Suffix rivalry in Russian: what low frequency words tell us

Natalia Bobkova Fabio Montermini
CLLE-ERSS, CNRS & Université de Toulouse Jean Jaurès
natalia.bobkova@univ-tlse2.fr fabio.montermini@univ-tlse2.fr

1. Introduction

The derivation of (relational) adjectives from nouns constitutes a particularly complex issue in Russian derivational morphology. Such adjectives are in fact constructed by a large set of different affixes, many of which constitute variants (and particularly extensions) of three main suffixes, *-sk-*, *-n-*, and *-Ov-*.[1] The competition between the three main affixes, and their variants, is apparently driven by a complex combination of factors, including phonological, morphological and semantic ones. In this article, we present a work in progress on the rivalry in question, and namely a series of quantitative observations on data extracted from the *National corpus of Russian language*. The data on which our observations are made include both the highest frequent lexemes in the corpus and hapaxes (frequency 1 lexemes). We are in fact persuaded that very low-frequency lexemes, if observed on a large scale, are likely to be good indicators of the creative use of morphological constructions. The observations presented in this article concern almost exclusively phonological factors, and namely the length of the base lexeme, the final segment of the base stem, and the presence of morpho-phonological phenomena typical of Russian derivation (vowel/Ø alternation and consonant mutation). If these factors play indeed a role in the choice of a suffix or of one of its variants, unsurprisingly they turn out to be insufficient to account for the choices made by speakers in a deterministic way. A multifactorial analysis is thus a necessary further step in our research we are committed to realize. In the final part of the article we present some hints about a theoretical modelization of the rivalry in question. In particular, we consider that a non-deterministic and hierarchically organized model, such as Construction Morphology (CxM, Booij 2010), constitutes the best way to account the complex interactions these rival (but in many respects overlapping) strategies have. Our CxM model is combined with a strictly stem-based and constraint-based approach to morphology, similar to the one elaborated by Roché and Plénat (2014), among others, about French.

2. The problem: denominal adjectives in contemporary Russian

2.1 Relational vs. possessive adjectives

As in other Slavic languages, denominal adjectives constructed by derivation in Russian belong to two main classes, that we may globally label as relational (1a-b) and possessive (1c-d):[2]

[1] By this notation we indicate the variation of the vowel in this suffix, that may correspond, phonologically, to different surface forms, and orthographically to <o> or <e>.

[2] The typographical conventions adopted in this paper are the following: Russian lexemes are presented in transcription and in their citation form; derivational exponents (either in their "bare" or in their extended form, see below) are marked in bold; inflectional endings (when different from zero) are in brackets. When a difference

(1) a. *um* → *um**stvenn**(yj)* 'mind'/'related to the mind'
b. *um* → *um**n**(yj)* 'mind'/'smart'
c. *otec* → *otc**ov*** 'father'/'of the father'
d. *lis(a)* → *lis(ij)* 'fox'/'of the fox'

The adjectives we label as relational have mainly the function to express the relation between the referents of two nouns (the base noun and the head noun in a syntactic construction). The actual range of meanings they can express, however, is quite large. Moreover, they can undergo lexicalization phenomena reducing their transparency, and can acquire, eventually, the properties of qualitative adjectives, and an arbitrary semantic relation with their base, as in the case of *umn(yj)* in (1b). Possessive adjectives, on the other hand, have a quite straightforward semantic interpretation, roughly corresponding to a genitive construction and limited to human, and, to a lesser extent, animate nouns (cf. Corbett 1987). The two classes are not only distinguished semantically: whereas relational adjectives follow the "canonical" declension patterns encountered also with simple adjectives, possessive adjectives follow specific declension patterns. Finally, although some derivational exponents (e.g. *-Ov-*) are common to the two, they are globally derived through different exponents and different derivational strategies in general (for instance, conversion, as in (1d) is only used for the derivation of possessive adjectives).

2.2 Morphological rivalry in relational adjective formation

The focus of this paper is on relational adjectives, which, as observed, constitute a semantically and inflectionally homogeneous set. However, they also display a great deal of variation , especially in the range of exponents (and namely suffixes) employed. In this respect, we may consider that they constitute a good testing ground for the study of the competition between rival derivational strategies for the same syntactic and semantic function (cf. Lindsay & Aronoff 2013; Aronoff 2016; Bonami & Thuilier 2018, among others). Russian grammars such as Townsend (1975) or Švedova (1980), for instance, list up to 25 different exponents that are used in Russian to form denominal relational adjectives. What is peculiar about Russian (and other Slavic languages) is that most of these exponents correspond to variants of the same affix extended with additional phonological material. Accordingly, we call these variants "extensions" or "extended variants" in what follows, in order to distinguish them from the "bare" ones. In particular, three main suffixes are identified in the literature (cf. Zemskaja 2015; Hénault & Sakhno 2015; Kustova 2018) as being productive in synchrony, -SK-, -N- and -OV-.[3] As we will show later, our data confirm this fact, since 96% of the low-frequency lexemes in our database contain one of these three exponents. All the suffixes in question may appear in their bare form, as in (2):

(2) a. *begemot* → *begemot**ov**(yj)* 'hippopotamus'
b. *universitet* → *universitet**sk**(ij)* 'university'
c. *kanal* → *kanal'**n**(yj)* 'canal'

between a lexeme and its stem(s) is made, we note the lexeme in small capitals and the stem(s) in IPA, using a broad transcription, in the lines, for instance, of Yanushevskaya and Bunčić (2015), where only some major phonological phenomena (namely unstressed vowel reduction and consonantal devoicing/assimilation) are marked. In the examples that follow, where only the noun is translated, the adjective is implicitly taken as meaning, predictably, "related to X".

