Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2025”

April 23–25, 2025

Derivation of Aktionsarten from (non-)directional verbs of motion: posing restrictions on restrictions

Chuikova O. Iu.
Herzen State Pedagogical University of Russia
oxana.chuykova@gmail.com

Gorbova E. V.
independent researcher
elenagorbova12@gmail.com

Abstract

The paper examines prefixal derivatives of Russian motion verbs that represent perfective Aktionsarten. The focus is on the statistical analysis of motion verbs as a specific subset of Aktionsarten characterized by their derivational features. Both sets (verbs of motion and Aktionsarten) are included in the database of Russian prefixed verbs developed by the authors based on the Dictionary of Russian language. One of the aims of the study is to verify the statement defended in previous works that the stem of a motion verb imposes restrictions on the possible Aktionsarten derived from it. This verification will be based on both dictionary data and the data from the Russian National Corpus and the Runet.

Keywords: verbs of motion; Aktionsarten; derived perfectives; prefixal and circumfixal derivation
DOI: 10.28995/2075-7182-2025-23-86-102

Деривация способов действия от (не)однонаправленных глаголов движения: ограничения и условия их смягчения

Чуйкова О. Ю.
РГПУ им. А. И. Герцена
oxana.chuykova@gmail.com

Горбова Е. В.
независимый исследователь
elenagorbova12@gmail.com

Аннотация

В статье рассматриваются приставочные дериваты русских глаголов движения – перфективные представители (морфемно-характеризованных) способов действия. В фокусе внимания находится статистический анализ глаголов движения как специфического по своим деривационным характеристикам подмножества способов действия. Оба упомянутых множества являются составной частью базы данных русских приставочных глаголов, созданной авторами по Словарю русского языка (Малому академическому словарю). Одной из задач исследования явилась также верификация известных по литературе вопроса ограничений на вхождение дериватов от глаголов движения в способы действия с опорой как на словарные данные, так на данные из Национального корпуса русского языка и Рунета.

Ключевые слова: глаголы движения; способы действия; производные перфективы; префиксальная и циркумфиксальная деривация

1 Preliminary notes

Russian verbs of motion (hereinafter, MVs) comprise a compact group of 15 or 14 pairs (see [7] for a review of interpretations) when treated narrowly as non-derivative verbs (simplexes). We accept the list of 14 pairs from [12]: *бежать – бегать, брести – бродить, везти – возить, вести – водить, гнать(ся) – гонять(ся), ехать – ездить, идти – ходить, катить(ся) – катать(ся), лезть – лазить, лететь – летать, нести(сь) – носить(ся), плыть – плавать, ползти – ползать, тащить(ся) – таскать(ся)*. The specificity of this group is that both members of a pair function as imperfectives, and their semantic distinction is interpreted through the binary opposition of (uni)directionality / multi/non-directionality and/or singularity / plurality, e.g., *идти*^{Dir} *– ходить*^{NonDir}, *ехать*^{Dir} *–*

ездить^{NonDir}, *бежать*^{Dir} – *бегать*^{NonDir}[1]. The subject of this study is prefixal and circumfixal derivatives from 14 pairs of MVs provided they are perfective and represent one of the Aktionsarten (hereinafter, As), e.g. *избегаться*^{PFV} (excessive-multiple A) from the simplex *бегать*^{NonDir} or *забродить*^{PFV} (inchoative A) from *бродить*^{NonDir}.

The peculiarities of derivation of a number of As from MVs were mentioned in the literature.

In [9: 319-322], a series of observations is presented regarding the relationship between *Dir/NonDir* derivatives and As (*совершаемости* in Isachenko's terms). From *Dir*, an ingressive *совершаемость* is formed with the prefix *по-*, as in *Он встал и пошёл к выходу* 'He stood up and went to the exit'. In certain forms, the corresponding derivatives express a resultative meaning, as in *Отца нет дома, он пошёл в город* 'The father is not at home, he went to the city' [9: 320]. For *NonDir* derivatives, [9: 323] notes the realization of the following As: inchoative (*заходить*); delimitative (*походить*); perdurative (*проходить*); evolutive (or ingressive-intensive; *расходиться*); and saturative (*находиться вдоволь*).

The interaction of *Dir*/*NonDir* with perfectivizing prefixes is discussed in [15: 100–103; 16: 6–11]. It is noted that *Dir* stems are compatible with all prefixes, while *NonDir* stems are not. When a prefix is compatible with both, their meanings differ: *Dir* typically conveys a spatial meaning (except for *по-*, which serves as an ingressive marker), whereas *NonDir* conveys a non-spatial meaning. Many prefixed derivatives of *NonDir* represent a specific A [15: 101]. An intersection between the semantics of *NonDir* prefixes and external prefixes, which are considered markers of As in the literature [14], is also noted [16: 11]. The authors observe that with *Dir* stems, the prefix *по-* has an ingressive meaning, while *NonDir* stems can take *за-* (inchoative), *по-* (delimitative), *от-* (finitive), *про-* (perdurative), and *на-* (cumulative). In fact, limitations regarding the compatibility of *Dir* and *NonDir* with certain prefixes[2] and their semantics have been formulated.

However, as far as we know, no comprehensive study of the peculiarities of the formation of As from 28 MVs along with the analysis of the system of perfective As (in any known edition) has been carried out so far.

This study aims to systematically examine the formation of perfective As from MVs using a continuous sample from the Small Academic Dictionary. The sample will be supplemented at certain stages with the data on derivatives not included in the dictionary but available in other sources, such as corpora and the Runet. Additionally, the study seeks to verify existing notions regarding limitations on forming perfective As based on the stem type of the root MVs. The ultimate goal is to develop a more accurate model of the intersection between MVs and aspectual grouping of As in Russian verb grammar.

Following the preliminary notes in Section 1, Section 2 describes the research material and outlines the methodology used in this study. Section 3 discusses the productivity of MVs as a basis for forming As, addressing limitations related to two types of stems and the impact of expanding linguistic sources on the findings. Finally, Section 4 summarizes the results of the study.

2 Materials and methods

The material of the study is a continuous sample of prefixal perfectives of 19952 lexemes (lexeme as "a word considered in one of its available meanings" [1: 55]), compiled according to the Small Academic dictionary, also known as MAS [4]. The sample includes 1378 MV-derivatives[3], 304 of which represent As. The derivation from the MVs was determined on formal grounds, so the sample also includes derivatives from non-literal meanings of MVs (e.g., *отбродить* (about wine), *провести* <an event>). The issue of semantic annotation of MVs and their derivatives as non-literal, or metaphoric, is covered in Appendix D.

We utilize the As system, which is based on specific principles (see [5; 3]). As denote "different types of semantic modifications of the verb expressed by certain formal means" [15: 110]. An A is defined as a linguistic sign comprising a signifier (a word-formation marker, i.e. an affix or a combination of

[1] Here we use the notations *Dir* and *NonDir* for (uni)directional (like *идти*) and non-directional (like *ходить*) MVs, borrowed from [16]; one of the important questions for [16] about the direction of derivation in a number of aspectual pairs of MVs derivatives: whether *приходить*^{PFV} is derived from *прийти*^{PFV}, as argued in [16] following [9], or from *ходить*^{NonDir}, as in [8], [10] or [6], is not addressed here, as it is irrelevant for the purely perfective material of the study.

[2] As for circumfixal As, the derivation from *Dir* only is postulated for durative-negative *до-...-ся* in [11].

