Phonology (2026), **Vol. 43**:e19 1–28
doi:10.1017/S0952675726100426

PHONOLOGY

ARTICLE

Rhythmic optimisation in the Russian comparative: a corpus study

Anton Kukhto[1] and Alexander Piperski[2]

[1]Massachusetts Institute of Technology, USA
[2]Department of Slavic and Baltic Studies, Finnish, Dutch and German, Stockholm University, Sweden

Corresponding author: Alexander Piperski; Email: alexander.piperski@slav.su.se

Received: 5 December 2022 UTC; **Revised:** 17 April 2025 UTC; **Accepted:** 3 May 2025 UTC

Keywords: rhythm; Russian; comparative; corpus phonology; stress

Abstract

This article addresses the effect of phrasal rhythm on synthetic adjectival comparative forms in Russian that exhibit optional truncation of the final vowel, such as *sil′née ~ sil′néj sloná* 'stronger than an elephant'. In a corpus study of such comparative forms, using a generalised mixed-effects model to analyse these data, we show that phrasal rhythm affects the distribution of truncation. It is dispreferred in contexts like *sil′née zébry* 'stronger than a zebra', where it would yield two adjacent stressed syllables (*sil′néj zébry*). This effect is categorical, and the study does not identify any gradient effect of distance on truncation. Stress placement in the comparative form itself also affects the distribution of *-ee vs. -ej*: comparative forms with stress on the suffix exhibit a lower proportion of truncation. These findings show that Russian exhibits rhythmic optimisation at the phrase level.

Contents

1. Introduction **2**
2. Comparatives in *-ee/ej* **3**
2.1. Overview of Russian comparative morphology . . . 3
2.2. Stress in comparative forms . . . 3
2.3. Phonetics of *-ee/ej* . . . 5
2.4. Distribution of *-ee vs. -ej* . . . 7
3. Hypothesis **7**
3.1. Rhythmic alternation . . . 7
3.2. Phonological representation of rhythm . . . 8
4. Data **9**
4.1. Data collection procedure . . . 9
4.2. Descriptive statistics . . . 11
5. Analysis **14**
6. Expanding the data set **18**
6.1. Clausal standards of comparison . . . 18
6.2. Spoken corpus . . . 20
7. Discussion **21**
8. Conclusions **25**
A. Individual adjectives included in the analysis **25**
References **27**

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

1. Introduction

It was proposed as early as in Sweet (1887: 92) that linguistic rhythm, manifesting itself as an alternation between stronger and weaker linguistic units, plays a role in grammar; see Schlüter (2005: 17–42) for a summary. In research on prosody, this tendency is known as the Principle of Rhythmic Alternation (Schlüter 2015: 179):

(1) *Principle of Rhythmic Alternation*
There is a fundamental tendency in the organisation of speech encouraging the alternation of strong and weak beats.

The examples in (2) from English use grid representations (Liberman & Prince 1977; Prince 1983) to illustrate rhythmic alternation across and within words, respectively.

(2) a.
3							×	
2	×						×	
1	×			×			×	
0	×		×	×	×		×	
	a	c	h	r	o	m	a	t i c l e n s

b.
3								×		
2	×							×		
1	×			×				×		
0	×		×	×	×		×	×	×	
	A	p	a	l	a	c	h	i	c	o l a

Both (2a) and (2b) exhibit a perfect binary alternation between stressed (σ́) and unstressed (σ) syllables, as indicated by the presence or absence of grid marks above level 0, respectively.

Since the late 19th century, various effects of rhythm on grammar have been documented and analysed, most prominently in word stress assignment (Hayes 1995), but also in other corners of phonology and beyond; see recent examples in Hammond (2016), Shih (2017b), Schlüter & Knappe (2018), Breiss & Hayes (2020) and the references therein. This growing body of literature has been to a large extent made possible by advances in linguistic corpus technology. Most of the studies cited above use linguistic corpora as their main source of data. The present contribution is no exception.

In this article, we present a study of one morpheme of the Russian language (Slavic), the comparative suffix *-ee.* We show that the variation between the disyllabic (*-ee*) and monosyllabic (*-ej*) shapes of this suffix is influenced by phrasal rhythm, which can be represented by interacting phonological constraints. This case is a new contribution to the growing body of literature on rhythm sensitivity in Russian.

Stress assignment in Russian is, by and large, determined by the interplay of accentual properties of morphemes and paradigmatic factors, and most words have only one stressed syllable (see §2.2). While it may seem that Russian would therefore be unlikely to exhibit effects of rhythm, recent literature indicates that it does. Gouskova & Roon (2013) find that the presence of a secondary stress in Russian compounds is conditioned by the distance between the primary stress and the potential location of the secondary stress. Kukhto & Piperski (2020) find that verbal forms with variable stress placement tend to select the stress that conforms to the Principle of Rhythmic Alternation in (1) depending on the phrasal context. These findings invite the following question: what is the common core of the grammar that underlies these different phenomena sensitive to rhythm? The goal of the present study is to inch towards the answer to this question. Looking beyond Russian, this study connects to the broader literature investigating the universal grammar of rhythm.

§2 introduces the comparative suffix *-ee*/*ej,* which is the focus of this article: the basics of comparative morphology in Russian are laid out first (§2.1), followed by stress placement in Russian adjectives (§2.2), and the variation between forms in *-ee* and in *-ej* (§2.3). §3 presents the general hypothesis that we put to test, namely that rhythm influences the distribution of *-ee* and *-ej* (§3.1), and its specific instantiations in terms of phonological constraints (§3.2). §4 describes the corpus and our data collection procedure (§4.1) and provides a summary of the data (§4.2). §5 presents the results of mixed-effects logistic regression modelling, with adjective as a random intercept and phonological/prosodic variables as fixed effects; forward model selection identified two significant predictors – rhythmic

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

alternation and stress in the comparative. In §6, we investigate two additional data sets: written (§6.1) and spoken (§6.2). §7 discusses the interpretation of these findings. Finally, §8 concludes the article.

2. Comparatives in *-ee/ej*

2.1. Overview of Russian comparative morphology

Russian adjectives have two strategies of forming the comparative degree: synthetic, as in (3b), and analytic, as in (3c).[1] Russian adverbs exhibit the same two strategies of comparative formation, and their synthetic comparative forms are morphologically identical to adjectival forms. In this article, we focus on synthetic forms of adjectival comparatives, setting adverbs aside.[2]

(3) a. *èlegant-n-yj* 'elegant' [elegant-ADJ-SG.M.NOM]
b. *èlegant-n-ee/ej* 'more elegant' [elegant-ADJ-CMPR]
c. *bolee èlegant-n-yj* 'more elegant' [more elegant-ADJ-SG.M.NOM]

Three allomorphs of the comparative suffix are traditionally singled out in synthetic forms: *-ee/ej, -e* and *-še* (Garde 1980; Es′kova 2011). Generally, an adjective forms its synthetic comparative with only one of these three allomorphs; only a few adjectives exhibit variation between *-ee/ej* and *-e*; see Table 1. To the exclusion of four adjectives, all synthetic comparatives are indeclinable (the exceptions being *xorošij* 'good' – *lučšij* 'better', *ploxoj* 'bad' – *xudšij* 'worse', *bol′šoj* 'big' – *bol′šij* 'bigger', *malen′kij* 'small' – *men′šij* 'smaller').

The formation of comparatives with *-ee/ej* by semantically compatible adjectives is generally unimpeded, except for adjectives with certain derivational suffixes which are 'not entirely free' to form comparatives (Es′kova 2011: 136–138). These are deverbal adjectives in *-l* (e.g., *uny-l-yj* 'dreary', *usta-l-yj* 'tired') and adjectivised participles in *-šč, -t* and *-em* (e.g., *blest-jašč-ij* 'glittery', *otkry-t-yj* 'open', *uvaža-em-yj* 'respected'). The suffix *-ee/ej* palatalises the final consonant of the stem it attaches to, as in *bel-yj* [ˈbʲelə̯ɨ] 'white-M.SG.NOM' – *bel-ee* [bʲɪˈlʲeə] 'white-CMPR'. (We return to Russian vowel reduction in §2.3.)

The allomorph *-ee/ej* accounts for the majority of synthetic comparative forms: of the 17,994 adjective stems marked as having non-restricted synthetic comparative in Zalizniak's (1987) grammatical dictionary, 17,857 stems (99.2%) take *-ee/ej*. The only class of stems that productively takes *-e* is velar-final; *-še* is restricted to a handful of stems (Itkin 2007: 201–202).

2.2. Stress in comparative forms

Russian exhibits a system of morphologically governed lexical stress assignment (Melvold 1989; Alderete 1999, among others): a word is assigned stress based primarily on the accentual specifications of its constituent morphemes. Most words only allow one stress to surface; secondary stresses occur in

Table 1. *Allomorphs of the comparative.*

Positive (M.SG.NOM)	Comparative	Translation
bel-yj	*bel-ee/ej*	'white'
dorog-oj	*dorož-e*	'dear, expensive'
lov-k-ij	*lov-č-e ~ lov-č-ee/ej*	'deft, nimble'
dolg-ij	*dol′-še*	'long' (of time)

[1] We use the traditional scholarly transliteration of Cyrillic in Russian examples throughout the text, accompanied by IPA transcriptions where necessary. Stresses are marked with acute accents when relevant.

[2] Some roots, like *èlegant-* in (3), require adjectivising suffixes; see Kukhto (2024) for a discussion.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 2. *Stress in positive and comparative forms of Russian adjectives.*

	Positive				
	Long	Short			
Pattern	F.SG.NOM	F.SG	PL	Comparative	Gloss
a/*a*	*rádost-n-aja*	*rádost-n-a*	*rádost-n-y*	*rádost-n-ee*/*ej*	'jolly'
a/*b*	*úm-n-aja*	*um-n-á*	*um-n-ý*	*um-n-ée*/*éj*	'clever'
a/*c*	*cél-aja*	*cel-á*	*cél-y*	*cel-ée*/*éj*	'whole'
b/*b*	*smeš-n-ája*	*smeš-n-á*	*smeš-n-ý*	*smeš-n-ée*/*éj*	'funny'
b/*c*	*živ-ája*	*živ-á*	*žív-y*	*živ-ée*/*éj*	'alive'

compounds (Gouskova & Roon 2013). In synthetic comparative forms, the suffixes *-e* and *-še* are never stressed. By contrast, the suffix *-ee*/*ej* can be either stressed or unstressed, depending on the accentual class of the stem it attaches to.

Russian adjectives have two subparadigms with different syntactic distributions and sets of desinences in the positive degree: the so-called long and short series (see Borik 2014 and the references therein). In the simplest case, stress placement in the two subparadigms of one adjective is identical modulo accentual specifications of individual desinences, see rows *a*/*a, b*/*b* and *b*/*c* of Table 2. For other adjectives, whether stress falls on the stem or on the desinence differs between long and short forms; see rows *a*/*b* and *a*/*c.* The first letter of these accentual class labels refers to the accentual class of an adjective's stem in long forms; the second letter, after the slash, refers to the accentual class of its short-form stem (see Zalizniak 1987), where *a* stands for accented stems, *b* for post-accenting and *c* for unaccented stems (for more on Russian adjectival stress, see Melvold 1989: 181 ff.).

The standard analysis maintains that stress assignment in synthetic comparatives follows stress assignment in feminine singular short positive forms (Zalizniak 1987: 35). Whenever the F.SG short positive form receives stem stress, so does the comparative; whenever the F.SG short positive form receives desinence stress, the suffix *-ée*/*éj* is stressed in the comparative. The stress of other forms, such as the long forms and the plural short form, may fall on different morphemes, which results in the five accentual classes of Russian adjectives shown in Table 2.

However, there exist adjectives with mismatches in stress placement between the F.SG short positive form and the comparative:

(4)

	Short F.SG	Comparative		
a.	*zdoróv-a*	*zdorov-ée*	'healthy'	(Zalizniak 1987)
b.	*vlást-n-a*	*vlast-n-ée*	'commanding'	(Es′kova 2011: 138)

Examples like (4) originated through restructuring within the system of adjectival stress, in particular in comparative forms. Following Es′kova (2011), we hypothesise that stress in comparatives in Modern Russian is assigned based on (a) length and morphological complexity of the stem and (b) paradigmatic correspondence to stress in the long positive form, along the lines of (5).

(5) In a synthetic comparative form,
   a. if the stem is monosyllabic or monomorphemic, stress the suffix;
   b. otherwise, copy the stress of the long positive form.

Monosyllabic adjectives with stem stress in the comparative (like *zrím-ee* 'more visible') and polysyllabic adjectives with suffix stress in the comparative (like *zelen-ée* 'greener') are vastly underrepresented (see §4.1); most adjectives either have monosyllabic/monomorphemic stems and stress the suffix in the comparative (like *um-n-ée* 'more clever' in Table 2) or have polysyllabic and morphologically complex stems and stress the stem itself (like *rádost-n-ee* 'jollier' in Table 2).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

The correlation between having a polysyllabic stem and stressing the stem is due to derivational morphology: most polysyllabic stems are morphologically complex, and many derivational suffixes introduce fixed stem stress (Melvold 1989). For monosyllabic stems, Es′kova (2011: 138) notes that mobile stress, characterised by desinence stress in short F.SG forms (like *cel-á* 'whole-F.SG' in Table 2) and suffix stress in the comparative (*cel-ée*), is typical. She hypothesises that their predominance results in the tendency to stress the comparative suffix with all monosyllabic stems, no matter their stress in the positive forms.

