

DOI: 10.7596/taksad.v6i5.1250

Citation: Galeev, T., & Habibulina, E. (2017). Frequency of Verbal Forms and Language Standard. Journal of History Culture and Art Research, 6(5), 90-98. doi:<http://dx.doi.org/10.7596/taksad.v6i5.1250>

Frequency of Verbal Forms and Language Standard

Timur I. Galeev¹, Elena V. Habibulina²

Abstract

The article offers the description of a modern experiment, which gives the possibility of complex information extraction about the cognitive structure of the linguistic evolution of Language Standard (Norm). The study was conducted using the Google Books Corpus, which provides unprecedented opportunities for linguistic studies. The purpose of the experiment was to identify the patterns of competing forms evolution within the center of the verbal paradigm (3Sg and 3Pl) on the basis of the data concerning the frequency of their use. The study was conducted on the material of excess verb forms with the variability of *a/o* vowels in a root (*обусловливать/обуславливать*). The graphs for variable word form competition clearly illustrate that the process of norm change consists of stages, each of which has numerical characteristics of two competing word form use. The chronological frameworks for an inflectional model change are established with the accuracy of up to 10 years. The graphs obtained as the result of the experiment make it possible to conclude that almost half of the verbs were not variative, although they previously considered. During the discussion of the obtained empirical data, a conclusion is made about the morphemic structure of a word, in which a root vowel changes. Possessing the information about similar processes in other verb paradigms, researchers are able to predict a possible change of inflectional models in the future and, as a consequence, the fixing of a new norm in lexicographical, orthographic and orthoepic sources.

Keywords: Verb, Paradigm, Ngram, Google books, Statistics, Norm.

¹ Institute of Philology and Intercultural Communication, Kazan Federal University, Kazan, Russia. E-mail: tigaleev@kpfu.ru

² Institute of Philology and Intercultural Communication, Kazan Federal University; Kazan, Russia. E-mail: elvikh@mail.ru

Introduction

The significance of the frequency for the morphological system of verbs has been studied very actively in recent years. Thus, using the statistical methods they proved that more frequent English verbs are less inclined to regularization - the transition from the category of irregular to regular ones (*arise: arose → arised*), than less frequent ones (Lieberman et al., 2007). A similar study with a more traditional linguistic approach was also performed on the basis of verb competition analysis from strong (~ regular) conjugation and weak (irregular) verbs of the German language (Carroll et al., 2012; Piperski, 2014). By the actual data these studies confirmed an intuitively obvious assumption that more frequent words retain the inflectional type well, while less frequent words tend to change under the influence of analogy.

Taking into account the fact that Russian language is inflectional, a special attention must be paid to the various morphological processes accompanying the inflectional paradigmatics. Such processes are actively studied by modern experts in Russian language (Zagidulina et al., (2016); Varlamova et al., (2014)).

The object of the study is represented by variable verbal forms with the variability of the vowels *a/o* and the root (*обусловливать/обуславливать*) according to the data of several modern dictionaries and collections (Graudina, 2004; Gorbachevich, 2009), as well as according to own files of examples compiled by the authors, with the total volume of about 68 units.

The subject of the study is the evolutionary dynamics of "competing" pairs of verbal forms based on the data of their usage frequency. The aim of the study is to reveal the regularities of variable form evolution within the center of the verbal paradigm.

Materials and methods

In order to study the competition of verbal variants it has been used a quantitative method. Using Ngram Viewer service, which allows to create graphs concerning the frequency of word usage using the material of the texts from the Russian-speaking part of the Google Books corpus, the main interaction models of particular two verbs are described.

Google Books is an information system that contains several corpora of marked book texts in 9 languages. The books were obtained from 40 university libraries of various countries. In addition, some publishers provided the copies of their products for the developers of the system, both in print and in electronic form. Russian segment of Google books contains nearly 600,000 books (texts) or 67 billion words. It is important that Google Books N-gram Viewer is a young and evolving system. It's hard to say how often changes take place in it: the volume of cases changes, new functions are added, etc.