[3] For details about the representation of affixes and their variants in this article, see below.

Whereas this is the only option available for -OV-, -SK- and -N- may also appear with different extended variants, as the examples in (3) and (4) show.

(3) a. *cikl → cikl**ičesk**(ij)* 'cycle'
b. *korol' → korol**evsk**(ij)* 'king'
c. *kladbišč(e) → kladbišč**ensk**(ij)* 'cemetery'

(4) a. *simvol → simvol**ičn**(yj)* 'symbol'
b. *vin(a) → vin**ovn**(yj)* 'guilt'
c. *stipendi(ja) → stipendi**al'n**(yj)* 'scholarship'

In Table 1 we present a rough list of the extended variants attested in our database (see 2 below). A modelization for the relations between the different variants is provided in Figures 4 and 5 in 5.1.

Table 1: extended variants of adjectival suffixes encountered in the database

-SK-	-N-	-OV-
-sk-	*-n-*	*-Ov-*
-esk-	*-Ovn-*	
-česk-	*-ičn-*	
-ičesk-	*-ivn-*	
-ističesk-	*-on(n)-*	
-ijsk-	*-en(n)-*	
-ansk-	*-(e)stven(n)-*	
-ensk-	*-ozn-*	
-insk-	*-al'n-*	
-istsk-	*-onal'n-*	
-Ovsk-	*-arn-*	
	-in-	

As it is clear from the list above, extended variants may have different sources. Some of them, for instance, are formally similar to derivational suffixes which are available in other cases in Russian. Nevertheless, we consider that, at least in some cases, they should be considered as fully integrated in the exponent of the denominal adjectival construction, as the following examples show:

(5) a. *Ø/xudožestv(o) → xudož**estvenn**(yj)* 'art'
b. *um/Ø → um**stvenn**(yj)* 'mind'
c. *bog/božestv(o) → bož**estvenn**(yj)* 'god/deity'

The sequence *-estv-* corresponds to a derivational suffix forming abstract nouns in Russian. For (5a), however, it should be considered as inherited from the base, since *xudožestv(o)* does not possess a simpler lexeme to which it is connected (although it is an abstract noun); for (5b), on the other hand, it is certainly introduced by the derivational construction, as no abstract noun in *-estv-* constructed on the basis of *um* is attested. (5c), finally, is ambiguous: whereas it may be considered as constructed on the basis of *božestv(o)* formally, the semantic noun to which *bož**estvenn**(yj)* is related is the simple form *bog*, rather than the abstract noun. We will go back to such cases of multiple motivation below.

Other extended variants correspond to sequences that have been introduced in Russian in order to adapt adjectives borrowed from foreign languages containing suffixes of Greek or Latin origin (cf. Dubkova 2004). Consider, for instance, the following examples:

(6) a. *simvol* → *simvol**ičn**(yj)* 'symbol'
b. *ulic(a)* → *ul**ičn**(yj)* 'street'
c. *god* → *god**ičn**(yj)* 'year'

(7) a. *Kub(a)* → *kub**insk**(ij)* 'Cuba'
b. *Jalt(a)* → *jalt**insk**(ij)* 'Yalta'
c. *sestr(a)* → *sestr**insk**(ij)* 'sister'

In (6a) the sequence *-ič-* has clearly the function of adapting the adjectival suffix found in several languages and derived from *-ik(ós)* in Greek or *-ic(us)* in Latin. In (6b) the same sequence is the outcome of a recurrent morphophonological operation of final consonant mutation (see 4.3 below), whereas in (6c) it has apparently no specific function and is simply indissociable from the extended variant of the suffix as a whole. Similarly, the sequence *-in-* in (7a) is parallel to the suffix derived from Latin *-in(us)* found in several European languages to form ethnic nouns or adjectives;[4] however, it may also be found in ethnics constructed on native Russian toponyms (7b). Finally, concerning (7c), the sequence *-in-* is possibly linked to the homophonous suffix forming possessive adjectives from nouns of the first declension (nominative singular in *-a*).[5] What we claim is that a clear-cut distinction between suffixes of foreign origin vs. native Slavic suffixes is hard to establish, and that variants such as *-ičn-* or *-insk-* (along with some others) owe their diffusion to a convergence between elements of foreign origin and native ones.

2.2 A two-level suffixal rivalry

Affixal rivalry in the derivation of denominal adjectives may thus be observed in Russian from a double point of view, i.e. by dealing with the alternation between one of the three main suffixes, or with the alternation between specific extensions, either within the same main suffix or across them. Some examples are presented in (8), where denominal adjectives are constructed from the same base with three variants belonging, respectively, to the three different main suffixes (8a) or to the same one (8b).