Mismatches between the short F.SG forms and the comparatives are tolerated in frequent adjectives such as *zdoróv-a* 'healthy', whereas for less frequent adjectives (such as *vnját-n-yj* 'intelligible'), synthetic comparatives are avoided because of the conflicting pressures of stressing the suffix and preserving the stress of the positive form (?*vnjat-n-ée* or ?*vnját-n-ee*),[3] hence analytic comparatives (*bolee vnját-n-yj*) are preferred (Es′kova 2011: 138–140). This state of affairs suggests a competition between faithfulness of stress in the comparative to stress in the positive forms (see Breiss *et al.* 2026 on the effects of frequency in output–output correspondence), and the tendency to assign stress based on the shape of the stem as per (5a).[4]

A full analysis of adjectival stress is tangential to our present goals, so we leave it for future research. What is important to us here is that most adjectives stress either the stem or the comparative suffix; variation exists (?*vnjat-n-ée* or ?*vnját-n-ee* above and examples in Es′kova 2011: 138–140), but adjectives that exhibit it were excluded from our data (see §4).

Crucially, the choice between stem and suffix stress is orthogonal to the choice between *-ee* and *-ej*: no matter whether an adjective stresses the stem or the suffix, the location of stress in its comparative form with *-ee* is the same as in the form with *-ej*.[5]

2.3. Phonetics of -ee/ej

Virtually any synthetic comparative form with *-ee*/*ej* can appear in either of the two shapes: the disyllabic *-ee* /eje/ or the monosyllabic *-ej* /ej/.[6] This variation has arisen as a result of truncation of word-final vowels over the history of Russian (Zalizniak 1992). The *-ee*/*ej* variation is found both in speech and in writing, yielding examples like (6).

(6) *čërnaja čaša neba vsë* ***polnee*** *i* ***polnej*** *nalivaetsja mutnym, toskujuščim svetom*
'the black bowl of the sky is fuller and fuller filled with a muddy, yearning light'
(Maxim Gorky, *Tales of Italy,* 1911–1913; the translation is ours)

The phonological similarity between the two shapes, *-ee* /eje/ and *-ej* /ej/, invites the question of whether there is a clear distinction between the two. Complicating matters, on the one hand, we have their phonetic similarity. In *-ee* /eje/, the final vowel is always unstressed and reduced to [ə] (or [ɘ];

[3]The authors of this article, both native speakers of Russian, do not agree on how this form is stressed. Curiously, the older author prefers the innovative form *vnjat-n-ée,* whereas the younger author prefers the conservative *vnját-n-ee.*

[4]Faithfulness of stress to positive forms essentially means faithfulness to long forms. First, long positive forms are by far the most frequent. In the manually disambiguated subcorpus of the Russian National Corpus (ruscorpora.ru, accessed 17 October 2023; see Plungjan *et al.* 2005 for a general description of the corpus), they account for 91% of all adjective tokens (465,081 out of 511,001). This proportion is due to the fact that long positive forms are the ones used attributively and in analytic comparatives, while short positive forms can only be used predicatively, and synthetic comparatives and superlatives are not so common either. Second, stress in short forms either coincides with the long forms (class *a*/*a*) or points in the same direction as stressing the comparative suffix (*a*/*b, a*/*c*), or both (*b*/*b, b*/*c*).

[5]It is conceivable that adjectives which exhibit stress variation could have different shares of forms in *-ee vs. -ej* in correlation with stress placement, for example, *vnjátn-ee* > *vnjátn-ej,* but *vnjatn-ée* < *vnjatn-éj.* However, such forms are rare and avoided, so testing this effect would be difficult.

[6]Es′kova (2011: 136) identifies five adverbial forms that only appear with the disyllabic *-ee*: *bol-ee* 'more', *men-ee* 'less', *dal-ee* 'further', *dol-ee* 'longer', *ran-ee* 'earlier'. Note that *bol-ee* 'more' and *men-ee* 'less' are used to form analytic comparatives as in (3c). Es′kova suggests that these forms are fossilised and that the suffix *-ee* in them might not be identified with the general comparative *-ee*/*ej.*

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 3. *Phonetic realisations of the* -ee/ej *comparative suffix.*

	Stressed		Unstressed	
Context	Disyllabic (*-ée*)	Monosyl. (*-éj*)	Disyllabic (*-ee*)	Monosyl. (*-ej*)
Emphatic	eje, eje	ej, ej	eje, eje	ej, ej
Phrase-medial	eə, eï	eɪ̯	ïə, ïï	ïɪ̯
Phrase-final	eə	eç, eɪ̯	ïə	ïç, ïɪ̯

the height of this vowel can vary); see Iosad (2012) on vowel reduction in Russian. Regular palatal glide deletion applies (Es′kova 2011: 30), yielding *-ée* [eə] if the suffix is stressed or *-ee* [ïə] if the suffix is unstressed; we follow Iosad (2012) in using [ï] to transcribe a raised and fronted schwa in the context of a palatalised consonsonant. (Recall that *-ee*/*ej* palatalises the final consonant of the stem it attaches to.) In *-ej,* the word-final glide is most commonly realised as a semivowel [ɪ̯] (Es′kova 2011: 30; Yanushevskaya & Bunčić 2015), yielding *-éj* [eɪ̯] if the suffix is stressed and *-ej* [ïɪ̯] if the suffix is unstressed. Clearly, both stressed *-ée* [eə] and *-éj* [eɪ̯] and unstressed *-ee* [ïə] and *-ej* [ïɪ̯] are alike, making it difficult to distinguish between them in running speech, be it by ear or using the tools of acoustic analysis. A reviewer points out that additional complications arise from coarticulation with linearly following material in the absence of a prosodic break. In particular, when *-ee*/*ej* is followed by a jV-initial word, as in *veselée jármarki* 'merrier than a market', *-ee* and *-ej* may be conflated to such a degree that it is impossible to draw a line between them. We return to this issue in §7.

On the other hand, however, the two shapes (i) admit pronunciations that set them apart more clearly and (ii) diverge in contexts where syllable count is at stake. First, if every syllable in a comparative form with *-ee*/*ej* is pronounced emphatically, the glide is restored in *-ee* and realised as either [j] or even the fricative [ʝ] in both, that is, *-ee* [eje] or [eʝe] and *-ej* [ej] or [eʝ]. Second, a phrase-final /j/ can be devoiced (Bolla 1981; Yanushevskaya & Bunčić 2015), resulting in *-éj* [eç] or unstressed *-ej* [ïç] phrase-finally, pronunciations not available to *-ee.* Table 3 summarises the phonetic realisations that *-ee/ej* yields.

Another piece of evidence for the distinction between *-ee* and *-ej* comes from verse. Russian poetry has been predominantly accentual-syllabic since the mid-18th century (Gasparov 1996), relying formally on both ictus (strong beat) placement and syllabic distances in between strong beats. In accentual-syllabic verse, *-ee* and *-ej* are clearly distinguished. Take the excerpt in (7); we use the acute accent to indicate strong beats here, which coincide with word stresses, except for the absence of ictus on *každyj* [ˈkazʑdəɪ̯] 'every' and *èti* [ˈeʲtʲï] 'these'.

(7) *Nikolai Gumilev, 'The Sahara', 1921; the translation is ours*

Tak i búdut brodít' do skončán'ja vekóv	'And so they will roam till the end of time,
Každyj čás vsë ***groznéj*** *i* ***groznée****,*	with every hour more and more baleful,
Golovój propadája sredí oblakóv	with their heads lost among the clouds,
Èti strášnye sérye zméi.	these fearsome grey snakes'.

In this anapaestic quatrain, the strong beats (and the corresponding stresses) in *groznéj i groznée* 'more and more baleful' are separated by two unstressed syllables: [grɐ.ˈzʲnʲe.(i)i.grɐ.ˈzʲnʲe.ə]. In the first instance, *groznéj,* only the monosyllabic *-ej* is possible; the extra syllable of *-ee* would violate the metre. The second instance, *groznée,* rhymes with *zméi* [ˈzmʲe.ï] 'snake.NOM.PL' in the last line. The monosyllabic *-ej* is impossible here, as it would violate both the ABAB rhyme scheme deployed throughout the poem and the specific rhyme with *zméi* [ˈzmʲe.ï]; the distinct form *zmej* [ˈzmʲeɪ̯] 'snake.GEN.PL', rhyming with a potential *groznéj* [grɐ.ˈzʲnʲeɪ̯], would be ungrammatical in the last line. This brief example illustrates how the disyllabic and monosyllabic shapes of *-ee*/*ej* contrast when syllable count is of the essence.

To reiterate, some instances of *-ee*/*ej* in casual speech may be difficult to classify as *-ee* or *-ej* unambiguously. We nevertheless conclude from this discussion that *-ee* and *-ej* are distinct variants of the allomorph at hand and that speakers of Russian can have intuitions as to whether one or the other

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

variant is used. Those intuitions can be manifested in writing, resulting in the choice of either *-ej* ⟨ей⟩ or *-ee* ⟨ее⟩. Henceforth, we will refer to forms with *-ej* as TRUNCATED (or monosyllabic), and forms with *-ee* as NON-TRUNCATED (or disyllabic). We assume in what follows that *-ee* /eje/ is the underlying representation of this allomorph of the comparative suffix, and that its monosyllabic variant is derived by apocope, in line with both the historical development of comparative forms (see above) and their frequency (see §4.2).

2.4. Distribution of -ee vs. -ej

Given that any adjective that forms its comparative with *-ee* may appear as ending in either *-ee* or *-ej* (bar the five adverbial forms mentioned in fn. 6), the question arises of what factors affect their distribution. This question has not received a lot of attention. We are not aware of any discussion of inter-speaker variation, that is, whether different (groups of) speakers exhibit different distributions of the truncated *vs.* non-truncated variants.

More attention has been devoted to intra-speaker variation. Most prominently, it has been suggested that the monosyllabic variant is more colloquial compared to the disyllabic one, especially in writing (Isačenko 1965: 210; Es′kova 2011: 136; cf. Chervonnaya 2017). In addition to that, Chervonnaya (2017) finds that the share of *-ej* is different in adverbial *vs.* connective uses of *točnee* 'more precisely; rather', and that attenuated comparatives with the prefix *po-* (Boguslavsky & Iomdin 2009) exhibit a higher share of *-ej* than their counterparts without *po-*; she examines a couple of other lexical factors, but does not find that they impact the choice of *-ee vs. -ej*.

Here, we focus on how rhythm affects the distribution of the truncated and non-truncated variants of *-ee.* To the best of our knowledge, the effect of rhythm has not been addressed in prior literature. The only phonological factor to have received some attention before is stress placement in the adjective. Es′kova (2011: 136) remarks that the ratios of stressed *-ée vs. -éj* and unstressed *-ee vs. -ej* might be different, even though she does not specify what that difference could be – an observation that our findings confirm (see §5). In the following section, we make explicit the rhythmic and stress-related factors that we put to the test.

3. Hypothesis

3.1. Rhythmic alternation

The central question we ask in this article is whether the variation between the disyllabic shape of the comparative suffix *-ee* and its monosyllabic, truncated shape *-ej* is influenced by rhythm. Recall that Russian adheres to the principle of culminativity rather strictly, allowing secondary stresses to surface only in compounds. Yet even though rhythm does not play a major role in Russian stress assignment (for discussion of its role in compounds, see §7), it is not mute in the phonology of Russian, and its effects can be seen under the right circumstances. In what follows, we show that the *-ee*/*ej* variation is exactly one of such cases. To reveal the influence of rhythm in it, we turn to the alternation of stresses within phonological units larger than individual words.

In the most general form, the hypothesis that we test here can be stated as follows: all else being equal, the variation between the disyllabic *-ee* and the monosyllabic *-ej* ensures rhythmic alternation (as per (1)), in the following manner. The comparative form itself and the linearly following word are assigned stresses based on their morphological composition. Given the stress placement and the words' lengths, the distance between the stresses will vary. If the stresses are too close to each other, the disyllabic *-ee* will be preferred to space them apart, and *vice versa,* if the stresses are too far apart, the monosyllabic *-ej* will be preferred to reduce that distance. This is illustrated in (8) using the comparative form *bystree* 'faster'.

(8) a. *bystr-ée zébry* > *bystr-éj zébry* 'faster than a zebra'
b. *bystr-ée sloná* < *bystr-éj sloná* 'faster than an elephant'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

In (8a), truncation yields a configuration where two stresses are adjacent (*bystr-éj zébry*), disrupting rhythmic alternation, whereas in (8b) truncation helps create perfect binary alternation between stressed and unstressed syllables (*bystr-éj sloná*). Therefore, by hypothesis, a configuration like (8b) will exhibit truncation to a greater extent than (8a).

The above is a general outline of how rhythmic alternation is expected to affect the variation in question. There are various approaches to analysing rhythmic effects in the current literature, which yield different particular predictions. In the following section, we provide a brief overview of those approaches and make explicit the particular effects we are testing.

3.2. Phonological representation of rhythm

Since at least Liberman & Prince (1977), rhythm has been analysed as part of the phonological grammar. Proposals differ in how rhythm should be represented and computed; in other words, what it means for a string to be 'rhythmic'.