The system does not support the grammatical normalization of the vocabulary, in other words, the search for lexical units (words and phrases) is performed only in a specified form (word-form search) and develops the diagrams of word forms and word combinations occurrence frequency. The relative frequency of the specified N-gram in a given year expressed in percent is developed on the vertical axis of the graph developed by the system. The horizontal axis of the graph shows the years included in the time interval specified by a user. Each curve of the graph is colored; it is indicated which N-gram (a word or a phrase) it corresponds to at the end of the curve. Below is one of 1020 obtained competition graphs concerning the excess forms of the verb group under study (Graph 1).

Graph No. 1. Frequency change of the verbs *оспоривать* and *оспаривать*



The graph with the frequency of the infinitive forms *оспоривать* and *оспаривать* demonstrates that initially in the first half of the 19th century the form *оспоривать* was more common: fluctuations in the range of 0.0005% with an average annual volume of texts of 300,000 word forms during this period demonstrate that more than 150 infinitives with a root vowel "o" were used annually. Infinitives with the root "a" were met 10 times less – less than 15 word forms per year. Thus, the total frequency of both types of infinitives was 170-180 word forms.

During the period of 1840-1860 there is the decrease in the frequency of the form with "o" and the increase in the frequency of its competitor – the form with "a". At that, the total number of word usage for both types of forms remains at the level of 170-180 per year.

The point of bifurcation occurs in the 1850's. In these years, the outgoing and established norm is characterized by an equal degree of penetration, some uncertainty appears.

The close location of the frequency curves at the end of the 19th century reflects the dominance of the form with "a". While maintaining the overall frequency of both options, their proportions are also preserved: the infinitive *оспаривать* is used more often than the competitor, just like the competitor himself, the infinitive *оспоривать* prevailed over the modern norm only a few decades ago.

As for the further functioning of the verbs, it should be noted that in the 20th century these verbs lose their "popularity" in the book speech: their share in the total number of words during the period under study is reduced by 2-3 times while maintaining a tenfold proportion of forms, i.e. the form with "a" is still used much more often than the form with "o".

Undoubtedly, the research is focused not on a detailed analysis of each case of word interactions, but on the identification of patterns by incomplete inductive methods which became traditional ones for linguistics: classification and structuring.

Results of the study

In the framework of a quantitative study of an excessive verbal paradigm structure and the "conservatism" of its individual parts, 68 pairs of verbal paradigms were analyzed with the variability of the vowels "o/a" in the root.

During the study of each case of variability, the paradigm with 15 forms was developed: the infinitive of the 1st, the 2nd, and the 3rd person singular and plural, active and passive participles of present and past

tense in the nominative case of the masculine gender (singular form), the verbal participle of the imperfective and perfective form, the imperative mood of the singular and plural.

The obtained results are presented in the form of a graph the vertical axis of which has the frequency coefficient, and the horizontal axis is the timeline.

The next step in the work with empirical data is the distribution of graphs based on the results of variational word form competition. When they analyzed verbs that were previously considered variative ones (Graudina 2004), only one of the forms was found in 46% of cases. One of such cases is illustrated by Graph 2.

Graph No. 2. Frequency change for the verb *одолживать* and *одалживать*



This situation was recorded in the forms of 31 paradigms. Most of these paradigms (21 cases) are represented by forms with the vowel "a" in the root. This means that the process of vowel change ended before the nineteenth century. In 10 paradigms the only detected form was the form with the vowel "o" in the root. These verbs do not participate in the process of the root vowel "o" replacement with "a" or resist it successfully, remaining "conservative" ones throughout the entire paradigm.

In most cases (54% of the total number of analyzed paradigms) the variability of various dynamics was recorded. The change in the norm occurred among 25 cases. One such case is depicted on Graph 3.

Graph No. 3. Frequency change for the verb *успокоивать* and *успокаивать*



In all these cases this is the transition to the vowel “a” in the root.

