th century there is a sharp surge of its use.

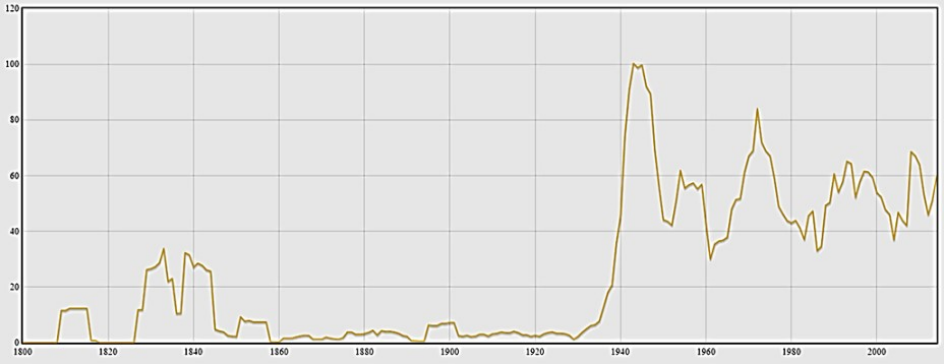



*Figure 1.* Graph of Frequency Distribution of Occurrences of the Word *Сержант*.

As can be seen from the graph above, the frequency of the fixation a lexical unit in works of fiction is directly determined (albeit with a certain delay) by social processes. Thus, in 1798, Emperor Paul I abolished the titles of сержант ‘sergeant’ and старший сержант ‘senior sergeant’, replacing them, respectively, with унтер-офицер ‘non-commissioned officer’ and фельдфебель ‘feldwebel’. The word gets almost completely out of use in literature, since during the 19th – early 20th centuries the reality associated with it appears under a different name. The surge from 1825 to 1845 is due, presumably, to an abundance of literature devoted not only to the Patriotic War of 1812, but also to the events preceding it. Its main veterans began service during the reign of Catherine II, even if—as ironically stated by A.S. Pushkin in The Captain's Daughter—during their infancy, thus avoiding the “Table of Ranks”: “*Матушка была еще мною брюхата, как уже я был записан в Семеновский полк сержантом, по милости майора гвардии князя Б., близкого нашего родственника. Если бы паче всякого чаяния матушка родила дочь, то батюшка объявил бы куда следовало о смерти неявившегося сержанта, и дело тем бы кончилось*”. In 1940, the rank of sergeant will return to the Russian–now Soviet–army, and the Great Patriotic War that followed in 1941 will provide its wide circulation in military literature.

After being borroed into the Russian language, the word *лейтенант* narrowed its set of meanings: this rank existed only in the navy and was equated to an army captain, which accounts for the relatively low, albeit sufficiently stable, frequency of occurrences of the unit in the literature corpus:

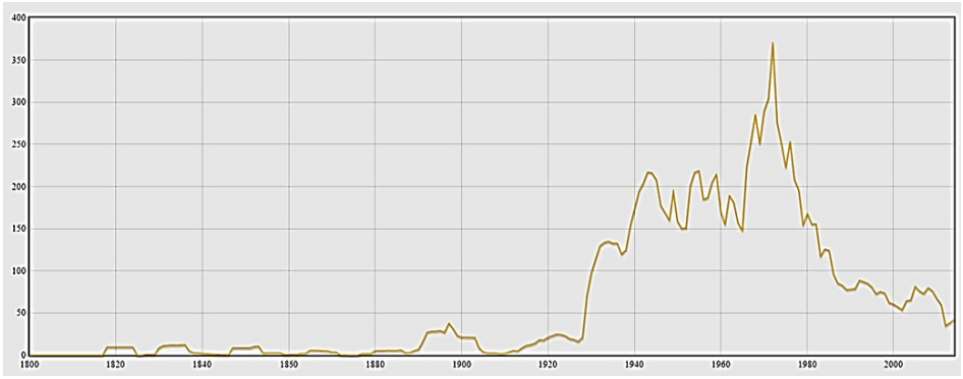



*Figure 2.* Graph of Frequency Distribution of Occurrences of the Word ***Лейтенант.***

As can be seen from the graphs, the dynamics of the occurrence of the units *сержан*т and *лейтенант* roughly coincide throughout the 19th century, albeit for various reasons: if the low occurrence of *сержант* was due to the administrative factors, the word *лейтенант* had a very limited sphere of use. The sharp increase in frequency happening in the 20th century is due to the same processes: the administrative decision of returning the rank of "сержант" and expansion of the meaning of word "*лейтенант*", which from 1935 was assigned to all graduates of military schools.