The common ground that most approaches to rhythm share is that the alternation applies to abstract phonological units: syllables, or sometimes morae (see Stanton 2019, 2021 for an alternative approach where rhythm is computed based on normalised duration, as well as discussion in §7). Two general classes of mechanisms that underlie the alternation of more and less prominent syllables (or morae) can be singled out: (i) those that parse them into larger prosodic units – feet, which are required to have their own internal structure and to be spaced apart (McCarthy & Prince 1993, among others) – and (ii) those that directly regulate distances between syllables, often represented as units of the metrical grid (Gordon 2002, among others). Here, we adopt the latter type of approach and investigate it using the formalism of constraint-based phonology.

For foot-free approaches, the basic constraints that regulate rhythm are *Clash (Prince 1983) and *Lapse (Nespor & Vogel 1989), defined in (9).

(9) a. *Clash: Assign a violation (∗) for each sequence of two adjacent stressed syllables.
b. *Lapse: Assign a violation (∗) for each sequence of two adjacent stressless syllables.

These constraints ensure binary alternation. Ternary alternation is accounted for by *ExtendedLapse (Gordon 2002: 502; for a foot-based approach, see Martínez-Paricio & Kager 2015).

(10) *ExtendedLapse: Assign a violation (∗) for each stressless syllable both preceded and followed by a stressless syllable (i.e., for each sequence of three adjacent stressless syllables).

Note that constraints of the *Lapse family incorporate a notion of distance between stresses in a way that *Clash does not: *Clash does not distinguish between the sequences σσσσ́ and σσσσσ́, whereas *Lapse and *ExtendedLapse do (1 *vs.* 2 violations and 0 *vs.* 1 violation, respectively); see Stanton (2021) for discussion of whether linear distance between prominences plays a role in rhythmic grammar, and if so, how distance should be defined. Versions of *Clash that are sensitive to distance have been proposed as well. To capture the finding that in Russian compounds, the further a potential location of secondary stress is from the syllable carrying primary stress, the more likely it is that the secondary stress will surface, Gouskova & Roon (2013: 425) introduce *Clash+, a positive gradient constraint, defined in (11).

(11) *Clash+: Assign a reward for every unstressed syllable that separates two stresses.

In a numerical OT framework, such as Harmonic Grammar (HG) or Maximum Entropy (MaxEnt) Grammar, constraints can assign both penalties and rewards, enabling gradient evaluation of candidate outputs. This numerical approach underpins the use of *Clash+ to account for distance effects in stress assignment. Part of the motivation behind introducing *Clash+ rather than accounting for the distance effect with a constraint of the *Lapse family is that Russian is tolerant of stress lapses elsewhere; we return to the discussion of gradience in §7.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

The constraints in the *Clash and *Lapse families evaluate adjacent prominence peaks and prominence troughs separately. Therefore, it is possible that they may be ranked differently with respect to a conflicting constraint, for example, *Clash ≫F ≫*Lapse. As an alternative, Shih *et al.* (2015; see also Shih 2017b) introduce the constraint EurhythmyDistance, which evaluates deviations from perfect rhythmic alternation cumulatively (12).

(12) EurhythmyDistance: Given two stressed syllables, assign $|n - 1|$ violation marks (∗), where n is the number of stressless syllables separating the two.

Thus, if two stressed syllables are adjacent (i.e., $n = 0$), EurhythmyDistance assigns $|0 - 1| = 1$ violation mark; if they are separated by one stressless syllable ($n = 1$), it assigns no violations; two intervening syllables incur one violation ($|2 - 1| = 1$), and so forth. In other words, for any inter-stress interval, EurhythmyDistance assigns the sum of the violations of *Clash and *Lapse which that interval would incur.

Our task is to model the relative frequencies of the monosyllabic and disyllabic shapes of the suffix *-ee*/*ej* in a corpus, which necessitates a stochastic constraint-based approach. The constraints in (9)–(12) form the main set of factors whose effects we will be testing in the rest of this article. In order to do so, we turn to MaxEnt grammar (see Hayes & Wilson 2008 and the references therein, and see Flemming 2021 for a comparison between MaxEnt and Noisy HG). The probability of each candidate is computed as an exponentiated weighted sum of its constraint violations and rewards. We implement the MaxEnt model in the more general form of a mixed-effects logistic regression, the equivalence between the two having been highlighted by Shih (2017a). This makes it possible to accommodate the variability inherent in our corpus data and handle fixed and random effects more easily than in classical implementations of MaxEnt grammar.

4. Data

4.1. Data collection procedure

In order to model the variation between the disyllabic *-ee* and the monosyllabic, truncated *-ej,* we rely primarily on data from Araneum Russicum Maius (ARM), a written corpus of one billion words comprised of texts collected from the web (Benko 2014; accessed 1 November 2022). We chose ARM over other available written corpora of comparable size, such as ruTenTen and the General Internet Corpus of Russian, because it is the only corpus which allows downloading complete concordances. These data are discussed in §4.2 and analysed in §5.

Our secondary source of data is the multimedia subcorpus of the Russian National Corpus (RNC), the largest openly available spoken corpus for Russian at 5.7 million words. It provides audio (and video) tracks in addition to semi-automatic orthographic transcription of each speech sample; samples come predominantly from Russian-language cinema, followed by recordings of public speaking, theatre productions and audiobooks, as well as informal conversation. The data from RNC are discussed in §6.2.

The rationale for focussing on written data is twofold. First, ARM is much larger than the multimedia subcorpus of RNC. Second, both written and spoken data come with their own trade-offs. On the one hand, because the truncated *-ej* is more colloquial and potentially regarded by speakers as non-standard in writing (see §2.4), its share may be depressed in written compared to spoken data, making it more difficult to detect any patterns in the distribution. On the other hand, we saw in §2.3 that, although the set of potential surface forms of *-ee*/*ej* is clear, it may not be possible to classify some instances of it as unambiguously mono- or disyllabic, especially before *j*V- or V-initial words, introducing noise into the analysis. As we further discuss in §7, this might suggest gradience in the surface realisation of *-ee*/*ej,* which is inevitably lost in writing, where we find either *-ee* ⟨ее⟩ or *-ej* ⟨ей⟩. However, if a phonological generalisation emerges from the written data, we maintain that it must be reflective of a phonological process observed in spoken data as well. In other words, while we may expect writing to obscure some

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 4. *Number of adjective types in the data by stem shape.*

	Stem shape	
	Monosyllabic	Polysyllabic
Stress on suffix (*-ée*/*éj*)	30	5
Stress on stem (*-ee*/*ej*)	0	30

phonological generalisations, we do not expect it to engender phonological generalisations not found in speaking. Even if we did find such a case, we should still be able to account for it. With these considerations in mind, we move on to describe our data collection procedure, which went as follows.

First, we extracted a list of the most frequent comparative forms in *-ej,* the less frequent of the two variants. We then annotated the list for position of stress based on Zalizniak (1987), confirmed by the authors' intuitions, and extracted the most frequent comparatives in *-ej,* with an arbitrary cap of 30 adjectives in each type: mono- *vs.* polysyllabic and suffix *vs.* stem stress. The types and the resulting numbers of adjective stems in each are shown in Table 4.

As Table 4 shows, comparatives with monosyllabic stems and stem stress (e.g., *zrím-yj* 'visible' – *zrím-ee* or *zrím-ej*) as well as comparatives with polysyllabic stems and suffix stress (e.g., *zdoróv-yj* 'healthy' – *zdorov-ée* or *zdorov-éj*) fell short of the cut-off point. That is, no adjective like *zrím-yj* 'visible' occurred in the comparative form with truncation (*zrím-ej*) once; only five adjectives like *zdoróv-yj* 'healthy' occurred in the comparative with truncation (*zdorov-éj*). Recall that in §2.2, we hypothesised (following Es′kova 2011) that this state of affairs is due to the emerging tendency to stress the comparative suffix with short and simplex stems and the stem itself with long and complex stems. On top of that, the comparatives *skorej, vernej, krasivej* and *gorjačej* were excluded because of stress variation (*krasívej* or *krasivéj* 'more beautiful'), abundance of homophonous forms (*gorjačej* 'hot.GEN/DAT/INS/LOC.SG.F' as well as *gorjačee* 'hot.NOM/ACC.SG.N') or predominance of set phrases (*skoree vsego* 'most likely') or adverbial uses (*vernee* 'rather'), and were replaced with four comparatives below the initial cut-off point. The resulting list of adjectives is shown in Table A1 in the Appendix.

For all comparatives in the resulting list, we extracted adjectival occurrences of their truncated and non-truncated form ($65 \times 2 = 130$ form types in total: *bystr-ée* and *bystr-éj* 'quicker', *složn-ée* and *složn-éj* 'more difficult', etc.) followed by a noun in the genitive case, optionally preceded by an adjective in the genitive case.[7] This procedure yielded a total of 19,208 examples. The standards of comparison (SoC) were then automatically annotated for stress based on Zalizniak (1987) and Wiktionary (wiktionary.org). Additionally, every SoC was annotated for the initial segment of the word linearly following the comparative form: whether it is a vowel or not (*j*V-initial words were classified in the latter group); recall from §2.3 that realisations of *-ee* end in a vowel, which creates a V#V sequence when the following word is V-initial, which could affect the distribution of *-ee vs. -ej*.

The ARM corpus is disambiguated and part-of-speech-tagged using TreeTagger (Schmid 1999), yet the resulting annotation contains errors, so all collected examples had to be checked manually. We excluded instances where the comparative was used adverbially or was not actually followed by a noun phrase in the genitive expressing the SoC. We also excluded all metrified examples since, as demonstrated in §2.4, rhythmic alternation is guaranteed in accentual-syllabic verse, making such examples uninformative for our purposes. Finally, we excluded examples where the SoC exhibits stress variation, as reported in Zalizniak (1987) or according to the authors' intuitions. A note is due on stress in compounds: we consistently preserved the automatic annotation, which only located the primary stress in compounds, to avoid factoring in our intuitions as to whether in each individual case

[7] A reviewer points out that there is more variety in what may follow a comparative form: a pronoun (*bystrée menjá* 'faster than I am'), adjectivised pronouns modifying a noun (*bystrée ljubógo sloná* 'faster than any elephant'), possessive pronouns (*bystrée moegó sloná* 'faster than my elephant'), etc. We did not include such cases because of the need to manually check our corpus search results and the time requirements of that process.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

a compound does or does not have secondary stress on the first root of the compound. After all that, we were left with 16,362 examples.

Recall from §2.1 that Russian uses two strategies to express comparison: analytic (*bólee složn-yj* 'more difficult') and synthetic (*složn-ée* or *složn-éj*). On top of that, synthetic comparatives introduce the standard of comparison in either of two ways: phrasally, as an NP with the head noun in the genitive, or clausally, with the conjunction *čem* 'than'. It is possible that the competition among all these forms is evaluated by phonology: *bystrée*/*éj zébry* 'faster than a zebra', *bystrée*/*éj čem zébra, bólee býstryj čem zébra*; see, for instance, Shih (2017b) on the English genitive. In this study, we limit ourselves to synthetic comparatives. Grashchenkov & Lyutikova (2017) show that synthetic and analytic comparatives in Russian differ both syntactically and semantically. In a feed-forward model of language (Chomsky 1965: 16), such differences entail that synthetic and analytic forms should not compete in the phonological module – the choice is made prior to phonology. This may be a simplifying assumption (see Breiss & Hayes 2020 for evidence of phonological influences on syntactic choices), but we adopt it for now; we return to this discussion in §7.

As for phrasal *vs.* clausal SoC (*bystrée*/*éj zébry* 'faster than a zebra' *vs. bystrée*/*éj čem zébra*), we assume by the same token that, since the syntactic structures of the SoC differ across these two types (Grashchenkov & Lyutikova 2017), they are not in direct competition in phonology. However, we use cases with clausal SoC (*bystrée*/*éj čem zébra*) as a test case for the model developed for the data with phrasal SoC (*bystrée*/*éj zébry*). We therefore extracted 4,529 adjectival occurrences of all 65 comparatives followed by a comma (optionally), *čem* and a noun in the genitive case, optionally preceded by an adjective in the genitive case, the same as for cases with phrasal SoC, and annotated them using the same procedure as above. These data are analysed in §6.1.

Finally, by way of comparison, we collected comparative forms in the multimedia subcorpus of the RNC. For all 65 adjectives in our list, we extracted all tokens of truncated and non-truncated comparatives linearly preceding *any* phrasal SoC. This choice was motivated by the scarcity of comparative tokens in this corpus. After manual filtering, we were left with 193 examples. For each example, we transcribed the comparative suffix as either disyllabic *-ee* or monosyllabic *-ej,* based on our perception. We limited ourselves to this procedure because a full-fledged phonetic study would take us too far afield; on top of that, the recordings available in the corpus are of varying quality and may not facilitate a careful acoustic analysis. We ultimately leave that to future research, but will return to the discussion of potential phonetic analysis in §7. The spoken data are discussed in §6.2.

4.2. Descriptive statistics

In this section, we present our primary data set, namely comparative forms with a linearly following phrasal SoC extracted from ARM.[8]

Table 5 provides the overall summary of the data with respect to stress in the comparative form and the SoC. It shows the cross-tabulation of all stress locations in the comparative (columns) and in the SoC (rows) that occurred in our data sample.[9] (13) shows the interpretation of a data cell of Table 5.