[3] The principle we established in the annotation is that a derivative at any stage of its derivation is considered an MV-derivative if its original stem is one of the 28 simplex verbs (14 pairs of *Dir* and *NonDir* from [12]).

affixes) and a signified, which represents an aspectually relevant semantic modification. Ambiguity in either component is analyzed through the identity of the linguistic sign. Multiple meanings associated with a single marker are classified as polysemy if there is non-trivial semantic similarity, or as homonymy if such similarity is absent. When different signifiers share similar semantics, synonymy is posited. Consequently, combining derivatives with different affixes within a single A is prohibited. Using this framework and classifications from [15: 110-135; 13: 141-167], we compiled a working list of 22 As (see Appendix A) based on prefix perfectives in MAS. While this list may not be theoretically ideal, it suffices for our research objectives. For perfective MV-derivatives, the list is reduced to 20, excluding the least productive *в*(*о*)*з*-delimitative (e.g., *взгрустнуть*) and *при*-attenuative (*приоткрыть*).

Further, a few composites (*снизойти*), denominal (*занавозить*) and polyprefixal derivatives like *по-вы-лезти, по-вы-таск-ивать* derived at the second or third derivation step from the prefixal perfective or the secondary imperfective (27 lexemes of *по*-distributive and 4 lexemes of *по*-attenuative) were withdrawn from the material. Since all the above derivatives are not directly formed from initial *Dir* or *NonDir*, their analysis in terms of the derivational stem is more complex than that of other MV-derivatives, so it is reasonable to consider them separately. For comparability of MV-derivatives with other prefixal perfectives, composites, nominal and polyprefixal derivatives were also removed from the original sample. As a result, the sample of prefixal perfectives was reduced to 18226 lexemes, of which MV-derivatives amounted to 1304 units, with 270 of them representing 19 As.

3 Productivity of As derived from MVs

3.1 A general overview of productivity of As derived from MVs

We will start by examining the overall picture of As productivity from MVs, then move on to the peculiarities of individual As.

First, we note that the proportion of verbs representing As in the MVs subset does not differ significantly from that for the whole set of prefixal verbs: about 20% (3771 out of 18226 in the whole sample and 270 out of 1304). However, individual As derived from MVs show higher or lower productivity than those derived from the rest of the verbs in the sample.

Let us consider the As for which this difference is statistically significant (see Appendix B). The increased level of productivity in the set of MVs derivatives is observed for the following As: *до*-completive[4] (*долезть*, *добродить*), *о(бо)*-distributive-summary (*облазить*, *обежать*), *по*-ingressive (*побежать*, *полететь*), *раз(о)-/рас*-distributive-summary (*растащить*, *разбежаться*) and *у-...-ся*-excessive-intensive (*убегаться*, *уходиться*). The following As are characterized by a lower level of productivity among the MV-derivatives: *за*-inchoative (*забегать*, *занестись* (about a chicken), *пере*-distributive (*перетаскать*, *перевестись* ‘disappear’), *пере*-repetitive (*переходить* <when playing chess>), *по*-delimitative (*полазить*, *поводить*), *по*-distributive (*потаскать* <everything, much>) and *под*-attenuative (*подвезти* ‘to have a bit of luck’).

Comment

1. Most of the above-mentioned As demonstrate the ability to derive from both *Dir* and *NonDir*. Thus, we can question the statement in the literature regarding the “fixation” of a number of As exclusively to one or another type of MV stem.

2. Ingressive *по-* shows increased productivity for MVs, while inchoative *за-* demonstrates decreased productivity; this indicates a more complex relationship between the two markers of the beginning of a situation than their complementary distribution over *Dir* and *NonDir* stems suggests. Overall, ingressive *по-* is characterized by limited productivity, with 46 lexemes according to MAS. Outside of MVs, this A is formed from verbs with the semantics of movement (see [15: 115-116]), often synonyms of *Dir* (*потопать*, *поскакать*, *пошлепать*, *повлечь*), or, in isolated cases, from verbs of state (*полюбить*, *поверить*). In contrast, inchoative *за-* forms one of the most numerous A (after the delimitative *по-*, see [2]) As, with 480 lexemes according to MAS.

It is also interesting that synonymous As with different prefixes show the same tendency to productivity within MV-derivatives: the distributive-summary *о(бо)-* and *раз(о)-/рас-* are characterized by increased productivity, distributive *по-* and *пере-*, attenuative *по-*, *под*(*о*)- and *pri-* – by decreased

[4] Here, the completive A is the result of combining the final-completive, completive-partitive, and terminative-locative As identified in [13].

productivity (up to the lack of attenuatives with *no-* and *при-* among MVs). The observed picture may be a consequence of both formal (at least it concerns *no*-distributive and attenuative As, mostly forming a subset of polyprefixal MVs, which are left out of consideration) and semantic reasons. The latter require further investigation, which is beyond the scope of this study.

3.2 Productivity of As from *Dir* и *NonDir* in MAS

We will now turn our attention to the peculiarities of the derivation of As from *Dir* and *NonDir*.

Overall, derivation from *Dir* is more productive, yielding 955 derivatives compared to 349 from *NonDir*. However, within the subset of As, *NonDir* produce more derivatives: 147 from *NonDir* versus 123 from *Dir*. The difference in the distribution of derivatives from the two stem types within the subset of As and beyond is statistically significant (see Table 1). This finding aligns with existing literature, which suggests that deriving As from *NonDir* is more common, while forming As from *Dir* is more challenging.

	Dir	***NonDir***	**Total**	**% *NonDir***
non As[5]	833	202	1034	19,54%
As	123	147	270	54,44%
Total	955	349	1304	26,76%
The difference is statistically significant, $\chi 2 = 131{,}332$, $p << 0{,}01$				

Table 1: Distribution of derivatives from *Dir* and *NonDir* in and beyond the As subset

We will now examine the set As derived from *Dir* and *NonDir* in more detail. Table 2 presents the number of lexemes for each A formed from *Dir* and *NonDir*, first sorted in descending order by the "*Dir*" column and then in ascending order by the "*NonDir*" column. Fig. 1 presents this data in a more visually accessible format.

	As	***Dir***	***NonDir***	**Total**
1	*до-*: completive	30	7	37
2	*раз(о)-/рас-*: distributive-summary	24	2	26
3	*по-*: ingressive	23	0	23
4	*на-*: cumulative	19	12	31
5	*о(бо)-:* distributive-summary	8	8	16
6	*раз(о)/с-...(-ся):* ingressive-intensive	7	7	14
7	*про-:* perdurative	3	24	27
8	*пере-:* distributive	2	8	10
9	*за-:* inchoative	2	13	15
10	*под(о)-*: attenuative	1	0	1
11	*за-...-ся*: excessive-durative	1	2	2
12	*до-...-ся*: durative-negative	1	4	5
13	*от(о)-*: finitive	1	16	17
14	*по-*: delimitative	1	23	24
15	*пере-*: repetitive	0	1	1
16	*по-*: distributive	0	1	1
17	*из(о)/с-…-ся*: excessive-multiple	0	2	2
18	*у-*: excessive-intensive	0	6	6
19	*на-*: saturative	0	11	11

Table 2: Number of As from *Dir* and *NonDir* according to MAS

[5] The MV-derivatives not included in the As are mainly perfectives with prefixes that realize their spatial meanings (*вбежать*, *забрести*, *отплыть*).

