

Robert Lagerberg

**The Relationship Between Frequency
and Word Stress in Russian:
A Case Study of the Verbal Suffix *-ировать*¹**

1. Introduction

In a previous article (Lagerberg 2003) a detailed analysis of the stress characteristics of the Russian verbal suffix *-ировать* was carried out. The article identified both a general historical shift (traceable in lexicological sources) towards the initial suffixal syllable from the final syllable, e.g. *лавировать* > *лави́ровать* ‘to tack’, as well as a current default or preferred non-final stress on the initial suffixal syllable (e.g. *телефони́ровать* ‘to telephone’), particularly evident in words more recently entering the language. However, in certain semantic categories (see below) this trend appears to be resisted quite strongly, so that in standard normative sources of stress (e.g. Zaliznjak 1977), many of these lexemes are given with final stress (e.g. *пломбирова́ть* ‘to fill (a tooth)’) or alternate stress positions (e.g. *татуи́ровать* ‘to tattoo’), and, therefore, with the result that this suffix represents a fluid area of Russian stress.

The fullest account of verbs in *-ировать* with final stress is given by Zaliznjak (1985, 105–106) as follows. The variant *-ировать* occurs:

1. With monosyllabic stems which have the meanings:
 - (a) to cover or fill with a certain substance, e.g. *лакирова́ть* ‘to varnish’;
 - (b) to cover one object with another, e.g. *маскирова́ть* ‘to mask’;
 - (c) to organise, exercise or improve, e.g. *группирова́ть* ‘to group’, *тренирова́ть* ‘to train’, *юсти́ровать* ‘to adjust (instruments)’.
2. In a few verbs with a polysyllabic stem with the meanings of 1a and 1b given above: *эмалирова́ть* ‘to enamel’, *костюмирова́ть* ‘to dress (in theatre)’.

¹ I would like to thank Nicolas Warren for carrying out the statistical analysis in this article, and also the School of Languages at the University of Melbourne for its award to me of a grant to cover the costs of this research assistance.

3. In some isolated cases which cannot be accounted for, e.g. *марширо́вать* ‘to march’, *серви́ровать* ‘to serve (meal)’, *кроки́ровать* ‘to rocket’.

In addition, a small number of verbs which satisfy the conditions required to belong to 1 (a, b, c) above (particularly those whose root ends in *-p*) have, in fact, initial suffixal stress, e.g. *хлори́ровать* ‘to chlorinate’, *гази́ровать* ‘to aerate’ (but *гази́рованный!*), *хроми́ровать* ‘to plate with chromium’ (but *хроми́рованный!*).² In verbs with polysyllabic roots and all other cases *-и́ровать* is the norm, as it is always in verbs with the compound suffix *-изи́ровать* (e.g. *автоматизи́ровать* ‘to automate’).

Two subsequent surveys by the present author (Lagerberg 2005 & 2006), however, demonstrated quite clearly that the normative description of the stress of *-ировать* was, in fact, not reflected for the most part in actual usage. According to the results of the former of these surveys, only 46.7% percent of words with final stress in Zaliznjak 1977 were given with this stress position by surveyed native speakers; in all other responses the stress was on the initial suffixal syllable or (very occasionally) both positions of stress were allowed (Lagerberg 2005, 45–46). In the latter survey, as few as 35% of words with normative final stress were actually given with this position by respondents (Lagerberg 2006, 200).

While the semantic categories mentioned above clearly appear to be playing a role in the retention of final stress, another factor identified was that of frequency. Indeed, frequency has often been connected with stress (see Kiparsky 1962, 295, in relation to this very suffix; see also Lagerberg 2007 for a discussion of frequency and stress variation in Russian), but not in any systematic (i.e. statistically quantified) way. In the case of *-ировать*, at a very general level at least, it seems that words belonging to one of Zaliznjak’s semantic categories associated with final stress (given above) are better able to retain final stress when combined with a relatively higher frequency, while such words with a relatively lower frequency are more likely to succumb, as it were, to the general tendency towards initial suffixal stress. This is, therefore, to all appearances, a case of rhyme analogy in progress, or, to use another approach, Anderson’s (1973) ‘abductive change’. According to this model, while the underlying linguistic model is one of rhyming stress uniformity for certain suffixes (i.e. all words with a given suffix have stress

² Both the forms marked here with an exclamation mark suggest derivation from forms with final stress, viz *газиро́вать**, *хроми́ровать**.

on the same syllable, in this case, *-úровать*), many speakers continue, to a greater or lesser extent (which is determined to a large extent, it would seem, by frequency and syllabic structure) to apply adaptive rules ('A-rules'); an A-rule is 'a stylistically motivated rule' (Andersen 1973, 773), in this case, a rule stipulating final stress (*-урова́ть*) in accordance with certain semantic/syllabic conditions. Andersen's model, if accepted, makes it clear that the older stress position (in the form of the A-rule), and, therefore, the stress variation itself, is ultimately doomed to extinction, whether sooner or later, in favour of the rhyming model, i.e. all verbs in *-уровать* will come to have initial suffixal stress in time.