(13) 4,681 Number (*n*) of examples with suffix stress in the comparative (column 1) and initial stress in the SoC (row 1), as in {*sil'née, sil'néj*} *zébry* 'stronger than a zebra'
1.8% Percentage of truncated examples (e.g., *sil'néj zébry*) out of *n*

[8] Descriptive statistical analysis was performed using Python; all plots were generated using the `seaborn` library (Waskom 2021). The mixed-effects logistic regression modelling presented in §5 was performed using R version 4.3.1 (R Core Team 2023) and, specifically, the `lme4` package (Bates *et al.* 2015).

[9] In this table and elsewhere (including Tables 12, 13 and 16), we use the following example words: *sil´née* 'stronger', *akkurátnee* 'more careful', *vnimátel´nee* 'more attentive'; *zébry* 'zebra.GEN.SG', *sloná* 'elephant.GEN.SG', *nosoróga* 'rhinoceros.GEN.SG', *vypusknikóv* 'alumni.GEN.PL', *universitéta* 'university.GEN.SG', *professionalízma* 'professionalism.GEN.SG', *samoidentifikácii* 'self-identification.GEN.SG'. This does not imply that all combinations of these adjectives and nouns actually occur in the data.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 5. *Share of truncated forms by stress pattern in the comparative and the SoC.*

	Stress in comparative							
	Suffix (*sil'née*)		Stem final (*akkurátnee*)		Stem penult (*vnimátel'nee*)		Total	
Stress in SoC	*n*	Trunc.	*n*	Trunc.	*n*	Trunc.	*n*	Trunc.
1 (*zébry*)	4,681	1.8%	918	14.5%	267	13.1%	5,866	4.3%
2 (*sloná*)	4,511	3.3%	1,380	12.0%	452	12.8%	6,343	5.8%
3 (*nosoróga*)	2,229	3.0%	738	12.2%	221	10.0%	3,188	5.6%
4 (*vypuskníkov*)	488	3.1%	217	8.3%	70	8.6%	775	5.0%
5 (*universitéta*)	101	3.0%	44	13.6%	9	11.1%	154	6.5%
6 (*professionalízma*)	20	5.0%	13	7.7%			33	6.1%
7 (*samoidentifikácii*)	2	0.0%	1	0.0%			3	0.0%
Total	12,032	2.7%	3,311	12.5%	1,019	12.0%	16,362	5.2%

That is, this cell tells the reader that there were a total of 4,681 examples of comparative forms with stress on the suffix linearly followed by a SoC with initial stress, like {*sil'née, sil'néj*} *zébry,* {*bystrée, bystréj*} *tígra* 'faster than a tiger', etc. 85 examples out of that total (1.8%) involved truncation: *sil'néj zébry, bystréj tígra,* etc.; the remaining 4,596 had a disyllabic, non-truncated comparative suffix, as in *sil'née zébry, bystrée tígra,* etc.

Below, we break the data down by stress placement in the comparative and the SoC and plot the distribution of the truncated *-ej* and disyllabic *-ee.*

4.2.1. Share of truncated forms in comparatives with suffix vs. *stem stress*

Comparatives with suffix stress are truncated in 2.7% of cases, while comparatives with stem stress are truncated in 12.4% (aggregating the data from columns 2 and 3 of Table 5). This is presented visually in Figure 1, which also includes 95% confidence intervals for proportions. Figure 1 only shows combinations of comparatives where the SoC has stress on one of its first four syllables; examples in which stress falls on the fifth syllable of the SoC (or later) are too rare to be plotted meaningfully.

4.2.2. Share of truncated forms in comparatives with suffix stress

Table 6 expands the first four cells of the first column of Table 5. Again, we ignore SoCs with stress on the fifth syllable or later.

Observe that the share of truncated forms is the lowest when stress in the SoC falls on the first syllable (95% confidence interval for SoCs stressed on the first syllable lies below the confidence intervals for the second and third syllable, and the confidence interval for the fourth syllable is too broad because of the sparseness of the data). In other words, there are fewer instances of truncation (*sil'néj*) in a configuration like {*sil'née, sil'néj*} *zébry* 'stronger than a zebra' than in cases where the SoC has stress further away from the beginning.

4.2.3. Share of truncated forms in comparatives with stem stress

Tables 7 and 8 present the proportions of truncated comparative forms in comparatives with stress on the final syllable of the stem (Table 7, shown in red in Figure 1), and with stress on the penultimate syllable of the stem (Table 8, green in Figure 1).

A slight decrease in the share of truncated forms with stress further from the beginning of the word in SoC can be observed in Tables 7 and 8. However, as the analysis in §5 reveals, this effect is not significant.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

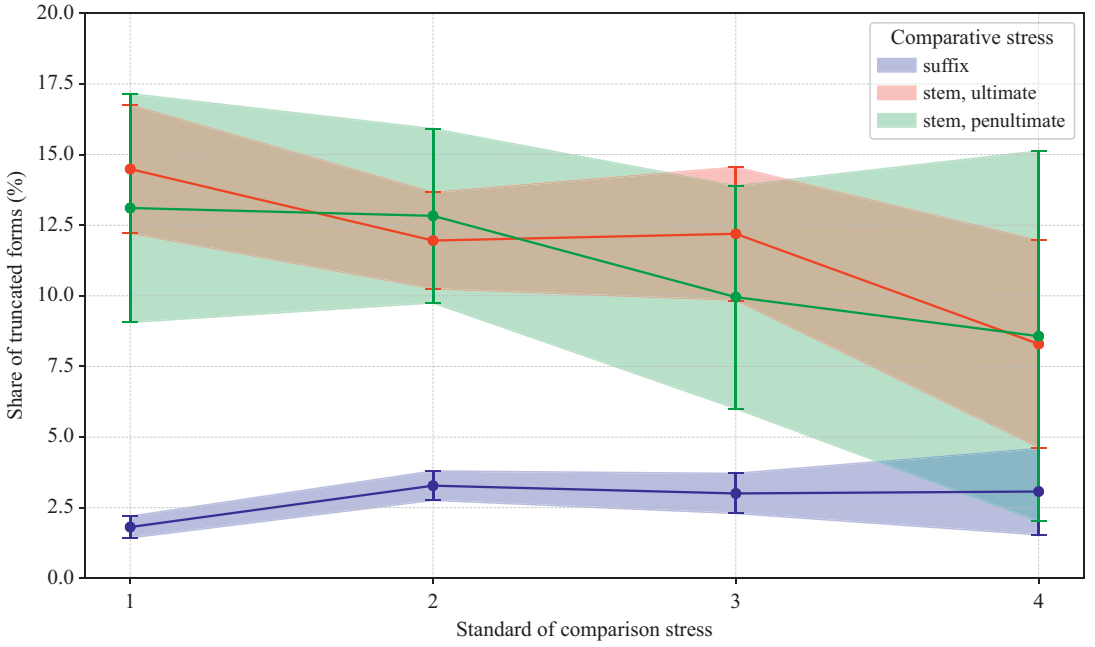


Figure 1. *Share of truncated forms by stress pattern in the comparative and the SoC.*

Table 6. *Share of truncated forms in suffix-stressed comparatives.*

	Truncated (*-éj*)		Non-truncated (*-ée*)		
SoC stress	*n*	%	*n*	%	Total
1	85	1.8%	4,596	98.2%	4,681
2	148	3.3%	4,363	96.7%	4,511
3	67	3.0%	2,162	97.0%	2,229
4	15	3.1%	473	96.9%	488
Total	315	2.6%	11,594	97.4%	11,909

Table 7. *Share of truncated forms in comparatives with stem-final stress.*

	Truncated (*-éj*)		Non-truncated (*-ée*)		
SoC stress	*n*	%	*n*	%	Total
1	133	14.5%	785	85.5%	918
2	165	12.0%	1,215	88.0%	1,380
3	90	12.2%	648	87.8%	738
4	18	8.3%	199	91.7%	217
Total	406	12.5%	2,847	87.5%	3,253

4.2.4. Share of truncation and vowel hiatus between the comparative and SoC

Recall from §2.3 that *-ee* ends in a vowel, [ə] or [ɪ̈], whereas *-ej* ends in a glide or a consonant: [j] or [i̯], or [ç] if devoiced. If the SoC is vowel-initial, this results in vowel hiatus with *-ee* but not with *-ej*. This is a potentially confounding factor if we seek to establish the presence of rhythmic effects. Table 9 shows the share of truncated comparative forms before vowel-initial *vs.* consonant-initial SoCs.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 8. *Share of truncated forms in comparatives with stem-penultimate stress.*

	Truncated (*-éj*)		Non-truncated (*-ée*)		
SoC stress	*n*	%	*n*	%	Total
1	35	13.1%	232	86.9%	267
2	58	12.8%	394	87.2%	452
3	22	10.0%	199	90.0%	221
4	6	8.6%	64	91.4%	70
Total	121	12.0%	889	88.0%	1,010

Table 9. *Share of truncated forms by initial segment of SoC.*

	Truncated (*-ej*)		Non-truncated (*-ee*)		
Initial segment	*n*	%	*n*	%	Total
C	730	5.2%	13,204	94.8%	13,934
V	124	5.1%	2,304	94.9%	2,428
Total	854	5.2%	15,508	94.8%	16,362

Note that in Table 9, we classify words that begin in a palatal glide [j] (i.e., words spelt with initial ⟨е⟩, ⟨я⟩, ⟨ю⟩, ⟨ё⟩ or ⟨й⟩) together with words that begin in [+consonantal] segments, as in both cases no vowel hiatus is created across the word boundary.

The data show nearly identical truncation rates (~5%) regardless of whether the SoC is V- or C-initial, suggesting that vowel hiatus is unlikely to have an impact on truncation.

4.2.5. Share of truncation in individual adjectives

The share of truncated forms in *-ej* out of all tokens of the comparative is different for individual adjectives, ranging from 0% to 27.3% in our data. For instance, the adjective *ujútnee* 'cosier' occurs 16 times in our data, four of them exhibiting truncation ($4/16 = 25\%$), whereas *komfórtnee* 'more comfortable' occurs 37 times, two of them exhibiting truncation ($2/37 \approx 5.4\%$), even though these two adjectives are similar in meaning and have the same syllable count and stress pattern. Details of truncation in individual adjectives are provided in Table A1 in the Appendix.

5. Analysis

We are now prepared to analyse the data explored in §4.2. Our objective is to find a model that most effectively predicts the stochastic behaviour of a binary variable: truncation (*-ej*) or absence thereof (*-ee*) in the comparative suffix. To that end, we employ mixed-effects logistic regression modelling, which is a special case of generalised linear mixed-effects modelling (GLMM); see Gries (2021) and the references therein.

First, we encode each data point in our data set as a comparative tableau (Prince 2002). Consider a toy tableau with three constraints: Max, *Lapse and *ExtendedLapse. Max is defined in (14); the other two constraints were introduced in (9b) and (10).

(14) Max: Assign a violation (∗) for every segment in the UR with no correspondent in the SR.

Take example (15), showing a monosyllabic stem with suffix stress in the comparative (*sil'néj*) and a SoC with peninitial stress (*sloná*). The tableau in (16) shows the violations of the three constraints, as well as the difference between candidates (16a) and (16b).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

(15) *Kít **sil'néj sloná** na porjadok.*
'A whale is way stronger than an elephant'.

(16)

/σ-ˈeje # σˈσ/	MAX	*LAPSE	*EXTENDEDLAPSE
a. σ-ˈeje # σˈσ		1	
b. σ-ˈej # σˈσ	1		
a ~ b	−1	1	0

A negative difference in violations indicates a constraint's preference for the non-truncated candidate; for example, MAX favours (16a). Conversely, a positive difference signifies a preference for the truncated candidate: *LAPSE prefers (16b). A zero difference means that a constraint does not differentiate between candidates: *EXTENDEDLAPSE does not distinguish between (16a) and (16b).

The bottom row of (16) serves as input to the model. That is, given just these three constraints, (15) would be represented as a vector of length 3 with the values (−1, 1, 0). The class y to which a context belongs is 1 if the comparative is truncated and 0 otherwise; for (15), $y = 1$. If (15) had no truncation (*...sil'née sloná...*), y would be 0, while the corresponding vector of predictor variables would stay the same. The model predicts the probability of truncation $\hat{p}$, which should ideally be as close as possible to the actual value of y. This form of logistic regression constitutes a MaxEnt model (Hayes & Wilson 2008 and the references therein), representing the simplest form of such a model with a binary response variable (Fauconnier 2015).

Now, we introduce the effects that were included in the model. They can be grouped into three categories. The categories in (17) and (18) include variables related to the comparative form and to the SoC, respectively. Among these, the variable *adjective* in (17a) is a random effect; the variables in (17b) and (18) are fixed effects.

(17) *Variables related to the comparative*
   a. *Adjective*: 65 values (see Table A1 in the Appendix for a list of all adjectives).
   b. *Stress in the comparative form* (*C*): 1 for comparatives with suffix stress (e.g., *sil'n-ée*), 2 for comparatives with stress on the ultimate syllable of the stem (e.g., *akkurátn-ee*), 3 for comparatives with stress on the penultimate syllable of the stem (e.g., *vnimátel'n-ee*).

Recall from §4.1 that our primary set of data included SoC that either contained a bare noun (as in *sil'née zébry* 'stronger than a zebra') or a noun modified by an adjective (as in *sil'née polosátoj zébry* 'stronger than a striped zebra'). Notice that in these two examples the stress of the head noun in the SoC is the same (*zébry* has initial stress), but stress placement in the word that linearly follows the comparative form is different: initial in the former, *zébry,* and postpeninitial in the latter, *polosátoj.* It is the location of stress in the word immediately following the comparative – and the distance between it and stress in the comparative itself – that we are interested in, hence the variable *S* is formulated as in (18a). The variable in (18b) encodes the presence or absence of an adjective in a given input in our data.