(8) a. *ieroglif* → *ieroglif**ičesk**(ij)/ieroglif**n**(yj)/ieroglif**ov**(yj)* 'hieroglyph'
b. *okean* → *okean**sk**(ij)/okean**ovsk**(ij)/okean**ičesk**(ij)* 'ocean'

In this paper we mainly deal with the first type of alternation, i.e. between the three main variants. When useful, we will occasionally present more fine-grained observations on the alternation between extended variants. In this case, each variant will be presented, canonically, in italics in the following format: *-ičesk-*, *-ičn-*, etc. In the most global case, we use capitals in

[4] The form *kub**insk**(ij)* cannot be claimed to be *stricto sensu* an adaptation of a specific lexeme in a foreign language, since the ethnic noun/adjective denoting an inhabitant of Cuba is formed by means of the suffix derived from Latin *-an(us)* in most Romance, German and Slavic languages (cf. English *Cuban*, Polish *kubański*, etc.). An influence of the Latin *-in(us)*, however, may be seen in Russian in such (non-fully transparent) adjectives as *argent**insk**(ij)*, *benedikt**insk**(ij)*, *filipp**insk**(ij)* or in some non-standard forms such as *florent**insk**(ij)* or *triest**insk**(ij)*.
[5] This hypothesis is supported by the observation of relational adjectives derived from feminine names or other names of animate entities in *-a*, such as *Elisavet(a)* → *elisavet**insk**(ij)* or *Satan(a)* → *satan**insk**(ij)*.

order to refer to the three main variants, -SK-, -N- and -OV-. Among others, this allows to avoid ambiguity with their concrete appearances as "bare", non-extended, variants, *-sk-*, *-n-* and *-Ov-*. Thus, the examples in (8a) illustrate a case of rivalry between the three main variants, whereas the examples in (8b) illustrate a case of rivalry between three extended variants of the main suffix -SK-.

3. The study of Russian denominal adjectives: empirical issues

3.1 Data extraction

The study we present is based on the *National corpus of Russian language*,[6] a corpus of modern Russian containing over 600 million words. This corpus is automatically annotated with morphological tags and only a small part if it has been verified manually, which makes the data extremely noisy. In order to extract exploitable data from the corpus, we ran a script based on the possible endings for denominal adjectives listed in Table 1 above. In total, 78,113 adjectives were extracted. Due to the quantity of noise in the data, we had to perform a manual cleaning in order to eliminate false positives, which led to leave out 93% of the lemmas extracted automatically. Finally, all the adjectives formed with the main suffixes -SK-, N- and -OV- (which represent 96% of all the adjectives extracted) were further divided in two subsets, corresponding respectively to high-frequency (frequency >100) and to low-frequency lexemes (frequency =1), which allowed to obtain two sets quite homogeneous in size. Table 2 provides the detail of high- and low-frequency forms included in the corpus for each suffix.

Table 2: Data distribution for high- and low-frequency adjectives

	-SK-	-N-	-OV-	Total
Hfreq	833	741	716	2290
LFreq	1107	479	557	2161
				4451

The data thus extracted were coded according to several formal properties, namely their length, the final segment of the base stem, and whether the base lexeme is subject to morphophological phenomena, such as vowel/Ø alternation or consonantal mutation. Of course, this implies that we are able to unambiguously identify a base lexeme, which is far from trivial. Section 3.2 is devoted to this problem. Eventually, this coding will hopefully allow to correlate the formal properties of nouns to the choice of a particular suffix (or extended variant) for the formation of a relational adjective. Naturally, (morpho)phonological properties should not be considered as the unique factors determining the choice of one strategy over the other.[7] It is certain that, at least, semantic properties also play a role (cf. Hénault & Sakhno 2015). A more detailed annotation, which includes semantic (and possibly other) factors, is thus a necessary step in our research.

In what follows, we propose some preliminary observations, based in particular on the observation of the two subcorpora. Taking very low-frequency items (i.e. hapaxes, in this case) is justified by the idea that this set is more likely to contain lexemes which are constructed "on the spot" by speakers (even though, of course, not all hapaxes are nonce formations), and thus

[6] http://www.ruscorpora.ru/.

[7] The stress pattern, both of the base and of the derivative, could be another relevant formal parameter to consider. However, this factor is not merely phonological, as it interacts with inflectional class membership. We leave thus this issue open for further research.

be less influenced by the range of well-known phenomena that blur the transparency of lexical items, such as polysemy, lexicalization, loss of formal and/or semantic transparency, etc. Comparing this set with high-frequency lexemes may thus give hints on the functioning of derivational morphology independent of lexicalization and similar phenomena, especially when comparing large sets of data. Moreover, high-frequency lexemes may be considered to belong to the core lexicon, and consequently that their construction is older than for low-frequency ones. If this is true, comparing the two subcorpora may also allow to make (rough) observations on diachronic tendencies.