вырабатывать → *выработывать*

вырабатываться → *выработываться*

заготовливать → *заготавливать*

зарабатывать → *зарабатывать*

обрабатываться → *обработываться*

обработывать → *обрабатывать*

отрабатывать → *отрабатывать*

разрабатывать → *разработывать*

разрабатываться → *разработываться*

дотрогиваться → *дотрагиваться*

затрогивать → *затрагивать*

притрогиваться → *притрагиваться*

заподозривать → *заподазривать*

обеспокоивать → *обеспокаивать*

освоивать → *осваивать*

присвоивать → *присваивать*

усвоивать → *усваивать*

оспоривать → *оспаривать*

приспособливать → *приспосабливать*

разрознивать → *разразнивать*

удвоивать → *удваивать*

уполномочивать → *уполномачивать*

условливаясь → *уславливаясь*

успокоивать → *успокаивать*

устроивать → *устраивать*

In general, these secondary imperfects with the suffix –и/ыва– in combination with the prefixes вы–, до–, за–, об–, от–, раз–, о–, при–, у–.

In 9 cases the competition of forms was much longer than 10-20 years. The forms with the vowels "a" and "o" could coexist in a root for many decades. Thus, the lack of uniformity in the book style throughout the 20th century was reflected in the graphs of verb *удостоивать* / *удостаивать* paradigm frequency.

Graph 4 shows a similar case.

Graph No. 4. Frequency change for the verb *удостаивавший* and *удостоивавший*.



Although it is difficult to define the only form that is approved as a result of a long-term competition, it should nevertheless be noted that during the last years of the 20th century the form with the root vowel "a" begins to predominate most often (as in the graph above). Such indicators were found in 8 paradigms of 9 similar ones.

Such graphs reflect the prolonged 3rd stage of the norm change ("bifurcation") and the started stage 4 ("inertia").

The result of more than 1000 graphs analysis concerning the dynamics of variative verbal forms revealed that almost half of the cases has no competition of forms and the only form is a bit more often - the form with the root "a".

There were no cases of two variant lexemes functioning in the same volume ("bifurcation points") during the last decade of the 20th century. This was contrary to the popular opinion about the linguistic chaos of post-Soviet Russia.

Conclusions

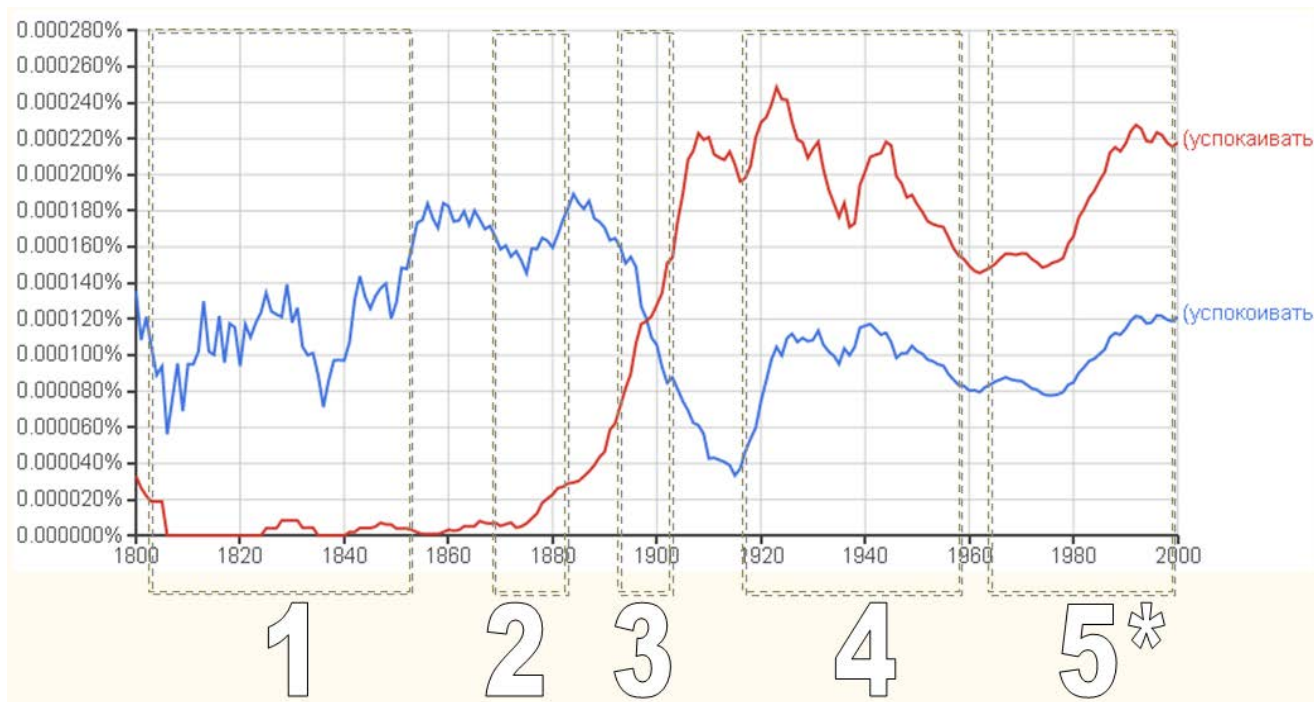
An invariable feature of Russian verbal variation for almost three centuries is its polemical nature, which needs to be settled by the application of the methods that bring positions closer and neutralize permanent disagreement.