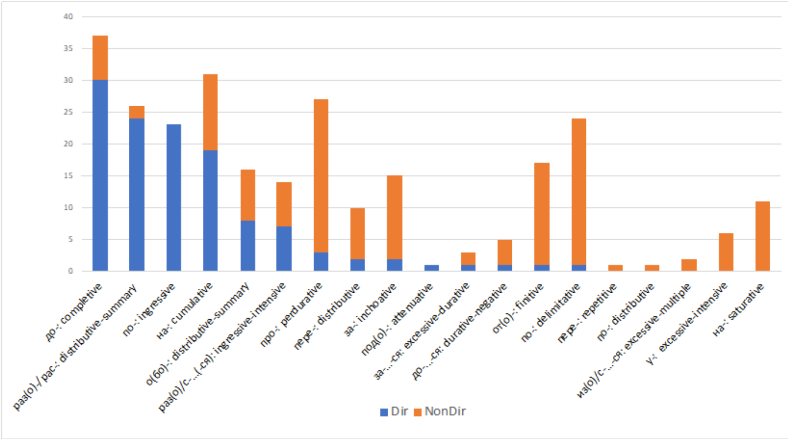

Figure 1: Distribution of As from *Dir* and *NonDir* according to MAS

Fig. 1 shows that there are more derivatives from *NonDir* than derivatives from *Dir*. According to MAS, lexemes belonging to 14 out of 19 As are found among derivatives from *Dir*, while representatives of 17 As are found among derivatives from *NonDir* (only the *по*-inchoatives and *под*-attenuatives are missing). At the same time, some of the derivatives from *Dir* belong to those As for which such derivation is postulated to be impossible in the literature. This applies for the *на*-cumulative (*набежать* 1. To congregate, accumulate, gather somewhere, *натащить* 1. Bring in some (usually large) amount of sth., *по*-delimitative (*понести* 5. Carry for a while sb, sth), *про*-perdurative (*проплыть* 3. Swim/sail for a period of time), *за*-inchoative (*занестись*[2] Start laying (about a chicken)).

The association plot[6] (Fig. 2) shows which As are significantly over- and underrepresented within *Dir* and *NonDir*. The As are labeled with numbers according to their numbering in Table 2.

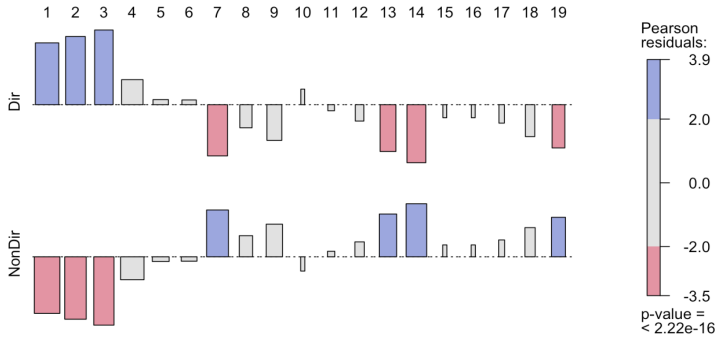

Figure 2: Association of *Dir/NonDir* and As according to MAS

According to Fig. 2, seven As show a significant association with the stem type. These are As that include non-unique lexemes (more than 10) with a strong tendency to derive from *Dir* or *NonDir*. The *до*-completive, *раз*(о)/*рас*-distributive-summary and *по*-ingressive are overrepresented when derived from *Dir* and underrepresented when derived from *NonDir*. In contrast, for the *про*-perdurative, *по*-delimitative, *от*(*о*)-finitive, and *на-...ся*-saturative, we observe an increase in productivity when derived from *NonDir*, while productivity decreases when derived from *Dir*.

To some extent, the data from MAS challenge the notion that certain As are strictly "fixed" to one of the stem types of MVs – either *Dir* or *NonDir*. Nevertheless, there remains a strong tendency for specific As to be formed predominantly from a single stem. Table 3 illustrates that when considering pairs of *Dir/NonDir* as the unit of analysis, instances of derivation from both types are relatively rare.

[6] Association plots are built with the assoc() function from the "vcd" package in R.

Specifically, among the 153 cases of As formation from a given *Dir/NonDir* pair documented in MAS, 79 cases involve derivatives exclusively from *NonDir*[7], 43 cases involve derivatives solely from *Dir*, and 31 cases include derivatives from both stem types.

As	***Dir* only**	***NonDir* only**	***Dir*&*NonDir***	**Total**
до-: completive	8	0	6	**14**
про-: perdurative	1	11	2	**14**
по-: ingressive	14	0	0	**14**
по-: delimitative	0	13	1	**14**
на-: cumulative	5	0	8	**13**
раз(*о*)-/*рас-*: distributive-summary	10	0	2	**12**
от(*о*)-: finitive	0	11	1	**12**
на-: saturative	0	11	0	**11**
за-: inchoative	0	8	2	**10**
раз(*о*)-/*рас-*…(*-ся*): ingressive-intensive	2	4	2	**8**
о(*бо*)-: distributive-summary	2	2	4	**8**
пере-: distributive	0	6	1	**7**
у-: excessive-intensive	0	6	0	**6**
до-....-ся: durative-negative	0	3	1	**4**
за-....-ся: excessive-intensive	0	1	1	**2**
пере-: repetitive	0	1	0	**1**
из(*о*)/с-…-*ся*: excessive-multiple	0	1	0	**1**
под(*о*)-: attenuative	1	0	0	**1**
по-: distributive	0	1	0	**1**
Total	**43**	**79**	**31**	**153**

Table 3: Derivatives from *Dir*/*NonDir* pairs within As in MAS

The existence of a strong relationship between the stem types (*Dir* or *NonDir*) and the As they produce is further corroborated through the application of machine learning algorithms.

First, the application of t-SNE (t-distributed stochastic neighbor embedding) to the As derived from *Dir* and *NonDir* reveals an almost unmistakable[8] division of 28 non-derivative stems into two distinct groups, as illustrated in Fig. 3.

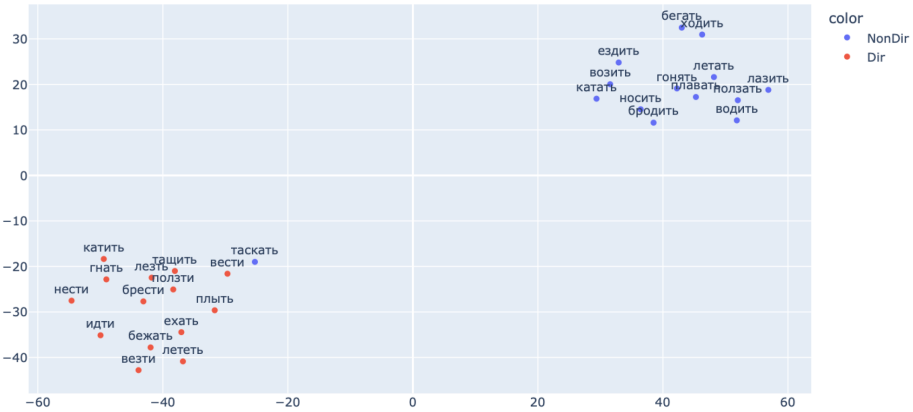

Figure 3: T-SNE projection for *Dir* and *NonDir* according to MAS

Furthermore, when training classification models (e.g., logistic regression and decision trees) on the data in which 270 lexemes recorded in MAS are treated as instances and As they belong to are used as features, the stem type (*Dir/NonDir*) is predicted with high metric values.

[7] Table 3 illustrates the presence of derivatives from paired *Dir* and *NonDir*, regardless of the number of lexemes involved.
[8] The only exception – *маскать* – requires further analysis.

	precision	recall	f1-score	support
NonDir	0.86	0.82	0.84	153
Dir	0.78	0.82	0.80	117
accuracy			0.82	270
macro avg	0.82	0.82	0.82	270
weighted avg	0.82	0.82	0.82	270

	precision	recall	f1-score	support
NonDir	0.86	0.83	0.84	152
Dir	0.79	0.82	0.80	118
accuracy			0.83	270
macro avg	0.82	0.83	0.82	270
weighted avg	0.83	0.83	0.83	270

Table 4: Classification reports for logistic regression (left) and decision tree (right) using As features with lexemes recorded in MAS as objects

Thus, the data exclusively from MAS confirm the relationship between the stem type and the derivation of As.