(18) *Variables related to the SoC*
   a. *Stress in the first word of the SoC* (*S*): values from 1 (initial stress) to 7 (stress on the seventh syllable from the beginning of the word).
   b. *Modified SoC*: 0 if SoC is a bare noun, 1 if SoC is an adjective–noun sequence.

The third group of variables relates to the set of rhythmic constraints whose effects on our data we will be testing, which were introduced in §3.2. Given the potential influence on the choice between *-ee* and *-ej* of vowel hiatus created between the comparative form in *-ee* and a vowel-initial SoC, we augment the set with the constraint ONSET (alternatively, NOHIATUS; see Casali 2011), defined in (19).

(19) ONSET: Assign a violation (∗) for every syllable that does not have an onset.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

With the constraint set in place, we are almost ready to encode it as variables. The tableaux in (20) through (22) illustrate the violation patterns of the constraints in (9)–(12) and (14) for inputs with suffix stress in the comparative and initial, peninitial or postpeninitial stress in the SoC.[10] Onset is not shown as its violations are orthogonal to syllabic count and stress placement.

(20)

/σ-ˈeje # ˈσσσ/	*Clash	*Lapse	Eurhythmy Distance	*Extended Lapse	*Clash+	Max
a. σ-ˈeje # ˈσσσ		1			−1	
b. σ-ˈej # ˈσσσ	1	1	1			1
a ~ b	−1	0	−1	0	−1	−1

(21)

/σ-ˈeje # σˈσσ/	*Clash	*Lapse	Eurhythmy Distance	*Extended Lapse	*Clash+	Max
a. σ-ˈeje # σˈσσ		1	1		−2	
b. σ-ˈej # σˈσσ					−1	1
a ~ b	0	1	1	0	−1	−1

(22)

/σ-ˈeje # σσˈσ/	*Clash	*Lapse	Eurhythmy Distance	*Extended Lapse	*Clash+	Max
a. σ-ˈeje # σσˈσ		2	2	1	−3	
b. σ-ˈej # σσˈσ		1	1		−2	1
a ~ b	0	1	1	1	−1	−1

To compute the values of the variables corresponding to individual constraints, we define the parameter D_{UR} in (23), which corresponds to the number of unstressed syllables between the two stresses in the underlying representation.

(23) $D_{UR} = C + S - 1$

The list in (24) shows the encoding for constraints from (9)–(12), (14) and (19).

(24) *Encodings of the relevant OT constraints*

a. Max: −1
b. Onset: 0 if SoC is C-initial, 1 if SoC is V-initial
c. *Clash: −1 if $D_{UR} = 1$, 0 if $D_{UR} \geq 2$
d. *Lapse: 0 if $D_{UR} = 1$, 1 if $D_{UR} \geq 2$
e. *ExtendedLapse: 0 if $D_{UR} \leq 2$, 1 if $D_{UR} \geq 3$
f. *Clash+: −1
g. Eurhythmy Distance: −1 if $D_{UR} = 1$, 1 if $D_{UR} \geq 2$

Unfortunately, we cannot include all these constraints in the model, especially because some of them are equivalent for the data we have. The tableaux in (20)–(22) illustrate that for the binary choice between *-ee* and *-ej,* the effects of Max and *Clash+ are the same: no matter the input, they prefer

[10] Recall that *Clash+ is a positive constraint (Gouskova & Roon 2013: 425) which assigns a reward for every syllable separating two stresses. Somewhat counterintuitively, we indicate the reward it assigns with the value −1 in (20)–(22) – that is, the opposite of a violation (1).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

the disyllabic *-ee* either because no segment is deleted (MAX) or because *-ee* results in an extra syllable between the two stresses (*CLASH+). In other words, if the disyllabic *-ee* has *n* violations of either MAX or *CLASH+, the monosyllabic *-ej* will have *n* + 1 violations. Since these constraints exert the same effect across all inputs, namely a preference for *-ee,* they cannot be distinguished from the intercept of the model.

Second, the relative weight of *CLASH, *LAPSE and EURHYTHMYDISTANCE cannot be established, since these constraints only make a binary distinction between a context with one syllable intervening between stresses in the UR (as in *sil'née zébry*) and all other contexts. In (20), only *CLASH and not *LAPSE distinguishes between the candidates with and without truncation; in (21) and (22), *LAPSE does but *CLASH does not; across all tableaux, EURHYTHMYDISTANCE exhibits the sum of violations of *CLASH and *LAPSE. Therefore, these three constraints cannot be simultaneously included in the model on account of their collinearity. In what follows, we include only the variable coded as *CLASH (note that we could have used any of the three encodings to the same effect) and dub it Rhythmic Alternation (RA). The resulting list of phonological variables is presented in (25).

(25) *Phonological variables*

a.	ONSET:	0 if SoC is C-initial, 1 if SoC is V-initial
b.	RHYTHMIC ALTERNATION (RA):	-1 if $D_{UR} = 1$, 0 if $D_{UR} \geq 2$
c.	*EXTENDEDLAPSE:	0 if $D_{UR} \leq 2$, 1 if $D_{UR} \geq 3$

To summarise, the variables that we included in the model-selection procedure are given in (17)–(18) and (25). To reiterate, we treat the categorical variable *adjective* in (17a), which has 65 possible values, as a random effect, and all other explanatory variables as fixed effects. We keep our model more readily interpretable by only introducing random intercepts but not random slopes for different adjectives. That is, we assume that the influence of fixed effects included in the model is the same for all adjectives, but each adjective can have its own baseline share of truncation.

To find the best model, we perform forward selection of input variables (Blanchet *et al.* 2008; cf. also a bottom-up procedure for grammar selection in Hayes *et al.* 2012: 713, although the implementation differs).[11] At Step 0, we start with a model that only includes *adjective* as a random effect and no fixed effects. At Step 1, we analyse a set of models obtained by adding one fixed effect and select the best model based on the lowest value of the Bayesian information criterion (BIC; Schwarz 1978).[12] At the subsequent steps, we increase the number of variables by one and also test the interaction between the variable that we are currently attempting to include and all fixed effects that are in the model already. We proceed until we cannot get a lower BIC value or any newly included factors prove to be non-significant for the model. The modelling code, together with the data, is available in the Supplementary Material.

As shown in Table 10, the forward selection process stops after adding two variables to the model, namely RA and Comparative stress (*C*). No interactions between variables have been found to improve model fit. The parameters of the resulting model are presented in Table 11. The intercept values for random effects (i.e., for individual adjectives) are included in Table A1 in the Appendix. Residual diagnostics performed using the DHARMa package revealed no significant deviations from uniformity, indicating a satisfactory fit of the GLM to the data (Hartig 2022).

The model above differentiates between four cases (setting random effects aside), as shown in Table 12. For individual adjectives, the model predicts truncation probabilities $\hat{p}$ between 0 and 21.3%. The maximum predicted value is for the adjective *prekrásnee* 'more beautiful', with standards

[11] Backward selection would be impractical in our case: a model including all variables would behave in an unpredictable way because many of the factors in (17), (18) and (25) encode basically the same information and are strongly intercorrelated.

[12] Alternatively, we could use the Akaike information criterion (AIC). We prefer BIC since it penalises model complexity more heavily, making it more stringent against overfitting with large data sets (Schwarz 1978). An application of AIC to our data resulted in a model with two additional predictors, neither of them significant even at the 0.1 level.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 10. *BIC values for different models.*

No fixed effects: 6,134.9	Step 1	Step 2	Step 3
Comparative stress (*C*)	6,124.0	**6,109.4**	
SoC stress (*S*)	6,143.9	6,121.5	6,115.6
Modified SoC	6,144.1	6,125.1	6,119.0
ONSET	6,143.4	6,122.2	6,116.1
RA	**6,115.6**		
*EXTENDEDLAPSE	6,143.8	6,124.8	6,117.5

Table 11. *Mixed-effects logistic regression: fixed effects of the resulting model.*

Variable	*b*	95% CI	SE	*z*	*p*
Intercept	−3.85	[−4.40; −3.30]	0.28	−13.7	$< 2 \times 10^{-16}$
Comparative stress (*C*)	0.72	[0.40; 1.04]	0.16	4.5	8×10^{-6}
RA	0.61	[0.36; 0.86]	0.13	4.8	2×10^{-6}

Table 12. *Probability of truncation ($\hat{p}$) by stress configuration.*

RA	*C*	Examples	$\hat{p}$
−1	1	*sil'née zébry*	2.3%
0	1	*sil'née sloná*	4.2%
0	2	*akkurátnee zébry, akkurátnee sloná*	8.2%
0	3	*vnimátel'nee zébry, vnimátel'nee sloná*	15.5%

of comparison bearing stress on the second, third and fourth syllables (that is, no possible clash configuration). The actual share of truncated forms in this case is $50/223 \approx 22.4\%$.

To sum up, our model demonstrates the effects of two predictors: Rhythmic Alternation (RA) and stress in the comparative form (*C*):

- The probability of truncation in a configuration with stress on the comparative suffix and the initial syllable of the linearly following SoC (*sil'née zébry*) is significantly smaller than the probability of truncation in all other cases.
- The probability of truncation is smaller for comparative forms with stress on the suffix (*sil'n-ée*) than in forms with stress on the final syllable of the stem (*akkurátn-ee*), and smaller than in comparative forms with stress on the penultimate syllable of the stem (*vnimátel'n-ee*).

Before we proceed to the interpretation of these findings in §7, we turn to the data sets with clausal SoC and spoken data.

6. Expanding the data set

6.1. Clausal standards of comparison

In this section, we address comparative forms with and without truncation followed by clausal standards of comparison with the complementiser *čem*. We followed the same data collection and annotation procedure for this data set as for the phrasal SoC, outlined in §4.1. After manual filtering of the extracted examples, we arrived at 4,529 data points of the shape comparative + *čem* (A) N, where N is the noun expressing the SoC and A is an optional adjective modifying N, as in *bystrée/éj čem* (*polosátaja*) *zébra* 'faster than a (striped) zebra'.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 13. *Share of truncated forms by stress pattern; clausal SoC with* čem.

	Stress in comparative							
	Suffix (*sil'née*)		Stem final (*akkurátnee*)		Stem penult (*vnimátel'nee*)		Total	
Stress in first word of SoC	*n*	Trunc.	*n*	Trunc.	*n*	Trunc.	*n*	Trunc.
1 (*zébry*)	902	3.1%	366	13.1%	85	11.8%	1,353	6.4%
2 (*sloná*)	1,026	4.0%	613	12.6%	190	11.1%	1,829	7.6%
3 (*nosoróga*)	554	2.2%	348	14.9%	99	11.1%	1,001	7.5%
4 (*vypusknikóv*)	141	0.7%	102	7.8%	32	15.6%	275	5.1%
5 (*universitéta*)	30	6.7%	25	12.0%	2	0.0%	57	8.8%
6 (*professionalízma*)	6	0.0%	5	20.0%	1	0.0%	12	8.3%
7 (*samoidentifikácii*)	1	0.0%	0		1	0.0%	2	0.0%
Total	2,660	3.2%	1,459	13.0%	410	11.5%	4,529	7.1%

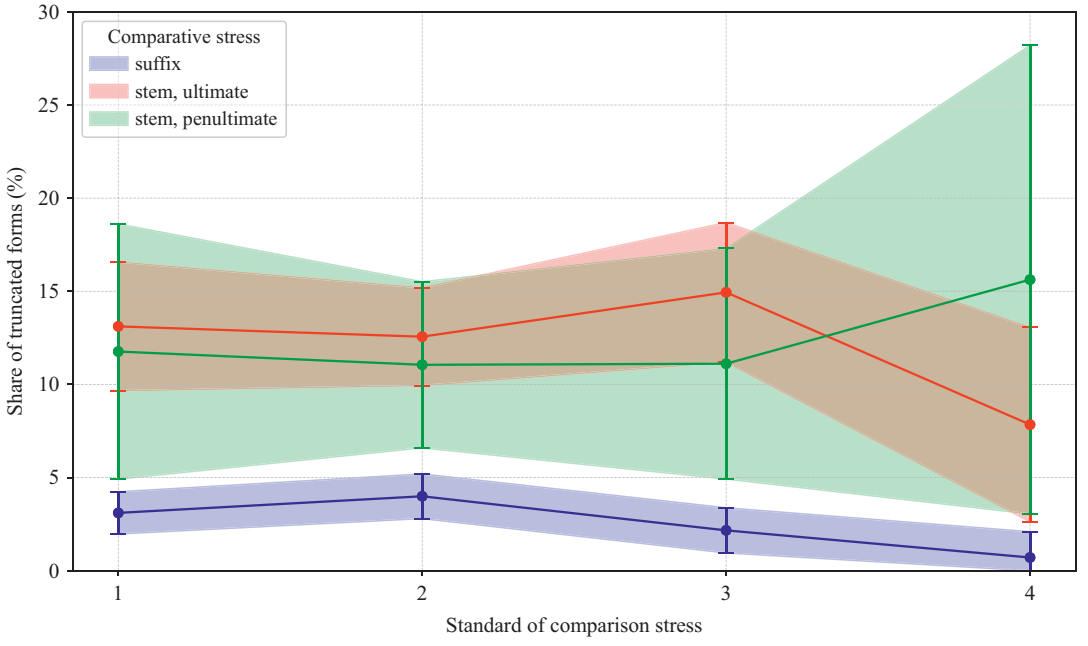


Figure 2. *Share of truncated forms by stress pattern in the comparative and the SoC; Clausal SoC with* čem.