3.2 The identification of derivational bases and their stems

As observed above, the identification of the base noun of a derived adjective is not always obvious. In many cases a derived adjective may be motivated by different possible nouns, either formally, or semantically or both.[8] Such multiple motivation corresponds to a tendency that has been observed for several different languages (cf. Roché 2010; Booij & Audring 2018, among others). Table 3 presents some examples of derived adjectives displaying different possible bases. Consider, for instance, the forms listed in (i). The simplest analysis consists in interpreting the adjective *marks**istsk**(ij)* as constructed by adding the bare suffix *-sk-* to the noun *marksist.* In some cases this is consistent with the meaning of the adjective: *marksistskij kružok*, for instance, denotes a circle of Marxists. The same adjective, however, could also be considered to be constructed on the name *Marks*, to which the extended variant *-istsk-* is added (cf. *marksistskie sočinenija* 'Marx's writings'), or to *marksizm* via a suffix substitution (cf. *marksistskij termin* 'a term of Marxism').

The cases in (ii) display the same uncertainty, with the difference that in this case there is no underived noun which motivates the adjective. In (iii) we observe the reverse case, where an adjective can be considered to be motivated by two nouns neither of which bears an explicit affix.[9] The examples presented so far correspond to non-native lexical elements. Those in (iv) and (v) show that the same uncertainty holds for the native lexicon.

We can then conclude that the precise identification of a unique base for derived lexemes is not only impossible, but also unnecessary. Methodologically, for the present study, we chose to code the adjectives in our database with all the potential base nouns.

[8] Note, moreover, that we only concentrate here on cases in which the ambiguity between two possible bases concerns nouns. However, it may also be cross-categorial; cf. the following adjectives, in which the base might be either lexeme in a verb/(event or agent) noun pair: *vzryv(at')/vzryv → vzryv**n**(oj)* ('explode'/'explosion'/'explosive'); *nagrev(at')/nagrevatel' → nagrevatel'**n**(yj)* ('heat'/'heater'/'related to heating').

[9] Describing the relation between *fizik* and *fizik(a)* in this manner implies that we consider it to be an instance of conversion. Other interpretations are possible, however, namely that the former is derived from the latter by subtraction (cf. Dressler 1987 and, for a discussion, Manova 2011: 178-179).

Table 3: Derived adjectives and their relation to nouns

	noun (underived)	noun (underived)	noun (derived)	noun (derived)	adjective
(i)	*Marks* 'Marx'	-	*marksist* 'Marxist'	*marksizm* 'Marxism'	*marks**istsk**(iy)*
(ii)	-	-	*kommunist* 'communist'	*kommunizm* 'communism'	*kommun**ističesk**(ij)*
(iii)	*fizik* 'physicist'	*fizik(a)* 'physics'	-	-	*fizič**esk**(iy)*
(iv)	*knigoved* 'bibliologist'	-	*knigovedeni(e)* 'bibliology'	-	*knigoved**česk**(ij)*
(v)	*bog* 'god'	-	*božestv(o)* 'divinity'	-	*bož**estvenn**(yj)*

Another problem that arises for base annotation concerns the stem allomorphies observed for some nouns, and in particular the cases of alternation between Ø and a full vowel. This point is important, since we consider that the length of the base stem (syllables) is a potentially significant parameter. Vowel/Ø alternation is encountered in nouns, both in inflection and in derivation (cf. Sims 2017). The alternation in question cannot be ascribed to a syncronically active phonological process, and thus we consider that in this case the lexical representation of a lexeme includes two stems. Note that short (vowelless) and long (with the vowel surfacing) stems correspond to different portions of the inflectional paradigm. The latter appear, in fact, in those inflected forms which do not bear any explicit affix, i.e. genitive plural for nouns of the first declension (nominative singular in *-a*), and nominal singular for nouns of the second declension, as shown in Table 4.

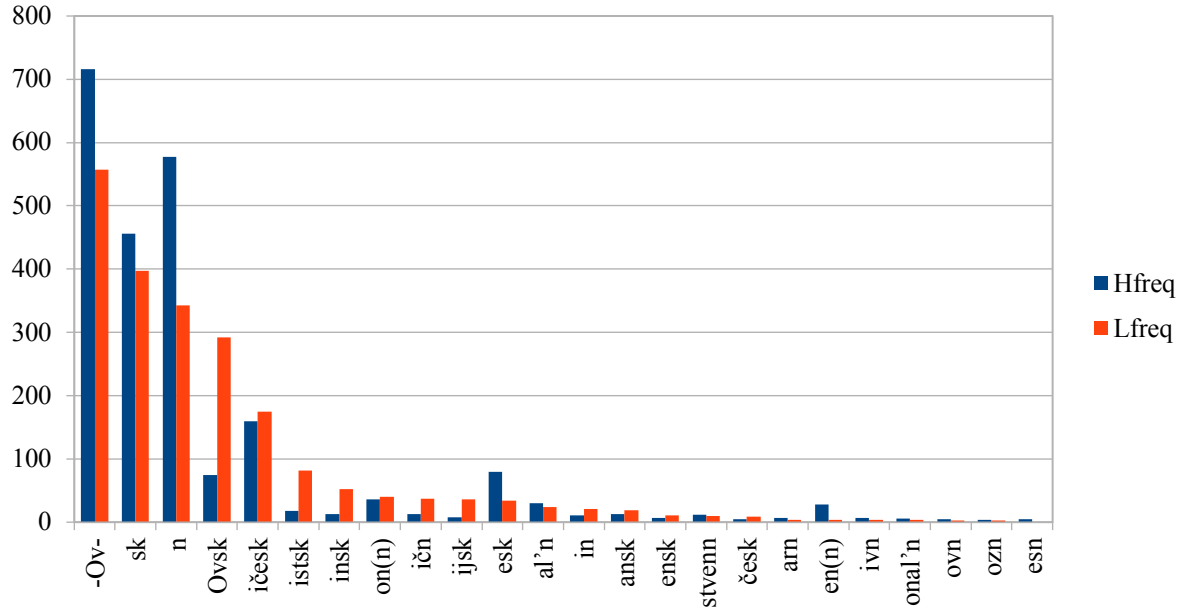
Table 4: Vowel/Ø alternation in inflectional stems and their distribution in the Russian nominal paradigms