3.3 Sample expansion through As derived from MVs not recorded in MAS

The analysis in section 3.2 relies exclusively on verb lexemes recorded in MAS. However, examples of MV-derivatives MV – As absent from MAS – can be found in the Russian National Corpus (RNC)[9] and on the Runet (Yandex and/or Google), as shown (1)–(2). Analyzing these instances allows for preliminary conclusions about the actual productivity of As derived from MVs and the extent to which the representation in MAS and literature aligns with real speech practices.

(1) *А вот именно взять за руку и* ***повести*** *какое-то время не в сторону, а туда же, куда она сама шла ... не нашлось Мужика*.[10] (http://yanka.lenin.ru/stat/koblov.htm)

(2) *За ухи их брали и в воду тащили, они* ***поплывут-поплывут*** *и ворочаются.* [Светлана Василенко. Ген смерти (1997-2000)]

As a result of additional searches in the RNC and on the Runet for derivatives of MVs not recorded in MAS, we analyzed 153 instances of As formation, with the unit of analysis being the pair of *Dir* and *NonDir* stems. The distribution shows that most pairs allow for derivatives from both stems (100), while instances formed solely from *NonDir* are the least numerous (24), as illustrated in Table 4.

As	***Dir* only**	***NonDir* only**	***Dir*&*NonDir***	**Total**
до-: completive	1	0	13	**14**
про-: perdurative	0	2	12	**14**
по-: ingressive	14	0	0	**14**
по-: delimitative	0	0	14	**14**
на-: cumulative	0	0	13	**13**
раз(*о*)-/*рас-*: distributive-summary	10	0	2	**12**
от(*о*)-: finitive	0	3	9	**12**
на-: saturative	0	2	9	**11**
за-: inchoative	0	1	9	**10**
раз(*о*)-/*рас-*…(*-ся*): ingressive-intensive	2	4	2	**8**
о(*бо*)-: distributive-summary	1	0	7	**8**
пере-: distributive	0	5	2	**7**
у-: excessive-intensive	0	5	1	**6**
до-....-ся: durative-negative	0	0	4	**4**
за-....-ся: excessive- intensive	0	0	2	**2**
пере-: repetitive	0	0	1	**1**
из(*о*)/с-…*-ся*: excessive-multiple	0	1	0	**1**
под(*о*)-: attenuative	1	0	0	**1**
по-: distributive	0	1	0	**1**
Total	**29**	**24**	**100**	**153**

Table 5: Derivatives from *Dir*/*NonDir* pairs within As in the extended sample

[9] ruscorpora.ru

[10] We concur with the anonymous reviewer that non-codified usages arise by analogy with codified forms found in dictionaries, such as «надышаться», «отпеть (свои песни)», and «налетать (тысячу км)». We also agree that their acceptability can vary significantly among native speakers.

Considering data on derivatives from the RNC and the Runet, the quantitative ratio of derivatives from *Dir* and *NonDir* changes. Table 6 and Fig. 4 show a single preferred type of stem: *Dir* for the *no*-ingressive and *раз(о)-/рас*-distributive-summary As.

	As	***Dir***	***NonDir***	**Total**
1	*до-*: completive	30	14	44
2	*раз*(*о*)-/*рас-*: distributive-summary	24	2	26
3	*по-*: ingressive	23	0	23
4	*на-*: cumulative	19	18	37
5	*по-*: delimitative	19	23	42
6	*про-:* perdurative	13	25	38
7	*о(бо)-:* distributive-summary	11	9	20
8	*на-...-ся*: saturative	11	11	22
9	*за-:* inchoative	10	13	23
10	*от*(*о*)-: finitive	10	16	26
11	*раз(о)/с-...(-ся):* ingressive-intensive	7	7	14
12	*до-...-ся*: durative-negative	5	4	9
13	*пере-:* distributive	3	8	11
14	*за-...-ся*: excessive-durative	2	2	4
15	*под*(*о*)-: attenuative	1	0	1
16	*пере-*: repetitive	1	1	2
17	*у-...-ся*: excessive-intensive	1	6	7
18	*по-*: distributive	0	1	1
19	*из*(*о*)/*с-…-ся*: excessive-multiple	0	2	2

Table 6: Number of As from *Dir* and *NonDir* according to the extended sample

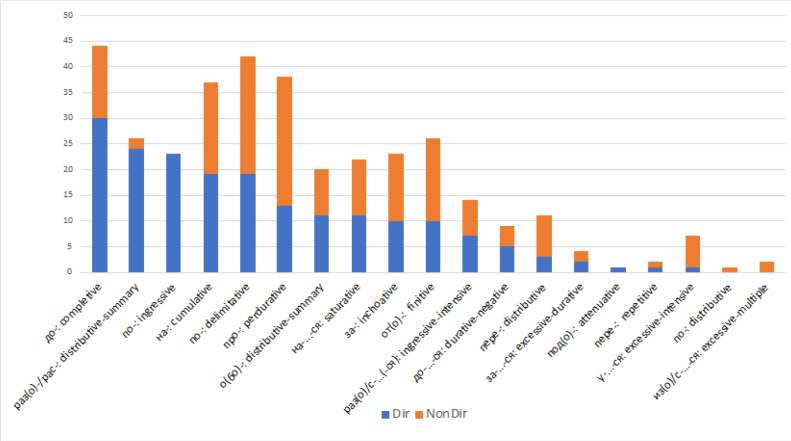

Figure 4: Distribution of As from *Dir* and *NonDir* according to the extended sample

Fig. 5 (with As numbered as in Table 6) confirms a significant association with the type of stem only for two As: the *no*-ingressive and *раз*(*о*)-/*рас*-distributive-summary. Both are overrepresented when derived from *Dir* stems and underrepresented when derived from *NonDir* stems. Notably, there are no As that demonstrate a high level of derivation from *NonDir*, contradicting the idea that many As are formed exclusively from *NonDir*.

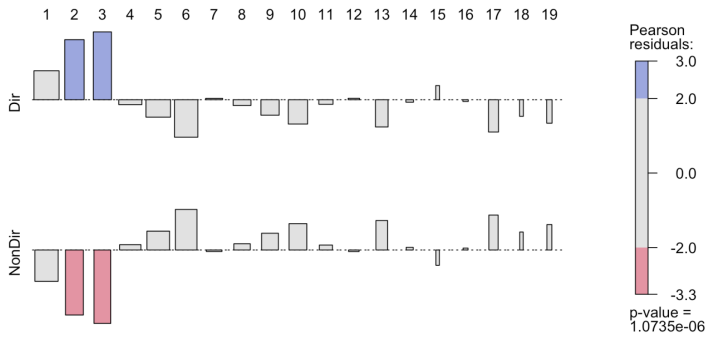

Figure 5: Association of *Dir/NonDir* and As according to the extended sample

Applying the same analytical tools used on the MAS data to the dictionary data supplemented with derivatives not recorded in MAS but found in the RNC and the Runet reveals a weakened dependence between the stem type and resulting As. Specifically, t-SNE analysis of the extended data shows that *Dir* and *NonDir* no longer exhibit a clear division into two groups (see Fig. 6).