Table 13 (which is organised similarly to Table 5 in §4.2) presents the aggregated counts by stress configuration in the comparative and in the standard of comparison. Recall that the rows indicate the location of stress in the first word of the SoC – that is, the word that immediately follows the complementiser *čem.*

Figure 2 plots the share of truncated forms for comparatives with suffix stress (blue), stress on the last syllable of the stem (red) and stress on the penultimate syllable of the stem (green) by stress in the first word of the SoC. As before, we only plot examples with stress on one of the first four syllables of the SoC. Similarly to the data with phrasal SoC (Figure 1), we observe that comparatives with stress on the suffix exhibit a lower share of truncation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 14. *BIC values for different models; clausal SoC with* čem.

No fixed effects: 2,210.2	Step 1	Step 2
Comparative stress (*C*)	**2,197.9**	
SoC stress (*S*)	2,218.5	2,206.1
Modified SoC	2,216.4	2,204.2
Onset	2,218.3	2,205.9
RA	2,210.2	2,197.9
*ExtendedLapse	2,216.5	2,205.9

Table 15. *Parameters of the resulting model; clausal SoC with* čem.

Variable	*b*	95% CI	SE	*z*	*p*
Intercept	−3.98	[−4.55; −3.41]	0.29	−13.6	$< 2 \times 10^{-16}$
Comparative stress (*C*)	0.84	[0.53; 1.15]	0.16	5.3	1×10^{-7}

In order to test which of the predictors we explored in §5 affect truncation of the comparative forms in configurations with clausal SoC, we used the same procedure for GLM selection. Table 14 shows the outcome of this procedure.

As the table shows, the BIC value is only improved after adding comparative stress (*C*) and no additional predictors. That is, the location of stress in the comparative itself is the only factor which significantly contributes to the resulting model, summarised in Table 15.

Before we move on to spoken data, recall that in §4.1 we made the assumption that comparatives with clausal and phrasal SoCs do not compete in phonology, this choice being made earlier in the derivation by the syntactic component of the grammar. We would like to argue that this assumption does not weaken the results obtained in §5. If the two constructions were in competition within phonology, we could expect to find a higher proportion of clausal to phrasal SoC in configurations with suffix stress in the comparative and initial stress in the SoC, namely the case where truncation is avoided. The avoidance of truncation in that context could not be a parasitic effect of the exclusion of clausal SoC from consideration – if anything, we could be underestimating the size of the observed effect.

6.2. Spoken corpus

Lastly, as mentioned in §4.1, we collected data from the multimedia subcorpus of the RNC, namely comparative forms from the same list of 65 adjectives followed by a phrasal SoC (e.g., *bystrée*/*éj tígra*). In doing so, we relied on manual transcription that classified a given example as either exhibiting truncation or not. Recall that these data are also different in that we included not only SoC expressed by bare nouns or nouns modified by adjectives, but any phrasal SoC whatsoever. Table 16 presents the results. As in Tables 5 and 13, the relevant stress in the SoC is that in the first word of the SoC, immediately following the comparative form.

One thing is immediately apparent in spite of the scarcity of these data, namely that the share of truncation (indicated by the percentage out of the total number of examples of each type) is much higher in the spoken RNC data than in the written ARM data. This is to be expected: recall from §2 that the disyllabic *-ee* is the prescribed spelling of this suffix, and spelling is something that speakers can easily control.

Figure 3 plots these data in the format already familiar to the reader. We do not include comparatives with stress on the penultimate syllable of the stem or SoC with stress on the fourth syllable from the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table 16. *Share of truncated forms by stress pattern; spoken data.*

	Stress in comparative							
	Suffix (*sil'née*)		Stem final (*akkurátnee*)		Stem penult (*vnimátel'nee*)		Total	
Stress in first word of SoC	*n*	Trunc.	*n*	Trunc.	*n*	Trunc.	*n*	Trunc.
1 (*zébry*)	72	20.8%	12	50.0%	1	0.0%	85	24.7%
2 (*sloná*)	64	37.5%	31	58.1%	0		95	44.2%
3 (*nosoróga*)	11	27.3%	0		1	0.0%	12	25.0%
4 (*vypusknikóv*)	1	0.0%	0		0		1	0.0%
Total	148	28.4%	43	55.8%	2	0.0%	193	34.2%

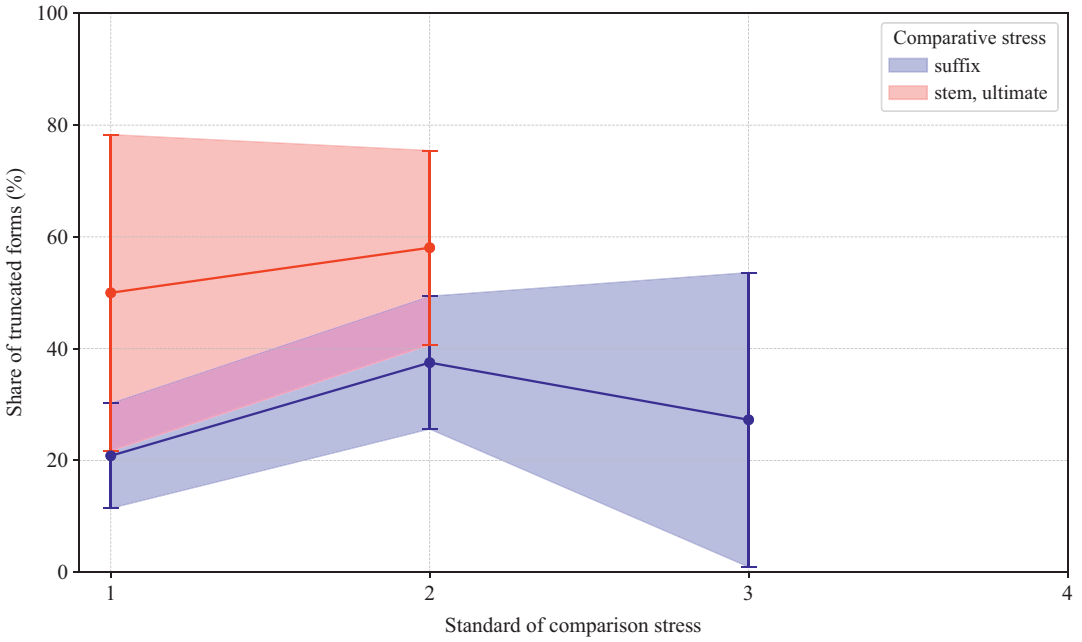


Figure 3. *Share of truncated forms by stress pattern in the comparative and the SoC; spoken data.*

beginning of the word (nor on the third for comparatives with stem-final stress) because of the scarcity of relevant data points.

The plot suggests a difference in the propensity to truncate the final syllable of the suffix *-ee* between comparatives with stress on the stem and on the suffix, as well as a lower share of truncation in comparatives with suffix stress followed by initial stress in the SoC (as in *bystrée*/*éj zébry*), but the paucity of these data does not allow for further analysis. We leave a closer investigation of spoken data for future research.

7. Discussion

To recapitulate, we found in the corpus study of comparative forms in *-ee*/*ej* that, first, the rate of truncation (i.e., the share of forms with *-ej* out of the total number of occurrences) depends on the position of stress in the comparative form itself: the probability of truncation is higher for comparatives

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

with stem stress compared to suffix stress (Table 12). This effect is observed throughout comparatives with phrasal SoCs (§5) and ones with clausal SoCs (§6.1). This finding confirms Es′kova's (2011: 136) hypothesis that stress placement could be a factor in the distribution of *-ee vs. -ej* (§2.4).

Second, we found that the probability of truncation is smaller in cases where a comparative form with the stressed suffix is followed by a word expressing the standard of comparison which has initial stress, for example, *bystrée*/*éj zébry* 'faster than a zebra' (§5) than in other configurations with a comparative form with suffix stress, or indeed with stem stress. The same dispreference for truncation is not observed with clausal SoC, where an unstressed syllable is guaranteed to separate the comparative form and the first stressed syllable in the SoC due to the presence of the unstressed complementiser *čem*, for example, *bystrée*/*éj čem zébra* (§6.1).

In interpreting the first finding, it is important to bear in mind that we did not find an effect of stress placement in the SoC, nor of an interaction between stress in the comparative form and the SoC. Therefore, the phenomenon in question cannot be attributed to the distance between the relevant stresses – it is conditioned by stress in the comparative form alone.

We would like to suggest that the higher propensity of unstressed *-ee* to truncate indicates greater reduction in final unstressed vowels that are not adjacent to stressed syllables. First, Barnes (2007) argues that 'radical' reduction in Russian – that is, roughly, reduction to schwa – which generally applies to vowels in all unstressed syllables except the first pretonic, results from reduced duration of these vowels. Vowels in the first pretonic syllable, on the contrary, can be as long as those in the stressed syllable (Kniazev 2006) and do not undergo radical reduction. Vowels in posttonic syllables do not come close to pretonic syllables in duration, but if they mirror the pattern of exhibiting greater duration when proximate to the stressed syllable, we can expect a lesser propensity to elision from the immediately posttonic vowel in *-ée.* Second, under vowel hiatus, a pretonic or posttonic vowel can be exempt from reduction, as in *koala* [koˈalə]~[kʷˈalə] 'koala', *poèt* [poˈɛt̪]~[pʷˈɛt̪] 'poet', *xaos* [ˈxaos̪]~[ˈxaəs̪] 'chaos' and *kakao* [kʷˈkao]~[kʷˈkaə] 'cocoa'. Regarding vowel hiatus in unstressed syllables, sources differ: for words like *kooperacija* 'cooperation' or *sootnošenie* 'ratio', some transcribe the hiatus sequence as [ʊ.ʊ], while others give [ə.ʊ] (see Iosad 2012: 524 and the references therein). If it is the case that vowels adjacent to the stressed vowel are longer and undergo less severe reduction, this would also align with the lesser probability of truncation in *-ée,* where at least in realisations with glide deletion, the final vowel is adjacent to the stressed vowel, in contrast with the unstressed *-ee.* We are unaware of instrumental studies that investigate either the relative duration of posttonic vowels or the duration and quality of vowels in hiatus, and we leave such an investigation for future research.

Alternatively, the observed effect could be attributed to NonFinality (Walker 2000): truncation of the final vowel in *-ée* yields word-final stress and thus incurs a violation of this constraint. While avoidance of word-final stress has been documented in many languages (see Gordon 2002 and the references therein), we are not aware of its reflexes elsewhere in Russian. In fact, for consonant-final words, the default location of stress is indeed on the final syllable (Lavitskaya & Kabak 2014), even though this default may be due to morphology (Crosswhite *et al.* 2003) and does not rule out the influence of NonFinality. In the absence of corroborating evidence, this interpretation remains speculative, and for now, we leave it at that.

We now turn to the second finding, namely the depressed probability of truncation in cases like *bystrée zébry.* We argue that this is an effect of rhythmic optimisation: the reduction in the probability of truncation is found in the only configuration where truncation yields a categorical stress clash, as in *bystréj zébry.* This phenomenon supports the central hypothesis of this study, formulated in §3, that phrasal rhythm impacts the distribution of the disyllabic *-ee* and monosyllabic *-ej.*

However, this tendency may be attributed to the effect of several alternative rhythmic constraints. The tableaux in (20)–(22) expose the crucial difficulty in the analysis of this finding. As §5 shows, our model is designed to predict the probability of one or the other outcome of a binary choice: truncation (*-ej*) or no truncation (*-ee*). Regardless of stress placement in any given surface representation evaluated by the model, it either has the same number of vowels (and thus syllables) as the corresponding underlying representation or is one vowel (syllable) shorter. The constraint *Clash only distinguishes

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

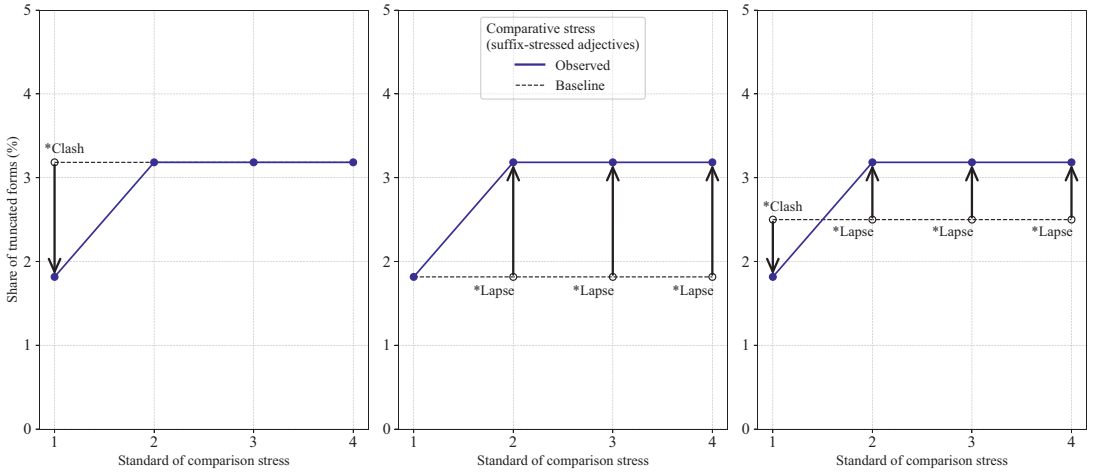


Figure 4. *Three interpretations of the difference in the probability of truncation.*

between *-ej* and *-ee* when the suffix is stressed and phrasal SoC has initial stress, as in (20): elsewhere, no stress clash arises. *Lapse distinguishes between *-ej* and *-ee* in all other cases, where *-ee* incurs one extra violation, as in (21) and (22), but never in cases like (20), where a stress lapse is not possible: there is no lapse in *bystrée zébry,* nor in *bystréj zébry.* (EurhythmyDistance is violated throughout; it prefers the faithful candidate in (20), which does not introduce a stress clash, like *Clash, and the truncated candidate elsewhere, like *Lapse.) In other words, the contexts distinguished by *Clash and *Lapse are complement sets of each other.