			inflected forms		
	noun	declension	N.SG.	others	G.PL.
(i)	*jubka* 'skirt'	1st	short /jupk/		long /jubak/
(ii)	*koren'* 'root'	2nd	long /kor^{j}in^{j}/	short /korn^{j}/	

3.3 Distribution of the data

A quick look at the distribution of the suffixes in our database shows that the bare variants are the most numerous both for high- and low-frequency lexemes, and particularly for the former (see Figure 1). As observed above, the comparison between the two sets may provide a rough indication of the diachronic tendencies that are active for the constructions in question. In this respect, we observe a decrease of the availability of all bare variants (and in particular of *-Ov-* and *-n-*), along with an important increase of some extented variants (in particular *-ičesk-*, *-istsk-* and *-insk-*). It is not sure, however, that this corresponds to a real diachronic tendency. Rather, we suggest that the big difference in the number of adjectives in *-Ovsk-* between the two subsets is due to the fact that this variant is the preferred one to be attached to proper nouns of foreign origin, a class of units which is more susceptible to give rise to derivatives displaying a relatively low frequency (cf. also section 4.2).

Figure 1: Distribution of adjectives in High and Low frequencies



The data presented also allow to better assess some assumptions about affixal productivity made in the literature. For instance, contrary to the observations of Švedova (1980), according to whom *-n-* is the highest productive suffix, we observe an important decrease, and an increase of such extended variants as *-Ovsk-* and *-istsk-*.

4. Analysis: phonological factors

In this section we analyse how different phonological factors may influence the choice of the suffix. In particular, we focus on stem length, on stem form (especially on its final phoneme), and on the modifications they undergo (final velar mutations).

4.1 Stem length

In their study on *-ify* suffixation in English, Lindsay and Aronoff (2013) observe a correlation between the length of the base and the choice of this suffix, when compared to *-ize*. The same tendency is pointed out by Lignon (2013) and Bonami and Thuillier (2018) for *-iser* and *-ifier* in French, although these two works use different methods.

Table 5 shows the percentage of adjectives formed with each suffix according to the length of the base. A strong correlation between the length of the base stem and the choice of the suffix, and this correlation is the same for high and low frequency adjectives. Adjectives in -OV- are constructed mainly on monosyllabic stems, while for longer stems -SK- is preferred. The distribution is more uniform for -N-. The preference of monosyllabic stems for -OV- may be explained on the basis of the fact that this affix allows to add one syllable to the derivative, whereas -SK- and -N- (in their bare forms) form adjectives with the same number of syllables as their stems. Although it is not true for some variants (e.g. *-ičesk*, *-onal'n-* both add two syllables), at the level of main variants, we observe that almost half of monosyllabic adjectives are constructed with -OV- and the proportion drops significantly when the number of syllables increases.

Table 5: Distribution of the three main suffixes by stem length (%)

	LFreq			HFreq		
syll.	-SK-	-N-	-OV-	-SK-	-N-	-OV-
>5	72.73	27.27	0.00	85.71	14.29	0.00
5	71.59	27.27	1.14	62.34	33.77	3.90
4	71.17	24.91	3.91	68.33	27.50	4.17
3	63.79	21.88	14.32	59.51	25.35	15.14
2	50.98	23.64	25.38	36.83	39.62	23.55
1	36.98	15.97	47.05	16.79	24.96	58.25

4.2 Stem form

As far as the last phoneme of the stem is concerned, vowel-ending stems and –to a lesser extent– stems ending with a sonorant consonant have a strong preference for -SK- suffixation. As for the other affixes, we observe a relatively homogeneous distribution, i.e. stems ending in a nasal, stop, affricate or fricative consonant for high-frequency adjectives, whereas for low frequencies -SK- is the preferred choice in all cases.[10] Table 6 gives the details of the distribution in our corpus.

Table 6: Distribution of the three main suffixes by stem form (type of final segment) (%)

	LFreq			HFreq		
	-SK-	-N-	-OV-	-SK-	-N-	-OV-
V	87.93	8.62	3.45	90.91	9.09	0.00
Nasal	61.62	17.37	21.01	46.08	22.57	31.35
Stop	50.96	23.25	25.80	35.32	31.69	33.00
Affr/Fr	48.06	23.89	28.06	27.90	36.88	35.22
Son	58.84	23.47	17.69	50.95	30.14	18.92

Concerning vowel-ending stems, we observe an important difference (1:5) between high- and low-frequency adjectives. This is mostly due to proper nouns of foreign origin, which are indeclinable, and for which the vowel is thus integrated into the stem. The relevant figures are given in Table 7.