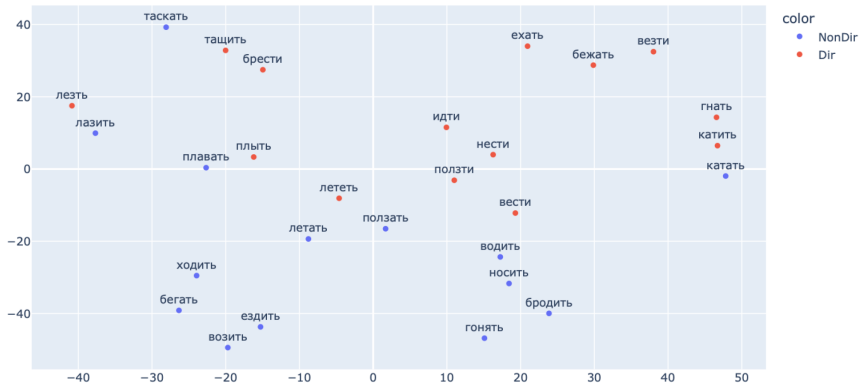

Figure 6: T-SNE projection for *Dir* and *NonDir* in the extended sample

The prediction quality of models trained on the extended data, considering only belonging to As, is significantly lower than that for the MAS data trained on the same features (see Table 7).

	precision	recall	f1-score	support
NonDir	0.58	0.63	0.60	150
Dir	0.71	0.66	0.68	202
accuracy			0.65	352
macro avg	0.64	0.65	0.64	352
weighted avg	0.65	0.65	0.65	352

	precision	recall	f1-score	support
NonDir	0.71	0.60	0.65	192
Dir	0.59	0.71	0.65	160
accuracy			0.65	352
macro avg	0.65	0.65	0.65	352
weighted avg	0.66	0.65	0.65	352

Table 7: Classification reports for logistic regression (left) and decision tree (right) using As features with lexemes the from the extended sample as objects

The analysis of As derived from MVs found in MAS, supplemented with derivatives from the RNC and the Runet, reveals that relying solely on the stem type (*Dir* or *NonDir*) is insufficient to determine the formation of a specific A.

3.4 Consideration of extra features

Thus, when considering derivatives not recorded in MAS but found in the RNC and the Runet, the distribution of *Dir* and *NonDir* stems, as well as their association with certain As, differs from that

obtained through analysis of derivatives exclusively in MAS. Additionally, in some cases, it is evident that when forming As from lexemes with non-literal meanings of MVs, derivation from *Dir* is the only option, particularly noticeable in examples from the RNC and the Runet, e.g., (3)–(4).

(3) *Они решили* ***погнать*** *самогона ночью. Как-то заткнули неправильно, аппарат взорвался, муж весь обварился.* (https://culture.wikireading.ru/2408)

(4) *А эти последние три года я мужа упрашивала поработать, когда нас просили провести мероприятие. И Стас шел на это, чтобы я напелась и* ***навелась*** *праздника от души!* (https://vk.com/wall386557704_5459)

Let us analyze the relationship between the source of documentation and semantic features on the one side and the formation of As from *Dir* and *NonDir* on the other.

First, it is important to note that existing dictionary descriptions do not provide a strict tool for identifying the metaphoric meanings of MV-derivatives, as discussed in Appendix D. Consequently, the authors independently annotated based on intuition, achieving a Cohen's kappa agreement coefficient of 0.76. The first author adopted a broader view of semantic metaphoricity, with non-literal meanings comprising 30.68% (108 lexemes), while the second identified metaphoric meanings in 21.31% of derivatives (75 lexemes). It was decided not to unify the annotations but to treat both variants as distinct features, as their association with *Dir* and *NonDir* stems may differ.

Training that incorporates sources of derivatives (MAS, the RNC, the Runet) and their semantic features results in significantly better prediction quality compared to training solely on the derivatives' belonging to As. The best outcomes from the logistic regression and decision tree models are achieved when both features are included.

Logistic regression achieved the best results when trained on the source features (with the "Runet" binary feature removed due to strong linear dependence with "MAS") and the semantic characteristics based on the second author's annotations. The metric values are shown in Table 8.

	precision	recall	f1-score	support
NonDir	0.84	0.78	0.81	174
Dir	0.80	0.85	0.83	178
accuracy			0.82	352
macro avg	0.82	0.82	0.82	352
weighted avg	0.82	0.82	0.82	352

Table 8: Classification report for logistic regression using As, source and semantics as features with lexemes from the extended sample as objects

The decision tree shows the best prediction results when trained with the source and semantic annotations from both authors (see Table 9), with the source being the most significant feature at 28.25% (by Gini criterion).

	precision	recall	f1-score	support
NonDir	0.85	0.88	0.87	157
Dir	0.90	0.88	0.89	195
accuracy			0.88	352
macro avg	0.88	0.88	0.88	352
weighted avg	0.88	0.88	0.88	352

Table 9: Classification report for decision tree using As, source and semantics as features with lexemes from the extended sample as objects

Next, we examine the interaction between a stem type and semantic features in forming As from MVs. Among derivatives from *Dir*, the proportion of lexemes with non-literal meanings is higher in both annotation variants ("non-literal _1" and "non-literal _2"). This difference is statistically significant only for derivatives supplemented with verbs not recorded in MAS but found in the RNC and the Runet, and

marginally significant for As from MVs based on the dictionary data (see Table 10). The specific stems associated with literal / non-literal meanings also differ in the two annotation variants (see Appendix C).

non-literal_1							
	literal			non-literal			Statistical test
	Dir	*NonDir*	*% Dir*	*Dir*	*NonDir*	*% Dir*	
MAS	43	31	58.11	80	116	40.82	$\chi2$=5.798, p=0.016
Extended sample	75	33	69.44	115	129	47.13	$\chi2$=14.12, p<<0.01
non-literal_2							
	literal			non-literal			Statistical test
	Dir	*NonDir*	*% Dir*	*Dir*	*NonDir*	*% Dir*	
MAS	33	24	57.89	90	123	42.25	$\chi2$=3.827, p=0.0504
Extended sample	51	24	68.00	139	138	50.18	$\chi2$=6.844, p=0.008

Table 10: Distribution of literal and non-literal meanings of As derivatives from *Dir* and *NonDir*

Thus, the analysis of As derived from MVs, encompassing lexemes from MAS, the RNC, and the Runet, underscores the significance of the source of the verb and its semantic characteristics (literal vs. non-literal meaning) in evaluating As formation from *Dir* and *NonDir* stems.

4 Conclusion

Let us summarize the findings of the study on prefixal and circumfixal MV-derivatives in relation to morpheme-characterized As.

The analysis above shows that the restrictions on forming certain As from MVs, presented in [15: 102–103; 16: 6-9] as the complementary distribution of prefix meanings based on the stem type (*Dir*, *NonDir*), represent a tendency rather than an absolute rule. Both the comprehensive sampling of As from MAS and additional searches in the RNC and the Runet reveal perfective MV-derivatives that challenge established limitations on deriving As from these stems. Therefore, we can assert that there are no strict prohibitions on specific derivations, only tendencies that vary in intensity.

For MV-derivatives recorded in MAS, there is a strong tendency to form certain As from only one stem. However, when analyzing derivatives not recorded in MAS, this tendency weakens.

It is reasonable to assume that the stem type is not the sole factor influencing the derivation of As. The source of the derivative's documentation (MAS, the RNC, and/or the Runet) and its proximity to the literal meaning of the root MV also affect the derivational potential of *Dir* and *NonDir*.

Acknowledgements

The research was funded by a grant from the Russian Science Foundation, № 24-78-00172 (https://rscf.ru/en/project/24-78-00172/), with Oksana Chuikova as the grant recipient.

The authors would like to thank Ekaterina Enikeeva for her helpful advice on the application of machine learning methods.

References

[1] Apresyan Yu. D. Foundations of systemic lexicography [Osnovaniya sistemnoj leksikografii] // Linguistic picture of the world and systemic lexicography [Yazykovaya kartina mira i sistemnaya leksikografiya]. Ed. by Yu. D. Apresyan. — Moscow: Languages of Slavic Culture [Yazyki Slavyanskoi Kul'tury], 2006. — P. 33–160.