As a result, a predictor encoded like (25b), our RA, is amenable to several competing interpretations. These are illustrated in Figure 4.

Plotted in blue in all three panels is the difference in propensity to truncate the final vowel in *-ée* across four environments, distinguished by the position of stress in the following SoC: initial (as in *bystrée zébry*), on the second syllable (*bystrée sloná*), on the third syllable (*bystrée nosoróga*) or on the fourth syllable (*bystrée vypusknikóv*). This difference is compatible with three interpretations. The left-hand panel shows a significant effect of *Clash which decreases the probability of truncation in the only case where it leads to stress clash and no significant effect of *Lapse. The centre panel shows a significant effect of *Lapse increasing the probability of truncation where it uniformly reduces the length of the lapse sequence by one unstressed syllable and no significant effect of *Clash. Finally, the right-hand panel shows that the same observed effect is compatible with a combined effect of both *Clash and *Lapse (or alternatively, an effect of EurhythmyDistance).

The compatibility of our finding with these three interpretations is due, first, to the binary choice between candidates in each context and the difference in the presence or absence of one syllable between those candidates. Where *Clash applies, *Lapse never does, and *vice versa*; candidates that delete or add extra syllables are categorically ruled out by Max and Dep and never surface. Russian exhibits several cases similar to the comparative *-ee*/*ej* in showing variation between the presence and absence of a final vowel in a word form or clitic: the feminine instrumental singular *-oj(u),* as in *ruk-oju* ~ *ruk-oj* 'hand-INS.SG';[13] the feminine instrumental plural *-(ja)mi,* as in *dočer-jami* ~ *dočer'-mi* 'daughter-INS.PL'; the clitics *b(y)* and *ž(e)*; and the complementiser *čtob(y)* – it is possible that all these are sensitive to rhythm, but they all have the same profile, in which the variants differ by one vowel and *Clash and *Lapse cannot be contrasted.

[13]This case is different from comparatives in *-ee*/*ej* in that the variant without truncation has become virtually obsolete, and the frequency of truncation is influenced by the syntactic position of the instrumental form (see Piperski 2017 and the references therein). Note also that the /jV#/~/j#/ alternation is not shared by all Russian desinences ending in /jV#/: for example, neither the adjectival feminine accusative *-uju* nor the nominative *-aja* exhibits such variation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Figure 4 illustrates the second component crucial to the disambiguation of the effects of these constraints: the baseline. If we know the baseline distribution of the competing variants modulo the impact of rhythmic constraints, the task then reduces to the comparison between the observed values in cases where there is a stress clash or lapse and the baseline. In the case of the comparative *-ee,* the baseline distribution is obscured. Phrase-medially, the influence of *Clash and/or *Lapse would always apply, so if we took the baseline to be the average share of *-ej vs. -ee* across phrase-medial contexts, that average would have those effects baked into it. Alternatively, we could take the distribution of *-ee vs. -ej* phrase-finally to represent the baseline. Unfortunately, this assumption is equally problematic: phrase-final vowels in Russian exhibit significant lengthening (Kniazev 2006), which may protect the final vowel in *-ee* from apocope; in other words, the phrase-final position comes with its own set of effects to obscure the baseline.

The baseline may be more accessible for other variable phenomena. Kukhto & Piperski (2020) report a preference for binary rhythmic alternation in an experimental study of past tense verb forms with variable stress placement. Their findings present the same issue of complementary distribution between *Lapse and *Clash: in the choice between *pródal pal'tó* and *prodál pal'tó* '(he) sold a coat' only *Lapse is violated, whereas in *pródal knígu* and *prodál knígu* '(he) sold a book' only *Clash is. They do not report the distribution of the accentual variants of the forms they analyse in contexts other than the ones that violate *Clash or *Lapse, but in the case of variable stress placement the phrase-final position may be more indicative of the baseline distribution. We suggest that variable stress is a case one could turn to in order to disambiguate the effects of *Clash and *Lapse.

Finally, as §5 shows, we have not found any significant gradient effects of distance between stresses on the distribution of *-ee* and *-ej.* The expectation that such an effect could be present in these data comes primarily from the findings of Gouskova & Roon (2013), who uncover evidence supporting an earlier intuition of Avanesov (1964). In Russian compounds, only a single, primary stress is guaranteed, which occurs on the rightmost stem. However, the first stem of a binary compound may exhibit secondary stress. The presence or absence of secondary stress is conditioned by several factors, including distance: the farther apart the two syllables are, the more likely the secondary stress is to surface. Recall from §3.2 that Gouskova & Roon (2013) model this effect with a *Clash+, a positive gradient constraint. Note that in our model, again because of the binary choice between candidates that it evaluates, the effect of this constraint cannot be distinguished from the intercept; see §5. A gradient distance effect could be reflected in the influence of stress in the SoC or its interaction with stress in the comparative, yet as §5 shows, those did not come out as significant.

The issue of gradience brings us to one last note. We argued in §2.3 that there is a categorical distinction between *-ee* and *-ej,* and we have treated them as categorically distinct in the remainder of the article. An alternative that a future study relying on spoken data could explore is to treat surface representations of /eje/ as a continuum in terms of their duration and the formant structure of the final segment (more glide-like or more schwa-like), and to remain agnostic whether any given instance of the suffix is *-ee* or *-ej.* That is, instead of predicting the outcome of a binary choice, as in the model we implemented here, one would predict the duration and formant values based on location of stresses, distance between them, and so on. In addition to sidestepping the need to classify realisations of /eje/, such an approach would make it possible to detect gradient effects of rhythm, if any, along the lines of the Phonetic Spacing Hypothesis (Hayes 1984; Stanton 2019, 2021), which posits that speakers adjust the temporal spacing between stressed syllables to enhance perceptual clarity. One additional case in Russian that could be useful in testing the predictions of this hypothesis comes to mind: nouns in *-ie.* These exhibit variation between syllabic [ï] and non-syllabic [i̯] (e.g., *čtenie* [ˈtɕtʲenʲï.ə] 'reading' ~ *čten'e* [ˈtɕtʲenʲi̯ə]), and they are similar to the comparative *-ee* in that it is non-trivial to classify all instances as falling squarely into one or the other category. For instance, the Phonetic Spacing Hypothesis would predict that the duration of *-ie* in *čténie knígˈ* 'book reading' should be greater than in *čténie monográfij* 'monograph reading' in order to space the two stressed syllables further apart. A study of effects like that would require acoustic analysis of spoken data, and we leave it for future research.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

8. Conclusions

In this article, we have investigated rhythmic optimisation in Russian using a subset of synthetic comparative forms. These comparatives, which use the suffix *-ee*/*ej,* exhibit optional truncation of the final vowel. In order to investigate these comparatives, we have primarily relied on data from the Araneum Russicum Maius, a written web-based corpus of Russian. For 65 adjectives, we have extracted all comparative forms occurring before a phrasal or clausal standard of comparison, with different stress placement across the comparatives themselves and the SoC, depending on the lexical item. We claim that these data can be modelled using a set of phonological constraints which have already been invoked in studies of rhythm and Russian stress in particular. Employing GLM modelling, we then show that the rate of truncation is significantly lower in configurations where a comparative form with suffix stress precedes a phrasal standard of comparison with initial stress in the first word (e.g., *bystr-ée zébry* 'faster than a zebra'). We have not found gradient effects of distance on truncation, even though such an effect could be expected given the findings of Gouskova & Roon (2013) for secondary stress placement in Russian compounds. The location of stress in the comparative form itself also affects the rate of truncation: it is lower in comparatives with suffix stress than in comparatives with stem stress (e.g., *bystr-ée* 'faster' *vs. dostúpn-ee* 'more accessible'). We take these findings to indicate that Russian exhibits rhythmic optimisation at the phrasal level and suggest directions for future research, which involve phenomena potentially similar to the variation in *-ee*/*ej* and acoustic analysis of spoken data for those as well as for the comparative.

A. Individual adjectives included in the analysis

Table A1. *Rates of truncation for individual adjectives.*

			Before phrasal standard of comparison					
			Truncated (*-ej*)		Non-truncated (*-ee*)			
Adjective	Gloss	Comparative	*n*	%	*n*	%	Total	Intercept
akkuratnyj	'careful'	*akkurátnee*	0	0.0%	8	100.0%	8	−4.92
aktivnyj	'active'	*aktívnee*	4	4.4%	87	95.6%	91	−4.29
bezopasnyj	'safe'	*bezopásnee*	22	13.5%	141	86.5%	163	−3.35
bol'noj	'sick'	*bol'née*	1	6.7%	14	93.3%	15	−3.68
bystryj	'quick'	*bystrée*	22	1.3%	1,664	98.7%	1,686	−4.75
važnyj	'important'	*važnée*	49	2.1%	2,302	97.9%	2,351	−4.36
verojatnyj	'likely'	*verojátnee*	0	0.0%	11	100.0%	11	−4.15
vesëlyj	'happy'	*veselée*	2	9.5%	19	90.5%	21	−3.47
vidnyj	'prominent'	*vidnée*	0	0.0%	1	100.0%	1	−3.86
vkusnyj	'delicious'	*vkusnée*	8	4.3%	179	95.7%	187	−3.69
vnimatel'nyj	'attentive'	*vnimátel'nee*	0	0.0%	8	100.0%	8	−4.25
vygodnyj	'advantageous'	*výgodnee*	29	9.1%	289	90.9%	318	−4.41
glavnyj	'main'	*glavnée*	3	3.5%	83	96.5%	86	−3.84
dlinnyj	'long'	*dlinnée*	11	1.8%	587	98.2%	598	−4.36
dobryj	'kind'	*dobrée*	4	14.3%	24	85.7%	28	−3.03
dostupnyj	'accessible'	*dostúpnee*	4	6.8%	55	93.2%	59	−3.98
zdorovyj	'healthy'	*zdorovée*	2	11.1%	16	88.9%	18	−3.41
zelënyj	'green'	*zelenée*	0	0.0%	5	100.0%	5	−3.93
interesnyj	'interesting'	*interésnee*	72	18.8%	311	81.2%	383	−2.94
komfortnyj	'comfortable'	*komfórtnee*	2	5.4%	35	94.6%	37	−4.07

(Continued)