Table 7: Number of vowel ending stems

	-SK-	-N-	-OV-
HFreq	10	1	0
LFreq	51	5	2

In this case, we observe a great deal of variation in the strategies adopted for adjective formation (concatenation, integration within a suffixal variant, vowel deletion), as illustrated by the examples in (9).

(9) a. *Korbjuz'e → korbjuz'e**ansk**(ij)* 'le Corbusier'
b. *Fellini → fellini**evsk**(ij)* 'Fellini'
c. *Goa → go**ansk**(ij)* 'Goa'
d. *regbi → regb**ijn**(yj)* 'rugby'

[10] If we take extended variants into consideration, the tendencies we observe are similar to the ones given in Table 6.

e. *Mussolini → mussolin**ovsk**(ij)* 'Mussolini'
f. *retro → retr**insk**(ij)* 'retro'

4.3 Consonant alternations (mutation)

Another phenomenon that concerns stem endings is the semi-productive morpho-phonological alternation in place of articulation between palatal and non-palatal consonants. As vowel/Ø alternation, these modifications do not correspond to a synchronically productive phonological phenomenon; unlike the former, however, they only surface in derivation for nouns. To refer to the alternation in question, we use the term "mutations" (cf. Sims 2017). As for vowel/Ø alternation, we consider that in this case lexemes are stored in the lexicon with two stems, which accounts for the fact that some lexemes display the alternation while others do not. Since this modification is limited to derivation, we label the mutated stem "Stem$_D$". Potentially mutating consonants and their alternates are illustrated in Table 8. However, cases in which a final /x/ undergoes mutation (iv) are marginal; in what follows, we focus then on the first three types.

Table 8: Stem allomorphy in Russian nouns

	noun	stem A	mutation	Stem$_D$	adjective
(i)	*knig(a)* 'book'	/knʲig/	/g/ – /ʒ/	/knʲiʒ/	*knіž**n**(yj)*
(ii)	*ruk(a)* 'hand'	/ruk/	/k/ – /t͡ʃ/	/rut͡ʃ/	*ruč**n**(oj)*
(iii)	*pjatnic(a)* 'friday'	/pʲatʲnʲit͡s/	/t͡s/ – /t͡ʃ/	/pʲætʲnʲit͡ʃ/	*pjatnič**n**(yj)*
(iv)	*ux(o)* 'ear'	/ux/	/x/ – /ʃ/	/uʃ/	*uš**n**(oj)*

The presence of a mutated stem is thus lexically determined, and also dependent on the derivational affix (Kapatsinski 2010; Sims 2017). For example, the bare suffix *-n-* disfavors velar-final stems, and almost systematically triggers mutation. In this respect, the choice of an extended variant may be a way of avoiding mutation (which introduces opacity in base-derivative relation) (cf. Zemskaja 2015). We discuss this point in detail below.

Table 9 shows the distribution of mutated stems with different suffixes. In low-frequency adjectives we observe a clear prevalence of *-n-* suffixation, whereas in the high-frequency ones *-sk-* is dominant for lexemes displaying the /k/ – /t͡ʃ/alternation.

Table 9: Distribution of the three main suffixes by mutating stems (%)

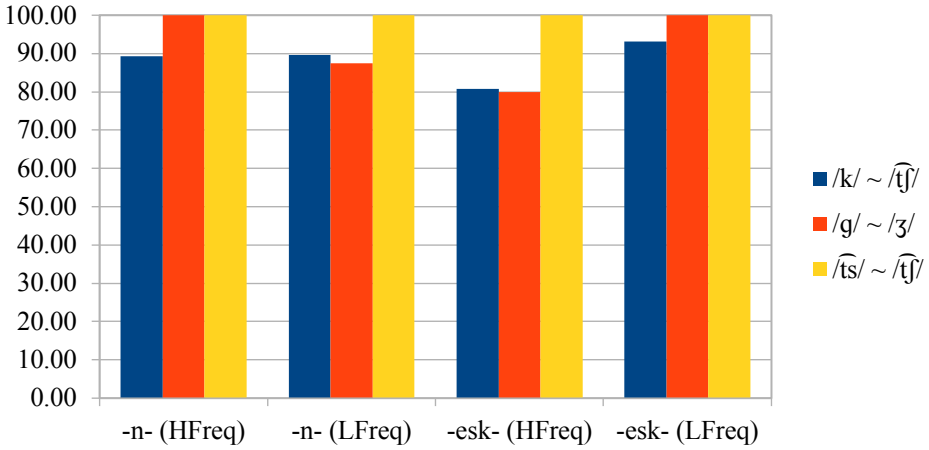
	LFreq			HFreq		
	-SK-	-N-	-OV-	-SK-	-N-	-OV-
/g/ ~ /ʒ/	46.15	53.85	0.00	31.03	68.97	0.00
/k/ ~ /t͡ʃ/	30.43	66.30	3.26	60.38	39.62	0.00
/t͡s/ ~ /t͡ʃ/	25.00	75.00	0.00	35.71	64.29	0.00

To go further into details, we observe that among all the extended variants of -SK-, *-esk-* triggers most of the mutations (78.95%). On the other hand, the bare *-sk-* variant shows variation in triggering or not mutation (cf. (10)), whereas among the variants of -N- it is the bare variant *-n-* that triggers mutation most of the time (96.00%).