The Google Books Ngram Viewer system provides unprecedented opportunities for linguistic research. Huge arrays of marked texts and the speed of result obtaining make this system an extremely valuable research tool.

Possessing the information about similar processes in other verb paradigms, researchers can predict a possible change of inflectional models in the future and, as a consequence, the fixing of a new norm in lexicographical, spelling and orthoepic sources.

1. The process of a norm change consists of stages, each of which is characterized by numerical characteristics of two competing word form use. Graph 5 shows all identified stages clearly.

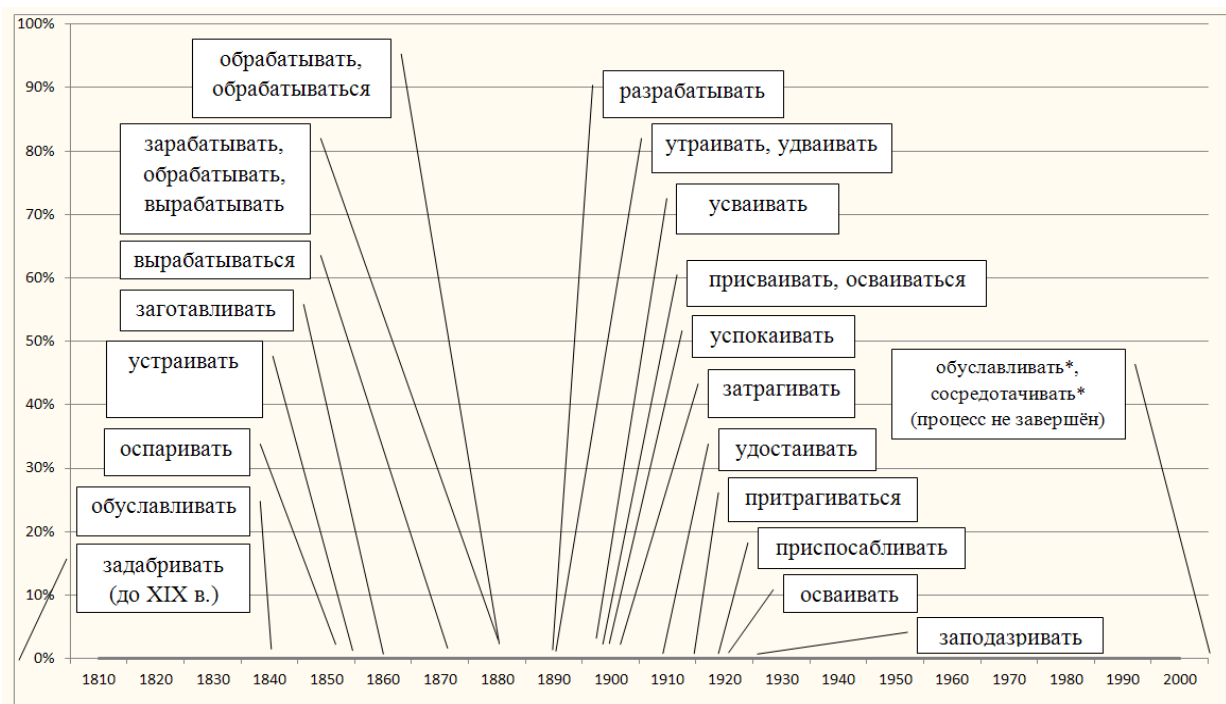
Graph 5. The main stages of language standard change.