[2] Chuikova Oksana Iu. (2020), On the secondary imperfectivation of *po*-perfectives in Russian [Ob osobennostyakh vtorichnoj imperfektivatsii glagolov s prefiksom *po-* v russkom yazyke], Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue", Vol. 19 (26), pp. 160–176.

[3] Chuikova Oksana Iu. (2022), Partitive Genitive and Aktionsarten in Russian (based on dictionary and corpus data) [Roditel'nyj partitivnyj i sposoby glagol'nogo dejstviya v russkom yazyke (po slovarnym i korpusnym dannym)], Russian Linguistics, no. 46(1), pp. 25–54.

[4] Evgenieva Anastasija P. (ed.). Dictionary of Russian Language in 4 volumes [Slovar' russkogo jazyka v 4 tomah], 2nd ed. — Moscow: Academy of Science of USSR, Institute of Russian Language, 1981–1984. Access mode: http://feb-web.ru/feb/mas/mas-abc/14/ma239217.htm
[5] Gorbova Elena V., Chuikova Oksana Iu., Sharygina Sofya S. (2021), Imperfectivability of Russian prefixal perfectives: regularity and peculiarities [Imperfektiviruemost' russkikh pristavochnykh perfektivov: regulyarnost' i specifika], Topics in the study of language [Voprosy Jazykoznanija], no. 4, pp. 91–130.
[6] Gorbova Elena. V. (2023), Russian prefixed verbs of motion: aspectual correlativeness and aspect formation [Pristavochnye glagoly dvizheniia v russkom iazyke: vidovaia korreliativnost' i vidoobrazovanie], Scando-Slavica, no. 69(2), pp. 219–243.
[7] Gorbova Elena. V. (in press), Prefixal derivatives of Russian verbs of motion: word and aspect formation [Prefiksal'nye derivaty glagolov dvizheniia v russkom iazyke: slovo- i vidoobrazovaniie]. Vestnik of Saint Petersburg University. Language and Literature.
[8] Grammar of Russian language [Grammatika russkogo yazika]. Ed. by V. V. Vinogradov V. 1. — Moscow: Izdatel'stvo Akademii nauk SSSR [Publishing house of Academy of Sciences of the USSR], 1952.
[9] Isachenko A. V. The grammatical structure of the Russian language in comparison with Slovak. Morphology [Grammaticheskij stroj russkogo jazyka v sopostavlenii so slovackim. Morfologija]. Vol. 2. — Bratislava: Slovak Academy of Sciences Publ. [Izdatel'stvo Slovatskoi akademii nauk], 1960.
[10] Janda Laura. Perfectives from indeterminate motion verbs in Russian // Hasko V., Perelmutter R. (eds.). New Approaches to Slavic Verbs of Motion. — Amsterdam / Philadelphia: John Benjamins, 2010. — P. 125–139.
[11] Mel'cuk, I. (1990), Slovoobrazovanie v lingvisticheskix modelyax tipa "Smysl ⇔ Tekst" (Word formation in linguistic models of the type "Meaning ⇔ Text" (preliminary remarks)] // Metody formalne w opisie języków słowiańskich. Ed. by Z. Saloni. — Białystok: Dział Wydawnictw Filii UW w Białymstoku, 1990. — P. 47–74.
[12] Russian grammar [Russkaya grammatika]. Ed. by N. Yu. Shvedova. — Moscow: Nauka, 1980. Vol. 2.
[13] Shelyakin M. A. Category of aspectuality of the Russian verb [Kategoriya aspektual'nosti russkogo glagola]. — Moscow: URSS, 2008.
[14] Townsend, C. E. Russian Word Formation. — Ohio: Slavica Publishers Inc., 1975.
[15] Zalizniak Anna A., Mikaelyan Irina L., Shmelev Aleksey D. Russian aspectology: In defense of the aspectual pair [Russkaya aspektologiya: v zashchitu vidovoi pary] — Moscow: Languages of Slavic Culture [Yazyki Slavyanskoi Kul'tury], 2015.
[16] Zalizniak Anna A. (2021), Russian Prefixed Verbs of Motion Revisited, Zeitschrift für Slawistik, no. 62(1), pp. 1–22.

Приложение A. List of As

As	Examples	Examples of MVs ГД
в(о)з(о)-/в(о)с-: delimitative	*всплакнуть, взгрустнуть*	———
до-: completive	*дочитать, долить*	*долезть*, *добродить*
до-...-ся: durative-negative	*доиграться, долежаться*	*добегаться, доездиться*
за-: inchoative	*закричать, заплакать*	*забегать*, *занестись*
за-...-ся: excessive-durative	*засидеться, залюбоваться*	*забегаться, заходиться*
из-...-ся: excessive-multiple	*изголодаться, изленитьcя*	*избегаться*
на-: cumulative	*накупить, наловить*	*налетать, натащить*
на-...-ся: saturative/sative	*наесться, наиграться*	*набегаться, навозиться*
о(бо)-: distributive-summary	*опросить, обзвонить*	*обегать, облететь*
от(о)-: finitive	*отшуметь, отмучиться*	*отбродить, отплавать*
пере-: distributive	*перебудить, перестрелять*	*перевестись, перетаскать*
пере-: repetitive	*пересочинить, перезаказать*	*переходить*
по-: attenuative	*пообождать, попривыкнуть*	(*поразогнать*)
по-: delimitative	*погулять, поиграть*	*полазить, понести*
по-: distributive	*попрятаться, понаписать*	*потаскать* (*повылезти*)
по-: ingressive	*помчаться, полюбить*	*поехать*, *повести*
под-: attenuative	*подкрасить, подсократить*	*подвезти*
при-: attenuative	*припудрить, приоткрыть*	———
про-: perdurative	*просидеть, проработать*	*проплавать, проездить*
раз(о)-/рас-: distributive-summary	*разворовать, растерять*	*растащить, разбежаться*
раз(о)/с-...(-ся): ingressive-intensive	*раскричаться, разволноваться*	*разбегаться, разойтись*
у-...-ся: excessive-intensive	*убегаться, упахаться*	*убегаться, уходиться*

Appendix B[11] The distribution of As within the subset of MVs and beyond it

	MVs			non-MVs			
	A	**non A**	**% A**	**A**	**non A**	**% A**	Statistical test
до-: completive	37	1267	2.84	209	16713	1.24	$\chi2=22.158$, $p<<0.01$
о(бо)-: distributive-summary	16	1288	1.23	15	16907	0.09	$\chi2=85.815$, $p<<0.01$
по-: ingressive	23	1281	1.76	46	16876	0.27	$\chi2=67.556$, $p<<0.01$
раз(о)-/рас-: distributive-summary	26	1278	1.99	117	16805	0.69	$\chi2=24.737$, $p<<0.01$
у-...-ся: excessive-intensive	6	1298	0.46	12	16910	0.07	$\chi2=14.853$, $p<<0.01$
за-: inchoative	15	1289	1.15	462	16460	2.73	$\chi2=11.245$, $p<<0.01$
пере-: distributive	10	1294	0.77	306	16616	1.81	$\chi2=7.108$, $p=0.0077$
пере-: repetitive	1	1303	0.08	211	16711	1.25	Fisher's exact test, $p<<0.01$
по-: delimitative	24	1280	1.84	532	16390	3.14	$\chi2=6.520$, $p=0.0107$
по-: distributive	1	1303	0.08	83	16839	0.49	Fisher's exact test, $p=0.0306$
под-: attenuative	1	1303	0.08	158	16764	0.93	Fisher's exact test, $p<<0.01$
до-...-ся: durative-negative	5	1299	0.38	36	16886	0.21	Fisher's exact test, $p= 0.21$
на-: cumulative	31	1273	2.38	368	16554	2.17	$\chi2=0.147$, $p=0.70$
от-: finitive	17	1287	1.30	130	16792	0.77	$\chi2=3.695$, $p=0.06$
про-: perdurative	27	1277	2.07	320	16602	1.89	$\chi2=0.124$, $p=0.72$
раз(о)-/рас-...-ся: ingressive-intensive	14	1290	1.07	180	16742	1.06	$\chi2=0$, $p=1$
из-...-ся: excessive-multiple	2	1302	0.15	27	16895	0.16	Fisher's exact test, $p=1$
за-...-ся: excessive-durative	3	1301	0.23	47	16875	0.28	Fisher's exact test, $p=1$
на-...-ся: saturative	11	1293	0.84	169	16753	1.00	$\chi2=0.160$, $p=0.69$