(10) a. *Sankt-Peterburg – sanktpeterbur**žsk**(ij)* 'Saint-Petersbourg'
b. *arxipelag – arxipelag**sk**(ij)* 'archipelago'

We then calculated the mutation ratio for these two variants and for each potentially mutating consonant. Figure 2 shows some interesting tendencies: in particular, *-n-* displays a higher rate of mutation in high-frequency adjectives, whereas the tendency is inverted for *-esk-*.

Figure 2: Mutation ratio for potentially mutating stems triggered by *-n-* and *-esk-*



Nevertheless, these cases are much more complicated than just a single consonant mutation. Consider the examples in Table 10. Example (i) shows the derivation of adjectives by means of *-n-* suffixation. However, there is one extra operation which intervenes here, the truncation of the last consonant of the stem (/k/). Mutation does not take place because there is no more potentially mutating consonant. Examples (ii) and (iii) illustrate the behaviour of the suffix *-esk-*. (ii) illustrates a case of /k/ ~ /ʃ/ mutation, which cannot be considered as productive in contemporary Russian. Examples (iiia-b) are the least obvious. The adjective *pedagog**ičesk**(ij)*, in fact, is motivated either with respect to *pedagogik(a)* ('pedagogy') or to *pedagog* ('pedagogue') (cf. *pedagogičeskaja nauka* 'pedagogic science' vs. *pedagogičeskie kadry* 'teaching personnel'). If the base is *pedagogik(a)*, we face indeed a case of mutation; if the base is *pedagog* we should consider that the extended variant *-ičesk-* is added.

Table 10: Mutation and mutation avoidance in potentially mutating derivatives

	noun	stem1	operation(s)	suffix	adjective
(i)	*kolgotk(i)* 'tights'	/kalgotk/	subtr. of /k/	*-n-*	*kolgot**n**(yj)*
(ii)	*mal'čik* 'boy'	/malʲt͡ʃɪk/	/k/ ~ /ʃ/	*-esk-*	*malčiš**esk**(ij)*
(iiia)	*pedagogik(a)* 'pedagogy'	/pʲidagogʲik/	/k/ ~ /t͡ʃ/	*-esk-*	*pedagog**ičesk**(ij)*
(iiib)	*pedagog* 'pedagogue'	/pʲidagok/		*-ičesk-*	*pedagog**ičesk**(ij)*

The examples above show that in the cases where mutation does not take place with the suffixes *-n-* and *-esk-* other factors may be invoked as an explanation. We conclude that mutation is thus obligatory with the *-n-* and *-esk-* variants. The relations between a derived adjective, the lexemes in its morphological family and the possible extended variants deserve more analysis, that we leave for future work.

5. Theoretical framework and modelisation

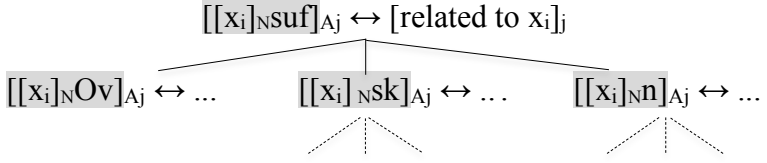
5.1 A Construction-Morphology model

As shown above, although some tendencies are clearly visible (as in the case of stem length), globally the phonological properties of base nouns alone do not allow to clearly discriminate between the different suffixes and variants. On the other hand, morpho-phonological properties, such as mutation, seem to work as better predictors. This suggests that a global lexical approach in which phonological constraints interact with morphological and lexical constraints is fitter for dealing with the competition between rival affixes in the formation of Russian denominal adjectives. Namely, the analysis we propose integrates a Construction Morphology (CxM) kind of representation (cf. Booij 2010) with a number of other approaches, which have been elaborated partially independently, but are compatible, and namely:

(i) a constraint-based approach, according to which the final form of a derivative is determined by the interaction of different, possibly conflicting, constraints (Hathout 2009; Roché & Plénat 2014);
(ii) a lexicalist approach, according to which canonical phonological constraints interact with more global lexical constraints that guarantee lexical homogeneity. The notions of morphological family (Bonami & Strnadová 2018; Schalchli & Boyé 2018) and of morphological series play a major role in this respect;
(iii) a strictly stem-based approach, in which formally morphological schemes do not simply connect a base with a derivative, but connect two stem spaces, i.e. organized sets of stems (Bonami *et al.* 2009);
(iv) an output-oriented approach, according to which derivational schemas consist in aligning a base stem to a specific template compatible with a series (Montermini 2018).

As CxM, the model we advocate is strictly lexeme-based, declarative (i.e. non-derivational and non-oriented) and hierarchical. For their part, constraints allow modelling morphology as an abstractive system (vs. a constructive one, cf. Blevins 2006), a feature which is particularly compatible with a CxM representation. Finally, both CxM and constraints are compatible with a variationist view of morphological creativity. In particular, here we focus on the formal (phonological) aspects of derivation. In this respect, our work constitutes an attempt to model the formal representation attached to morphological constructions, i.e. the one highlighted in grey in Figure 3, in which we propose a first, partial, formalization of the denominal adjectives constructions in Russian adopting a CxM formalism. Concerning semantics, which lies outside the scope of the present paper, we just give a very rough, informal, representation. To simplify the representation, the semantic portion is represented by ellipses in individual constructions. We thus consider, that the three main suffixal constructions we have identified for denominal adjectives in Russian are hierarchically linked to a superordinate, more abstract, construction, from which they inherit their categorial and semantic properties.