We see a different picture on the example of the word *капитан*, which entered the Russian language in the 17th century, together with the members of the foreign military hired to serve in Russia. The first fixation in fiction happened in the early version of the comedy "The Minor" by D.I. Fonvizin (1764). Since the end of the 18th century the use of the unit in literary works is consistent; there is a steady upward trend in the frequency due to the absence of significant external factors (the rank of captain existed in the army of the Russian Empire, it was preserved in Soviet times and exists today):




*Figure 3.* Graph of Frequency Distribution of Occurrences of the Word ***капитан***.

Analyzing the material of military vocabulary, we notice that only a relatively small number of French words became archaisms: *инфантерия, ретирада, спекулятор, бивуак, ведет.*

The word *инфантерия* has been in use in French since the 14th century. It is recorded in various forms and variants: enffanterie (1502), fanterie (1547), infanterie (1553) in the meaning of troupes à pied 'foot troops'. It was borrowed into the Russian language in the 18th century, simultaneously with the designations of other types of troops (cavalry, artillery), but did not receive a wide distribution; it consistently appears in private correspondence and official documents along with the original word *infantry*, but in the literature it is represented by single entries:

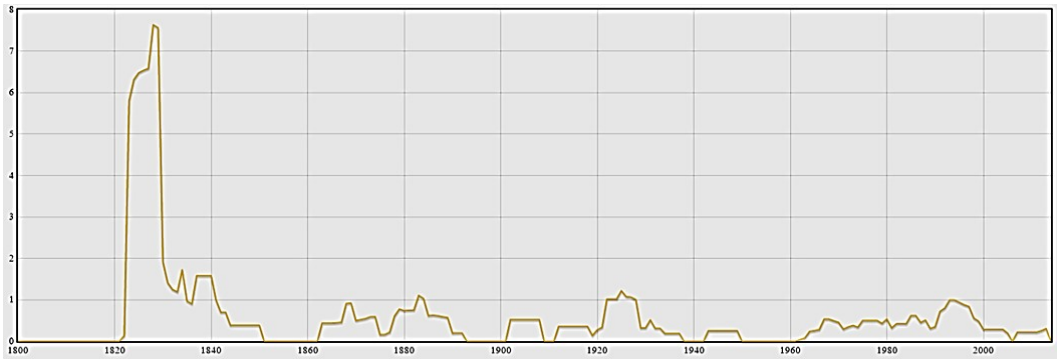



*Figure 4.* Graph of Frequency Distribution of Occurrences of the Word ***инфантерия***

For comparison, we will analyze the graph of the frequency of occurrences of the original word.




*Figure 5.* Graph of Frequency Distribution of Occurrences of the Word **пехота**

As can be seen from the comparison of both curves, the extrema of the first curve fall on the late 1920s - early 1930s, and is followed by a sharp decline. To this day, the frequency of this word rarely exceeds one or two occurrences per million word forms. The first third of the 19th century was the time of the flourishing of literature about the War of 1812 written by its eyewitnesses and veterans; most of

them belonged to the generation born at the turn of the 18th-19th centuries, when the term инфантерия was in use among the military. As N.A. Durova writes in the"Angle"*: «И кавалерия и инфантерия раскланивались очень учтиво с сидящими в карете, но легкая, чуть приметная усмешка, которою сопровождался поклон их, была как-то неуместна, судя по блистательной наружности экипажа тех особ, к которым относилась»*. "The Cavalry Maiden" easily embeds the gallicism in her writing (unlike the later works, here it is quite appropriate in its neutral, purely nominative function). Further, the unit will be included in the literary context either exclusively in the military clerical phrase «генерал от инфантерии» (ср. *Собственно говоря, я даже не знаю, кто меня будет читать: может быть, прапорщик, а может быть, генерал от инфантерии…*, or for stylistic purposes (as we can find in Turgenev's works: *«…один в инфантерии, другой в кавалерии, третий сам по себе…»* or in Chekhov's: *«Не пехота какая-нибудь, не инфантерия, а флотский!»* (Galdi, 1958).