[11] Each row in the table of the Appendix B represents a transformation of a contingency table with positive and negative values for two features: belonging to a specific A (indicated in the first column) – "A" and "non-A", and belonging to MVs – "MVs" and "non-MVs". Thus, the representation of this A within the set of MVs is compared to similar data for this A outside of MVs (the sum of values in the "A" and "not A" columns within MVs is consistently 1304, representing the total volume of the analyzed MVs sample; the sum of values for "A" and "non-A" outside of MVs is consistently 16922, which equals the total number of analyzed prefixed perfectives minus the MVs (18226 minus 1304). The last column of the table presents the results of applying Pearson's chi-squared test (with Yates' correction, two-tailed variant) or Fisher's exact test (two-tailed variant; used when at least one value does not exceed 5). The color labeling indicates the presence or absence of statistical significance in the distribution differences (at $p < 0.05$) and the direction of the trend (increased or decreased productivity of the examined A within the group of MV-derivatives): in cases of statistical significance, this is determined by comparing the proportions of verbs for this A within the set of MVs and outside it (grey shading indicates no statistical significance, green indicates significant differences where the proportion of A within MVs is higher than that outside, and orange indicates significant differences where the proportion of A within MVs is lower than that outside).

Appendix C Frequency literal / non-literal (metaphoric) meanings of As derivatives in relation to root MVs

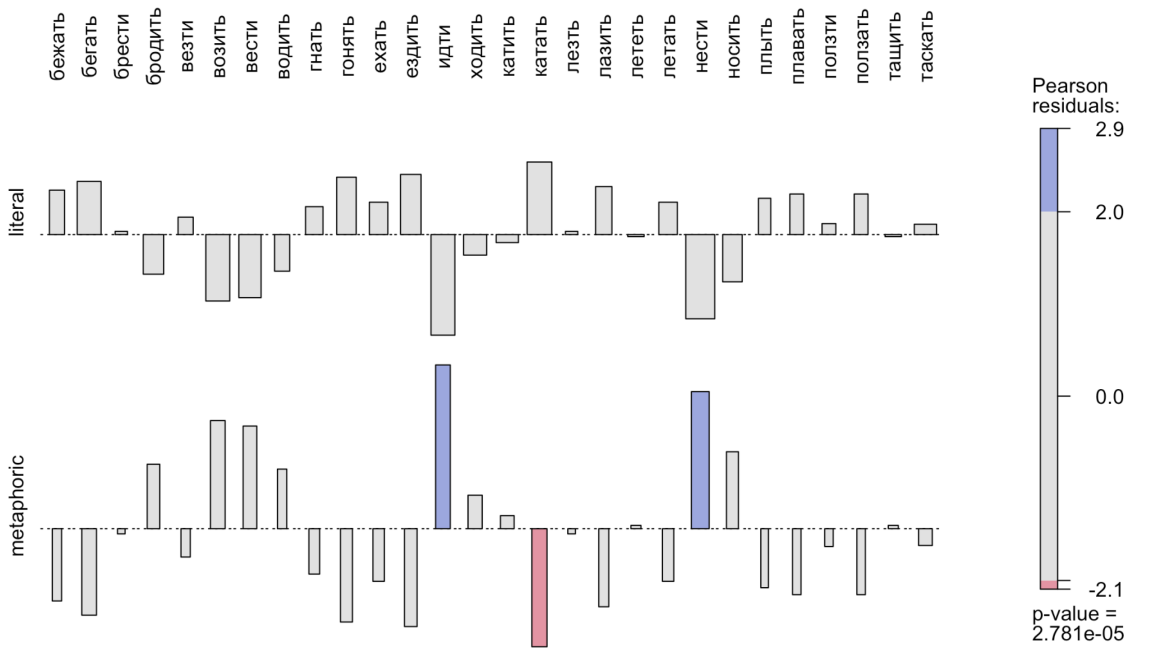

Figure C1: Association plot of root MVs and semantic characteristics based on MAS (Author 1)

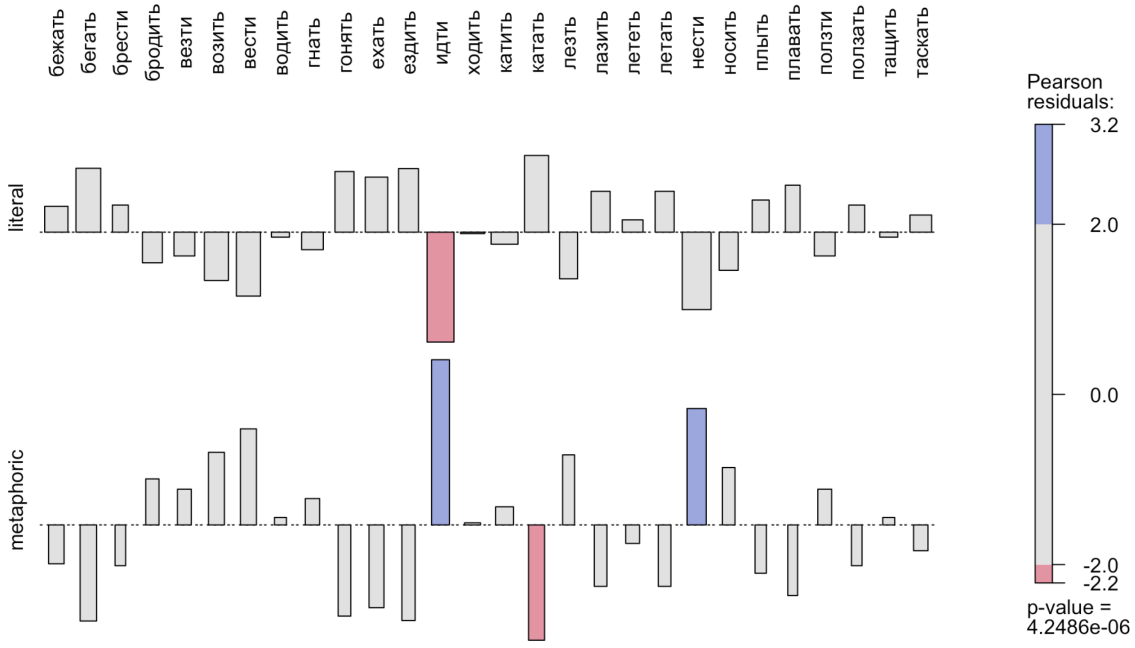

Figure C2: Association plot of root MVs and semantic characteristics based on the extended data (Author 1)