So much is clear at a general level; the aim of this article is to examine more precisely the relationship between stress position and frequency using actual data and statistical analysis. If there is indeed a link between frequency and stress position, then verbs in *-уровать* which fall into one of Zaliznjak's categories associated with final stress present an ideal testing ground for such a theory: this sub-group of verbs is characterised by a range of frequency values (see below) and a choice of two stress positions, one of which (the non-normative one, i.e. final stress) is clearly affected by a certain factor or, perhaps, factors, often resulting in its shift (among native speakers) to the initial suffixal syllable.

2. Analysis of Data

To begin with, the full set of data in alphabetical order of the survey carried out in Lagerberg 2005 is given below in Table 1. This data will be used as the basis of the study.³ Fifteen respondents gave information (the total for each row). ST1 (i.e. stress type 1) represents initial suffixal stress (*-úровать*), ST2 (stress type 2) final stress (*-урова́ть*). Rows with .5 values indicate that one or more respondents allowed both stress positions, with the value 1 shared between both columns. The stress positions given in the first column correspond to those given in Zaliznjak 1977, which is used as a base point for normative stress position.

The purpose of this survey (described in Lagerberg 2005) was to examine a sample of verbs in *-уровать* with a preponderance of ST2, in order to ascertain the retention rate of the latter stress type (i.e. final stress). A number of ST1 verbs

³ Although surveys, particularly those on smaller scales, can never offer totally conclusive evidence, the evidence of this survey was markedly similar to another carried out by the present author (Lagerberg 2006).

Word	English	ST1	ST2
баллотировать	to vote	15	0
бронировать	to armour	10	5
бронировать	to reserve	13	2
вальсировать	to waltz	15	0
газировать	to aerate	12	3
глазировать	to glaze	10	5
гофрировать	to corrugate	15	0
гримировать	to make up (theatre)	2	13
группировать	to group	0	15
декатировать	to sponge (wool)	15	0
демаскировать	to unmask	6.5	8.5
дренировать	to drain	10.5	4.5
дрессировать	to train (animals)	0	15
копировать	to copy	12	3
костюмировать	to dress (theatre)	15	0
крокировать	to roquet	12	3
манкировать	to neglect, be absent	11	4
маркировать	to mark	7.5	7.5
маскировать	to mask	2	13
нивелировать	to level	15	0
нормировать	to regulate	8.5	6.5
пикировать	to nose-dive	14	1
пломбировать	to fill (tooth)	11.5	3.5
премировать	to award a prize to, give a bonus to	10.5	4.5
проектировать	to project, plan, design	15	0
сталировать	to plate with steel	12	3
татуировать	to tattoo	5.5	9.5
франкировать	to prepay	15	0
хлорировать	to chlorinate	15	0
хромировать	to plate with chromium	12	3
экипировать	to equip	6	9
юстировать	to adjust (instruments)	15	0

Table 1: All data from 2005 survey

were included in order to provide a counter-sample of verbs with initial suffixal stress in order to confirm initial suffixal stress as the default stress type as seen by its overall retention (which was indeed confirmed).

The main conclusions of this survey can be summarised as follows:

1. the survey demonstrated the high potential for variation in verbs with this suffix: only 12 of the 32 words given in the survey received uniform stress placement (i.e. stress in all responses either on the initial suffixal syllable *-ировать* or on the final syllable *-и́ровать*);
2. the survey provided confirmation of the clear tendency for ST1: 70% of all responses had stress on the initial suffixal syllable, a high percentage given that almost half of the words chosen for the survey were given as ST2 in Zaliznjak 1997. Of all words given with ST2 in Zaliznjak 1977, only 46.7% of responses favoured this stress position.

From this base corpus the words in Table 2 can be selected as belonging to one of Zaliznjak's categories associated with final stress.⁴ In one case (*демакировка́ть*), the interpretation is rather broad, but the word is included as a semantic counterpart to *маскирова́ть*.

With the accentual data now assembled, it remains to be seen how frequency and stress relate to each other. Although hitherto Zatorina (1977) has generally been regarded as the fundamental reference book on frequency in Russian, Sharoff's online Russian frequency list (RFL) now comprises the largest corpus to date, as well as being more contemporary in terms of the usage which it is based on.⁵ By way of comparison, Zatorina (1977) offers vastly inferior data in terms of the suffix being examined, and can be ignored for the purposes of the present study: the only words in *-ировать* to be listed in it are *баллотировать(ся)* (with a frequency of 1), *премировать* (4) and *экипировать* (2), as opposed to sixteen words listed in RFL.