The frequency curve of the word *пехота*, by contrast, begins to grow since the 1830s, and this growth is impressive: the maximum reaches the point of about thirty occurrences, and the minimum never falls below one. The original word, therefore, not only was hard to replace, but even left the borrowing on the periphery. This phenomenon fits logically into the theory of linguistic contacts: gallicism did not meet any of the requirements that would allow it to gain a foothold in the language, since at the time of its borrowing the recipient language had an original equivalent — monosemic, short and already common in the language.

Being unable to compete with the original word, the gallicisms *бивуак, ведет,* and *ретирада* are not actually fixed in the Russian military vocabulary.

Quite a significant number of military terms underwent a process of historization, i.e. became out of use due to the loss of related realities. This vocabulary includes military tactics, structures and orders of battle (*каре, пикет, редут, etc.*), and also weapons (*мушкет, пика, пистоль, etc*.).

## 5. Summary

Thus, we come to the conclusion that the vocabulary of the military sphere accurately illustrates both the features of the process of linguistic contacts and the dynamics of the intralingual development of previously borrowed vocabulary, which is determined, besides the linguistic patterns, by socio-cultural vectors of evolution. We are interested in both the specificity of the contacts which have caused the borrowing of foreign elements into Russian and their further consolidation in the language. The military vocabulary of French origin demonstrates how the idea and understanding of the Russian army in the Petrine and post-Petrine eras were formatted by the foreign realities; we can see how dynamic this formatting was from the speed

of mental assimilation of a gallicism, its transition from a closed terminological group into a literary language and the spread in it.

## 6. Conclusions

Corpus text analysis methods provide exceptional possibilities for studies of foreign language vocabulary in a diachronic perspective. They make it possible not only to trace active processes in the borrowing and assimilation (first fixation, formation of a form, activation of new meanings, entry in language use, archaization, reactivation), but also to study the extralinguistic components of language development–economic, socio-historical and cultural specifics of the social evolution.

**Acknowledgements**

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

## References

Adamou, E., Breu, W., & Scholze, L. (2016). Borrowing and contact intensity: A corpus-driven approach from four Slavic minority languages. *Journal of language contact*, *9*(3), 513-542.

Ageeva, A. V, & Abdullina, L. R. (2018). Semantic analysis of the French word “pistol” and its origin in Russian literary language. *Journal of Social Sciences Research, 1*, 532-535.

Ageeva, A. V., Abdullina, L. R., & Latypov, N. R. (2015). Semasiological relations between the lexical parallels in the French and Russian languages (a case study of the French borrowed vocabulary). *Journal of Sustainable Development*, *8*(4), 53-60.

Ageyeva, A. V., Vasilyeva, V. N., & Galeyeva, G. I. (2015). Language situation in the Russian society at the start of the 19th century: Bilingualism or diglossia? *Journal of Language and Literature, 6*(1), 322-325.

Craig, C. (2017). Borrowing, incomplete lexical diffusion and the High German tenues shift. *Nowele-North-Western European Language Evolutions*, *70*(1), 233-244.

Cubberley, P. (1993). The phonologica dynamics of foreign borrowings in Russian. *Austral. Slavonica an East Europe Studies, 7*(1), 49-74.

Daniels, D. R. (2017). A method for mitigating the problem of borrowing in syntactic reconstruction. *Studies in Language*, *41*(3), 577-614.

Gabdreeva, N. V. (2011). *Istoriya Frantsuzskoj Leksiki V Russkikh Raznovremennykh Perevodakh (The history of French vocabulary in Russian translations of different epoch).* Moscow: Lenand Publications.

Galdi, L. (1958). *Slova Romanskogo Proiskhozhdeniya V Russkom Yazyke (Words of Roman origin in Russian)*. Moscow: Publishing House of Moscow University.

Kalegina, T. E., Takhtarova, S. S., & Zaglyadkina, T. Y. (2015). Denglish and Franglais in the framework of the modern European linguistic landscape. *Journal of Language and Literature*, *6*(3), 195-198.

Pennycook, A. (2016). Reflecting on Borrowed Words. *Tesol Quarterly*, *50*(2), 480-482.

Shmelev, D. N. (1973). *Problemy Semanticheskogo Analiza Leksiki (Problems of the lexico-semantic analysis).* Moscow: Nauka.