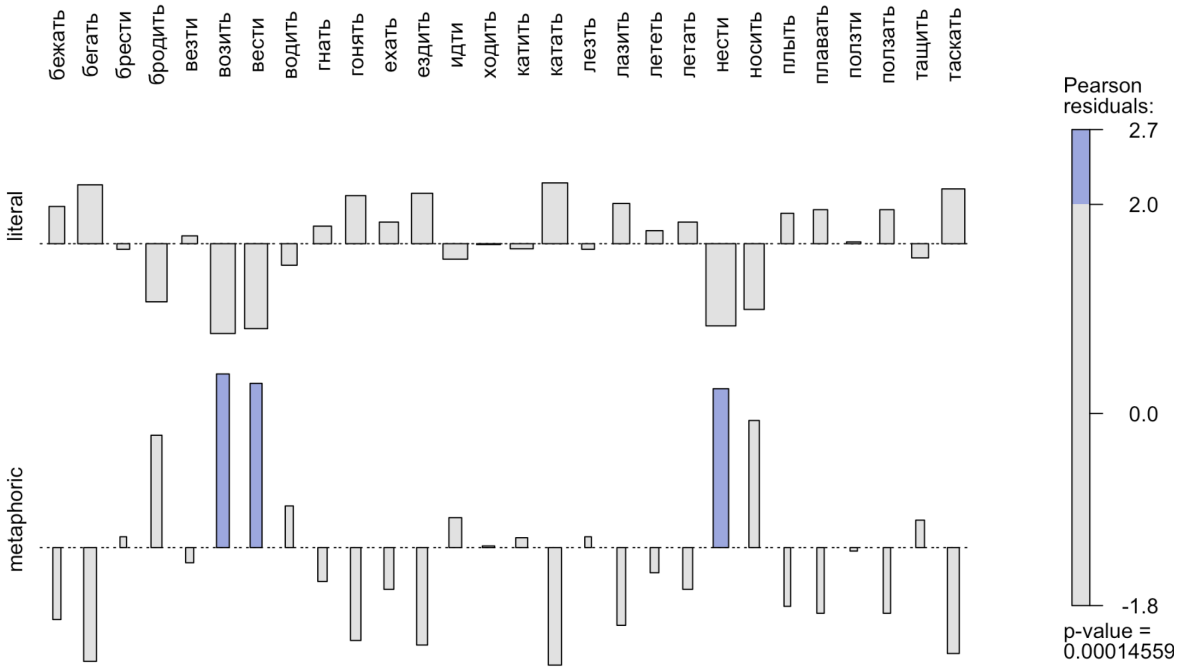

Figure C3: Association plot of root MVs and semantic characteristics based on MAS (Author 2)

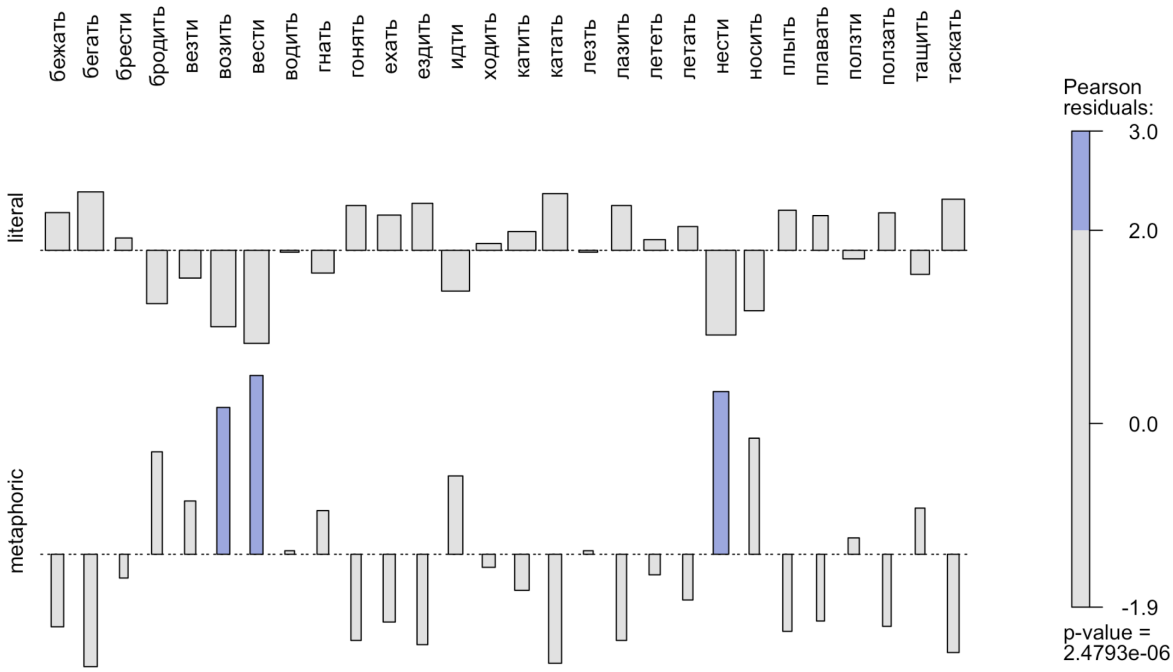

Figure C4: Association plot of root MVs and semantic characteristics based on the extended data (Author 2)

Appendix D The Issue of qualifying the meaning of the lexeme as non-literal (metaphoric)

When attempting to apply the traditional distinction in lexical semantics between literal and non-literal (metaphoric) meanings of a lexical unit, certain difficulties arise due to the following reasons.

1. The label "metaphoric" (*перен*.) is present in Russian language dictionaries created in the 20th century but is absent in more recent dictionaries. For example, it is used in "The Explanatory Dictionary of the Russian Language" by D. N. Ushakov (1935), "The Dictionary of the Russian Language" by S. I. Ozhegov (1949), and "The Dictionary of the Russian Language in 4 volumes", edited by A. P. Evgenyeva (MAS), but not in "The New Explanatory Dictionary of Synonyms of the Russian Language" edited by Yu. D. Apresyan or "The Large Dictionary of the Russian Language" by S. A. Kuznetsov. Additionally, the concept of metaphoric meaning is not relevant for semantic tagging in RNC, see https://ruscorpora.ru/page/instruction-semantic/.
2. Even in cases where the label "metaphoric" (*перен*.) is included in the dictionary's labeling system, the principles for its application are not specified in the relevant sections of the dictionaries (usually found in "How to Use the Dictionary"). This is particularly true for MAS, whose dictionary serves as the basis for this study, where it is stated only that: "§ 16. The metaphoric meaning (or shade of meaning) is accompanied by the label *перен*., which precedes all other labels (if any) and the definition of meaning".
3. Utilizing the decisions made by the dictionary compilers, particularly those of MAS, as a tool for annotation in this research proved to be unfeasible. This is because their decisions, lacking any explicit algorithm for their formulation, can only be regarded as intuitive judgments, which may not necessarily align with the intuitions of other researchers. This can be illustrated with two examples. First, one of our esteemed anonymous reviewers suggested that the derivative *занестись* from *нестись* 'to lay eggs (of birds)' in MAS and 'about birds: to lay eggs' in "The Dictionary of the Russian Language" by S. I. Ozhegov is derived from the metaphoric meaning of the original verb. However, this claim is not supported by either MAS or Ozhegov, as *нестись* does not carry the label *перен*. in either dictionary, even though it is listed as the third meaning of a polysemous word in MAS and appears as a separate entry in "The Dictionary of the Russian Language" by S. I. Ozhegov. The second example pertains to our subset of derivatives that represent As. According to MAS, only 9 of these derivatives are labeled *перен*., while Authors 1 and 2, in their independent assessments of metaphoric meanings within this subset, identified several dozen derivatives as metaphoric lexemes. However, their agreement coefficient during annotation (Cohen's kappa) was substantial but not perfect, measuring 0.76.