Table 3 gives the distribution of responses to the words surveyed in Lagerberg (2005) together with the frequencies recorded in RFL.⁶ While the data received

⁴ The alphanumericals under 'Category' indicate which type according to the explanation given above, e.g. 1a indicates a monosyllabic verb denoting 'to cover or fill with a certain substance'; 2b indicates a polysyllabic verb with the sense 'to cover one object with another'.

⁵ *Общий частотный словарь лемм для современных (>1950) текстов текущей версии НКРЯ* at: <http://corpus.leeds.ac.uk/serge/frqlist/>

⁶ The value 0 indicates that the word is not in RFL.

Word	Category
бронировать	1a
газировать	1a
глазировать	1a
гофрировать	1a
гримировать	1a
группировать	1c
демаскировать	2b
дренировать	1c
дрессировать	1c
костюмировать	2b
маркировать	1a
маскировать	1b
нормировать	1c
пломбировать	1b
сталировать	1a
татуировать	2b
хлорировать	1a
хромировать	1a
экипировать	2b
юстировать	1c

Table 2: Words from 2005 survey with one of Zaliznjak's semantic categories (indicated in alphanumericals) associated with final stress

Word	RFL	ST1	ST2
маскирова́ть	336	2	13
экипи́ровать	91	6	9
дресси́ровать	86	0	15
татуи́ровать	76	5.5	9.5
марки́ровать	73	7.5	7.5
грими́ровать	58	2	13
демаски́ровать	49	6.5	8.5
брони́ровать	0	10	5
глази́ровать	0	10	5
дрени́ровать	0	10.5	4.5
пломби́ровать	0	11.5	3.5
гази́ровать	0	12	3
стали́ровать	0	12	3
хроми́ровать	0	12	3
гофри́ровать	0	15	0
костюми́ровать	0	15	0
хлори́ровать	0	15	0
юсти́ровать	0	15	0

Table 3: List of words with descending frequency values in RFL and stress types recorded in 2005 survey

from a survey of this or any other kind can never be considered absolutely definitive in terms of stress position, it does give a more representative picture of stress at any given time than a lexicological source, which can only offer one of two stress positions, or, in some cases, alternate stress positions (without any percentage of distribution between the two positions, though, perhaps, with stylistic labels of the kind ‘non-standard’, ‘archaic’ etc.). Lexicological sources, therefore, are unable to offer quantitative data of the kind required by a study of this kind, as well as being generally conservative and lagging behind the true picture of actual stress usage.

As mentioned above, the general assumption is that verbs in *-уютъ* associated with stress on their final syllable as a result of their semantic category which have a relatively higher frequency (and *relatively* is the key word here, since we are basically not concerned with absolute frequency values, at least for the time being – below it is postulated that there might be a critical amount of frequency beyond which it becomes statistically meaningless) are more likely to be able to maintain this stress position in the face of a general tendency towards initial suffixal stress. The question remains then of how accurately this is borne out statistically.

On the basis of the data presented in Table 3, the following statistical tests were carried out:

Test 1: Hypothesis test

Firstly, a hypothesis test was performed on words divided into either ‘common’ or ‘uncommon’ depending on whether they obtained a strictly positive or null score in RFL or not respectively using a null hypothesis. In hypothesis testing we assume a null hypothesis, and then use the data to see if we are justified in rejecting it. In our example we assume that ST1 is the same as ST2, and, set up a statistical test to measure how likely the observed data is given that assumption. In this case, we reject the null hypothesis, because our p-value is so small.

Result: rejected with a 1% confidence level and a p-value of .000002.

Interpretation: this is an extremely strong result for concluding that there is a difference associated with the relative frequency of a word in the language and the stress pattern observed. We would expect to be wrong only two out of one

million times (if, for example, this pattern had arisen from chance because of the sample size).

Next, two linear models were carried out to quantify the relationship between frequency and stress patterns, one with the verb *маскировать*, one without: the verb *маскировать* is, with its frequency count of 336, a statistical outlier and unduly affects the results obtained (based on lying well outside Cook's distance). This does not affect the inference, as we can still fit a plausible model and a strong correlation is still observed, though the magnitude of this correlation is reduced by a factor of about three.