- 1) A stable period of an old form dominance (the form with the root "o");
- 2) Consecutive increase in the number of new form use (usually with the root "a") at a corresponding frequency decrease of the form with the root "o";
- 3) Then comes the moment of uncertainty, at which the frequency of the old and the new norm is compared – the so-called bifurcation point;
- 4) The inertial phase, at which the old form continues to be used for some time, is used rarer in times than the newly established one;
- 5*) In the future, the old form can cease to be used or to continue by occupying its "niche" in the language (it acquires semantic meaning, stylistic coloring, it is transferred into the sphere of professionalisms) or it can be used by mistake.

2. Accurate chronological indicators of the norm change in each pair of verbs are obtained. These results would be interesting to correlate with conversational speech in order to find out how much time is necessary to reflect in books the changes that occurred in speech. The illustration below shows the chronological framework of norm change in each pair of verbs. Chronological description of the norm change was made up with the accuracy up to 10 years (Graph 6).

Graph 6. Chronology of language standard change.



Speaking of other study results, it can be said that, on the basis of a super-large body of texts of the book style, the classification of the verbal pairs was constructed depending on the dynamics of their use frequency. The methods describing the dynamics of variable forms of verbs can be applied later on other cases. Thus, they will allow not only to describe and explain linguistic phenomena, but also to make informed quantitative predictions of linguistic forms development.

These and other results and illustrations are kept electronically in the cloud storage and are accessible via the following link because of their large volume: <https://cloud.mail.ru/public/9Ecf/oUb7TZChd>

Acknowledgements

The work is performed according to the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, agreement № 34.5517.2017/6.7, by Russian Government Program of Competitive Growth of Kazan Federal University and by RFBR (project №16–06–00165 A).

References

Carroll, R.; Svare, R. & Salmons, J. (2012). Quantifying the evolutionary dynamics of German verbs. *Journal of Historical Linguistics*, 2(2), 153-172.

Galeev, T. I. & Solovyev, V. D. (2016). Methods of application of modern text corpora in the study of the morphological system of Russian verbs unification of I productive (irregular) class of verbs. *Quantitative model based on google books. Modern Journal of Language Teaching Methods*, 177-180.

Gorbachevich, K. S. (2009). Variability of the word and language norm: On the basis of modern Russian language. Moscow: LIBROKOM.

Graudina, L. K. (2004). Grammatical correctness of Russian speech. Stylistic dictionary of variants. Moscow: Astrel.

Lieberman, E.; Michel, J. B.; Jackson, J.; Tang, T. & Nowak, M.A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449, 723-716.

Piperski, A. Ch. (2014). Alternation in the root as a pledge of sustainability: from the history of the strong verbs in German language. *Acta Linguistica. Proceedings of Institute for linguistic Studies*, 10(1), 821-831, Petersburg: Science.

Solovyev, V. D. & Kibrik, A. A. (2015). How can computer technologies help linguistic typology? *Herald of the Russian Academy of Sciences*, 85(1), 33-39.

Varlamova, M.; Palekha, E. & Miftakhova, A. (2014). Visual listening in theory and practice of effective foreign language teaching. *INTED2014 Proceedings. 8th International Technology, Education and Development Conference*, March 10th-12th, 2014, Valencia, Spain.

Zagidulina, A. F.; Gilazetdinova, G. Kh. & Islamova, E. A. (2016). Associative fields of lexemes "Homeland" and "Vatan" in Russian and Tatar linguistic consciousness. *Journal of Language and Literature*, 7(2), 289-292.