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Table A1. *Continued.*

			Before phrasal standard of comparison					
			Truncated (*-ej*)		Non-truncated (*-ee*)			
Adjective	Gloss	Comparative	*n*	%	*n*	%	Total	Intercept
krupnyj	‘large’	*krupnée*	4	0.8%	519	99.2%	523	–4.93
medlennyj	‘slow’	*médlennee*	11	5.2%	199	94.8%	210	–4.86
milyj	‘cute’	*milée*	15	18.3%	67	81.7%	82	–2.32
moščnyj	‘powerful’	*moščnée*	5	2.9%	165	97.1%	170	–4.01
mudryj	‘wise’	*mudrée*	1	1.6%	62	98.4%	63	–4.17
nadëžnyj	‘reliable’	*nadëžnee*	18	10.1%	161	89.9%	179	–3.66
nežnyj	‘gentle’	*nežnée*	1	2.2%	44	97.8%	45	–3.97
nužnyj	‘required’	*nužnée*	1	4.5%	21	95.5%	22	–3.77
opasnyj	‘dangerous’	*opásnee*	26	8.2%	291	91.8%	317	–3.85
ostorožnyj	‘careful’	*ostoróžnee*	1	12.5%	7	87.5%	8	–3.73
ostryj	‘sharp, spicy’	*ostrée*	5	10.4%	43	89.6%	48	–3.09
plotnyj	‘dense’	*plotnée*	2	3.0%	64	97.0%	66	–3.91
podrobnyj	‘detailed’	*podróbnee*	1	12.5%	7	87.5%	8	–3.73
poleznyj	‘useful’	*poléznee*	46	15.3%	254	84.7%	300	–3.19
polnyj	‘full’	*polnée*	0	0.0%	14	100.0%	14	–4.04
ponjatnyj	‘understandable’	*ponjátnee*	1	6.7%	14	93.3%	15	–3.92
populjarnyj	‘popular’	*populjárnee*	10	8.1%	113	91.9%	123	–3.86
pravil'nyj	‘right’	*právil'nee*	0	0.0%	13	100.0%	13	–4.41
predpočtitel'nyj	‘preferred’	*predpočtítel'nee*	66	18.1%	298	81.9%	364	–3.67
prekrasnyj	‘beautiful’	*prekrásnee*	50	22.4%	173	77.6%	223	–2.74
privlekatel'nyj	‘attractive’	*privlekátel'nee*	16	15.1%	90	84.9%	106	–3.88
prijatnyj	‘nice’	*prijátnee*	28	16.0%	147	84.0%	175	–3.16
pročnyj	‘lasting’	*pročnée*	8	2.2%	361	97.8%	369	–4.27
svetlyj	‘light-coloured’	*svetlée*	6	1.9%	304	98.1%	310	–4.24
svobodnyj	‘free’	*svobódnee*	3	27.3%	8	72.7%	11	–3.22
ser'ëznyj	‘serious’	*ser'éznee*	11	16.2%	57	83.8%	68	–3.24
sil'nyj	‘strong’	*sil'née*	71	3.2%	2,141	96.8%	2,212	–3.93
složnyj	‘difficult’	*složnée*	16	3.6%	430	96.4%	446	–3.89
smelyj	‘bold’	*smelée*	0	0.0%	8	100.0%	8	–3.95
spokojnyj	‘calm’	*spokójnee*	5	16.1%	26	83.9%	31	–3.37
strašnyj	‘scary’	*strášnee*	30	6.4%	438	93.6%	468	–3.23
strojnyj	‘slim’	*strojnée*	1	4.5%	21	95.5%	22	–3.73
sčastlivyj	‘happy’	*sčastlívee*	8	13.6%	51	86.4%	59	–3.43
tëmnyj	‘dark’	*temnée*	1	0.3%	369	99.7%	370	–5.07
tëplyj	‘warm’	*teplée*	3	1.5%	202	98.5%	205	–4.37
trudnyj	‘difficult’	*trudnée*	2	5.3%	36	94.7%	38	–3.72
tjažëlyj	‘heavy’	*tjaželée*	21	2.7%	757	97.3%	778	–4.02
udobnyj	‘comfortable’	*udóbnee*	23	8.6%	244	91.4%	267	–3.81
umnyj	‘smart’	*umnée*	11	2.6%	411	97.4%	422	–4.14
ujutnyj	‘cosy’	*ujútnee*	4	25.0%	12	75.0%	16	–3.15
xolodnyj	‘cold’	*xolodnée*	2	2.1%	93	97.9%	95	–4.10
celesoobraznyj	‘expedient’	*celesoobráznee*	0	0.0%	11	100.0%	11	–4.16
cennyj	‘valuable’	*cennée*	7	3.0%	225	97.0%	232	–3.95
èffektivnyj	‘effective’	*èffektívnee*	74	9.9%	674	90.1%	748	–3.66
jasnyj	‘clear’	*jasnée*	4	14.3%	24	85.7%	28	–2.90

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0952675726100426.

Acknowledgements. We would like to thank Adam Albright, Daniel Asherov, Canaan Breiss, Edward Flemming, Maria Kholodilova, Donca Steriade and the audiences at FASL 32 for helpful discussion. We are also indebted to the three anonymous reviewers as well as the editors of this special issue of *Phonology* for their insightful comments and suggestions. All errors are our own.

Competing interests. The authors declare no competing interests.

References

Alderete, John D. (1999). *Morphologically governed accent in Optimality Theory*. PhD dissertation, University of Massachusetts Amherst.

Avanesov, Ruben I. (1964). *Udarenie v sovremennom russkom jazyke [Modern Russian stress]*. New York: Macmillan.

Barnes, Jonathan (2007). Phonetics and phonology in Russian unstressed vowel reduction: a study in hyperarticulation. Ms, Boston University.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48.

Benko, Vladimir (2014). Aranea: yet another family of (comparable) Web corpora. In Petr Sojka, Aleš Horák, Ivan Kopeček & Karel Pala (eds.) *Text, Speech and Dialogue: 17th International Conference, TSD 2014*. Cham: Springer, 247–256.

Blanchet, F. Guillaume, Pierre Legendre & Daniel Borcard (2008). Forward selection of explanatory variables. *Ecology* **89**, 2623–2632.

Boguslavsky, Igor & Leonid Iomdin (2009). Semantics of attenuated comparatives in Russian. In David Beck, Kim Gerdes, Jasmina Milićević & Alain Polguère (eds.) *Proceedings of the fourth International Conference on Meaning-Text Theory*. Montreal: Observatoire de linguistique Sense-Texte, 65–76.

Bolla, Kálmán (1981). *A conspectus of Russian speech sounds*. Budapest: Académiai Kiadó.

Borik, Olga (2014). The argument structure of long and short form adjectives and participles in Russian. *Lingua* **149**, 139–165.

Breiss, Canaan & Bruce Hayes (2020). Phonological markedness effects in sentence formation. *Lg* **96**, 338–370.

Breiss, Canaan, Hironori Katsuda & Shigeto Kawahara (2026). Modelling frequency-conditioned paradigm uniformity in Japanese voiced velar nasalisation. *Phonology* **43**, article no. e12 (22 pp.).

Casali, Roderic F. (2011). Hiatus resolution. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume & Keren Rice (eds.) *The Blackwell companion to phonology*, volume 3. Oxford: Wiley-Blackwell, 1434–1460.

Chervonnaya, Darya Yu. (2017). Raspredelenie pokazatelej sravnitelnoj stepeni *-ej* i *-ee* v russkom jazyke. Ms, National Research University Higher School of Economics.

Chomsky, Noam (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Crosswhite, Katherine, John Alderete, Tim Beasley & Vita Markman (2003). Morphological effects on default stress in novel Russian words. *WCCFL* **22**, 151–164.

Es′kova, Natalia A. (2011). *Izbrannye raboty po rusistike*. Moscow: Jazyki slavjanskix kultur.

Fauconnier, Jean-Philippe (2015). What is the relationship between log linear model, MaxEnt model and logistic regression? *Quora* post. Available online at https://www.quora.com/What-is-the-relationship-between-Log-Linear-model-MaxEnt-model-and-Logistic-Regression.

Flemming, Edward (2021). Comparing MaxEnt and Noisy Harmonic Grammar. *Glossa* **6**, article no. 141 (42 pp.).

Garde, Paul (1980). *Grammaire russe, tome premier: phonologie – morphologie*. Paris: Institut d'études slaves.

Gasparov, Mikhail L. (1996). *A history of European versification*. Oxford: Clarendon Press.

Gordon, Matthew (2002). A factorial typology of quantity-insensitive stress. *NLLT* **20**, 491–552.

Gouskova, Maria & Kevin Roon (2013). Gradient clash, faithfulness, and sonority sequencing effects in Russian compound stress. *Laboratory Phonology* **4**, 383–434.

Grashchenkov, Pavel V. & Ekaterina A. Lyutikova (2017). O sintaksise komparativa v russkom jazyke. *Russkij jazyk v naučnom osveščenii* **33**, 116–141.

Gries, Stefan Th. (2021). (Generalized linear) mixed-effects modeling: a learner corpus example. *Language Learning* **71**, 757–798.

Hammond, Michael (2016). Input optimisation: phonology and morphology. *Phonology* **33**, 459–491.

Hartig, Florian (2022). DHARMa: residual diagnostics for hierarchical (multi-level / mixed) regression models. R package, version 0.4.6. https://CRAN.R-project.org/package=DHARMa.

Hayes, Bruce (1984). The phonology of rhythm in English. *LI* **15**, 33–74.

Hayes, Bruce (1995). *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *LI* **39**, 379–440.

Hayes, Bruce, Colin Wilson & Anne Shisko (2012). MaxEnt grammars for the metrics of Shakespeare and Milton. *Lg* **88**, 691–731.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426

Iosad, Pavel (2012). Vowel reduction in Russian: no phonetics in phonology. *JL* **48**, 521–571.
Isačenko, Alexander V. (1965). *Grammatičeskij stroj russkogo jazyka v sopostavlenii so slovackim*. 2nd edition. Bratislava: Vydavateľstvo Slovenskej akadémie vied.
Itkin, Ilya B. (2007). *Russkaja morfonologija*. Moscow: Gnozis.
Kniazev, Sergei V. (2006). *Struktura fonetičeskogo slova v russkom yazyke: sinxronija i diaxronija*. Moscow: MAKS-Press.
Kukhto, Anton (2024). On Russian comparative morphology, Nanosyntax, and pointers. *WCCFL* **39**, 584–591.
Kukhto, Anton & Alexander Piperski (2020). Lexical stress variation and rhythmic alternation in Russian: a pilot study. *Linguistic Variation* **20**, 33–55.
Lavitskaya, Yulia & Barış Kabak (2014). Phonological default in the lexical stress system of Russian: evidence from noun declension. *Lingua* **150**, 363–385.
Liberman, Mark & Alan Prince (1977). On stress and linguistic rhythm. *LI* **8**, 249–336.
Martínez-Paricio, Violeta & René Kager (2015). The binary-to-ternary rhythmic continuum in stress typology: layered feet and non-intervention constraints. *Phonology* **32**, 459–504.
McCarthy, John J. & Alan Prince (1993). Generalized alignment. In Geert Booij & Jaap van Marle (eds.) *Yearbook of morphology 1993*. Dordrecht: Springer, 79–153.
Melvold, Janis L. (1989). *Structure and stress in the phonology of Russian*. PhD dissertation, MIT.
Nespor, Marina & Irene Vogel (1989). On clashes and lapses. *Phonology* **6**, 69–116.
Piperski, Alexander (2017). *Opomnis', bog s toboju*: variativnost tipa *mnoj* / *mnoju* obuslovlena fonologičeski i grammatičeski. In Darya F. Mishchenko (ed.) *Četyrnadcataja konferencija po tipologii i grammatike dlja molodyx issledovatelej. Tezisy dokladov*. Saint Petersburg: Nestor-Istorija, 136–139.
Plungjan, Vladimir A., Tatjana I. Reznikova & Dimitrij V. Sičinava (2005). Nacionalnyj korpus russkogo jazyka: obščaja xarakteristika [The National Corpus of the Russian Language: general characteristics]. *Naučno-texničeskaja informacija [Scientific and Technical Information], Series 2* **3**, 9–13.
Prince, Alan S. (1983). Relating to the grid. *LI* **14**, 19–100.
Prince, Alan S. (2002). Arguing optimality. *University of Massachusetts Occasional Papers in Linguistics* **28**, 269–304.
R Core Team (2023). *R: a language and environment for statistical computing*. Computer software. https://www.R-project.org/.
Schlüter, Julia (2005). *Rhythmic grammar: the influence of rhythm on grammatical variation and change in English*. *Number 46 in Topics in English Linguistics*. Berlin: de Gruyter Mouton.
Schlüter, Julia (2015). Rhythmic influence on grammar: scope and limitations. In Vogel & van de Vijver (2015), 179–206.
Schlüter, Julia & Gabrielle Knappe (2018). Synonym selection as a strategy of stress clash avoidance. In Sebastian Hoffmann, Andrea Sand, Sabine Arndt-Lappe & Lisa Marie Dillman (eds.) *Corpora and lexis*. Leiden: Brill, 69–105.
Schmid, Helmut (1999). Improvements in part-of-speech tagging with an application to German. In Susan Armstrong, Kenneth Church, Pierre Isabelle, Sandra Manzi, Evelyne Tzoukermann & David Yarowsky (eds.) *Natural language processing using very large corpora*. Dordrecht: Springer, 13–26.
Schwarz, Gideon (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.
Shih, Stephanie S. (2017a). Constraint conjunction in weighted probabilistic grammar. *Phonology* **34**, 243–268.
Shih, Stephanie S. (2017b). Phonological influences in syntactic alternations. In Vera Gribanova & Stephanie S. Shih (eds.) *The morphosyntax–phonology connection*. Oxford: Oxford University Press, 223–252.
Shih, Stephanie S., Jason Grafmiller, Richard Futrell & Joan Bresnan (2015). Rhythm's role in predicting genitive alternation choice in spoken English. In Vogel & van de Vijver (2015), 207–234.
Stanton, Juliet (2019). Phonetic lapse in American English *-ative*. *Glossa* **4**, article no. 55 (37 pp.).
Stanton, Juliet (2021). Phonetic rhythm in American English *-ization*. Ms, New York University. Available online at https://ling.auf.net/lingbuzz/005815.
Sweet, Henry (1887). *A handbook of phonetics*. Oxford: Clarendon Press.
Vogel, Ralf & Ruben van de Vijver (eds.) (2015). *Rhythm in cognition and grammar: a Germanic perspective*. Berlin: de Gruyter Mouton.
Walker, Rachel (2000). Mongolian stress, licensing, and factorial typology. Ms, University of California, Santa Cruz. ROA #172.
Waskom, Michael L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software* **6**, article no. 3021 (4 pp.).
Yanushevskaya, Irena & Daniel Bunčić (2015). Russian. *Journal of the International Phonetic Association* **45**, 221–228.
Zalizniak, Andrei A. (1987). *Grammatičeskij slovar' russkogo jazyka*. Moscow: Russkij jazyk.
Zalizniak, Andrei A. (1992). Pravilo otpadenija konečnyx glasnyx v russkom jazyke. In André Clas (ed.) *Le mot, les mots, les bons mots. Word, words, witty words: hommage à Igor Melčuk à l'occasion de son soixantième anniversaire.* Montréal: Les presses de l'Université de Montréal, 295–303.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 11 Nov 2026 at 06:30:28, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0952675726100426