Test 2: Model with verb *маскировать*

```
Call:      lm (formula=per ~ RFL)

Residuals:   Min       1Q   Median       3Q      Max
              -50.4082  -8.5110   -0.1833   21.4834   30.1009

Coefficients: Estimate Std. Error t value Pr (>|t|)
Intercept    73.5166    26.40309   11.481  3.89e-09 ***
RFL          -0.26870    0.07123   -3.772  0.00167 **

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The intercept gives the percentage occurrence of ST1 for a word that occurs with zero frequency. This means that if a word does not occur in RFL (i.e. is 'uncommon'), then we would expect approximately 74% of native speakers to use ST1.
- The RFL gives us the slope of the line at -0.26% . Thus we could say that for every time a word occurs 4 times in the word bank, its ST1 usage would decrease by 1%; 40 times, 10% less ST1 use; 120 times, 30% less ST1 use.
- The p-value obtained here is also very good, representing just a 0.2% chance of there being no correlation between frequency and ST1 (i.e. we are 99.8% confident that there is a non-zero correlation).

Test 3: Model without verb *маскировать*

Call:	lm (formula=pertrunc ~ RFLtrunc)				
Residuals:	Min	1Q	Median	3Q	Max
	-28.70	-12.81	-2.81	17.19	21.17
Coefficients:	Estimate	Std. Error	t value	Pr (> t)	
<i>Intercept</i>	82.8100	4.9441	16.749	4.05e-11	***
<i>RFLtrunc</i>	-0.7031	0.1130	-6.222	1.63e-05	***
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 '' 1

- The intercept, as for the previous model, gives the usage of ST1 for ‘uncommon’ words. In this case it is higher at about 83%; this means if a word does not occur in RFL (i.e. is ‘uncommon’) we would expect 83% of native speakers to use ST1.
- The RFLtrunc gives us the slope at -0.7% , so that we would say as a rule of thumb that we would expect the frequency usage of ST1 to decrease by 2% for every three times the word occurred in the RFL; 30 times, 20% less usage; 90 times, 60% less usage. This is a substantially larger correlation factor than in the previous model.
- The p-value is also an improvement: it indicates that over this range (RFL from 0 to approximately 90) there is a 0.0016% chance of there being no correlation, and is an extremely strong result.

Conclusion

In summary, there is very strong statistical evidence to support the hypothesis that there is a negative correlation between RFL and ST1 usage; in other words, for the threshold examined, the more common a word is, the less common ST1 usage is (or the more common ST2 usage is). Because of the metric used to measure ST1 usage (i.e. percentage) and the results from the second correlation test, the most likely structure of the data would be a linear region from 0 RFL to about 90 RFL, with a slope as given previously, followed by a flatter region after this: after a certain point, it seems, frequency becomes largely irrelevant.

The findings to come out of the above analysis are significant, notwithstanding the limited data that we are using, and represent the first attempt at a systematic

study of the relationship between frequency and stress types based on actual statistics. The approach taken may also serve as a possible methodology for the future in certain fluid areas of language where variation exists for no apparent reason. Even where the relative frequency may not be the original cause of the variation, it may be able to offer valuable insights into future directions and thereby explain anomalies. In this paper we have seen that in one area of Russian, namely the stress of verbs containing the suffix *-ировать*, higher frequency is certainly linked to a conservative, essentially redundant and anomalous type of stress (ST2), its survival given more or less longevity, albeit of a temporary nature, by the degree of familiarity of speakers with words which are characterised by it.

References

- Andersen, H.: 1973, 'Abductive and deductive change', *Language*, 49/4, 765–793.
- Kiparsky, V.: 1962, *Der Wortakzent der russischen Schriftsprache*, Heidelberg: Carl Winter Universitätsverlag.
- Lagerberg, R.: 2003, 'Татуировать or татуировать? Towards a comprehensive account of the stress of Russian verbs containing the suffix *-ировать*', *Russian Linguistics*, 27, 349–362.
- Lagerberg, R.: 2005, 'Towards a comprehensive account of the stress of Russian verbs containing the suffix *-ировать*: a survey of Russian speakers', *Russian Linguistics*, 29, 39–47.
- Lagerberg, R.: 2006, 'Towards a comprehensive account of the stress of Russian verbs containing the suffix *-ировать*: a survey of Russian speakers in Melbourne', *Australian Slavonic and East European Studies*, 20/1–2, 195–201.
- Lagerberg, R.: 2007, 'Variation and Frequency in Russian Word Stress', *Australian Slavonic and East European Studies*, 21/1–2, 165–176.
- Sharoff, S.: 2005, 'Methods and tools for development of the Russian Reference Corpus', in D. Archer, A. Wilson, P. Rayson (eds), *Corpus Linguistics Around the World*, Amsterdam, Rodolpi.

Zaliznjak 1985 = Зализняк, А. А.: 1985, *От праславянской акцентуации к русской*, Москва: Наука.

Zaliznjak 1977 = Зализняк, А. А.: 1977, *Грамматический словарь русского языка*, Москва: Русский язык.

Zasorina 1977 = Засорина, Л. Н.: 1977, *Частотный словарь русского языка*, Москва: Русский язык.