Figure 3: Hierarchy of denominal adjective constructions in Russian



In their turn, each individual construction (except the one formally marked by *-Ov-*) is hierarchically linked to a set of subordinate constructions, which are phonologically more

specific, and correspond to different extended variants. The hierarchies in question are represented, respectively, in Figures 4 and 5.

Figure 4: Construction hierarchy for -SK- suffixation

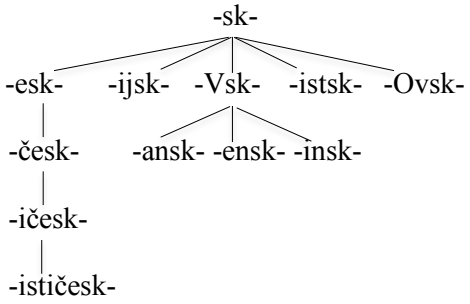
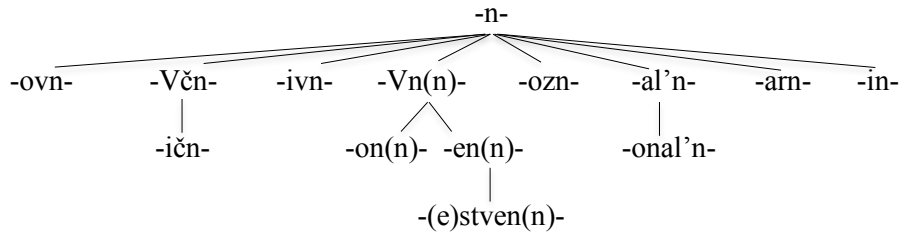


Figure 5: Construction hierarchy for -N- suffixation



5.2 Stem selection

The model we propose is lexeme-based, in the sense that it takes lexemes (words) as the atoms of morphological processes. Lexemes are complex and structured entities, whose lexical representation includes formal (phonological) information corresponding to a stem, or a set of stems. In the simplest cases, a lexeme is represented in the lexicon with a unique stem, but it may also comprise multiple stems, each having a specific (inflectional or derivational) function. From the point of view of derivational morphology, stems constitute both the formal bases of constructions, i.e. phonological objects on which the operation connected with a derivational construction is performed (e.g. the suffixation of *-sk-*, *-Ov-* or *-n-*), and their formal outputs. In (11) we represent the formal part of the derivation of *begemot**ov**(yj)* from *begemot* ('hippopotamus'), where the input and the output lexeme both include only one stem.

(11)	Lexeme		BEGEMOT		BEGEMOTOVYJ
	Stem /bʲigʲimot/	→	-Ov- suffixation	→	/bʲigʲimotav/

Of course, not all cases are as simple as the one exemplified, since lexemes may display various allomorphies, as described in sections 3.2 and 4.3. In these cases, all allomorphs which cannot be considered as derived by active regular phonological rules are listed in the phonological representation of a lexeme. The formal representation of a lexeme, thus, contains a set of stems (a structure that has been called "stem space", e.g. in Bonami & Boyé 2003, and subsequent works). The function of a morphological construction is thus potentially to connect the stem space of the base with the stem space of the derivative (Bonami *et al.* 2009). In (12) we present the quite simple case of the derivation of *ugl**ov**(oj)* from *ugol* ('angle'), a noun displaying vowel/Ø alternation (cf. 3.2).

(12)	Lexeme	UGOL		UGLOVOJ
	Stem A	/ugal/		
	Stem B	/ugl/	→ -Ov- suffixation	→ /uglav/

In the case of *ugol*, Stem A corresponds to the long stem, and Stem B to the short one (cf. Table 4 for their distribution in inflection). It is also the stem which is selected for the construction of an adjective, in this case via the suffixation of *-Ov-*. In cases like this one, in its formal part a morphological construction should contain not only an instruction on the operation to be performed, but also an indication of the stem to be selected.

A slightly more complex case is exemplified in (13-14). Here, we illustrate the derivation of the adjective *komponovo**čn**(yj)* from the action noun *komponovk(a)* 'assembly', which is itself derived from the verb *komponov(at')* ('assemble'). (13) illustrates the first derivational step (from verb to action noun), and (14) illustrates the second step (from noun to adjective):

(13)[11]

Lexeme KOMPONOVAT'				KOMPONOVKA
Stem A /kampanav/	→	-k- suffixation	→	Stem A' /kampanovk/
	→	-Ok- suffixation	→	Stem B' /kampanovak/
	→	-Oč- suffixation	→	Stem_{D} /kampanovat͡ʃ/
Stem B /kampanu/				

(14)

Lexeme KOMPONOVKA				KOMPONOVOČNYJ
Stem A' /kampanovk/				
Stem B' /kampanovak /				
Stem C' /kampanovat͡ʃ/	→	-n- suffixation	→	Stem A" /kampanovat͡ʃn/

In the first case (13), both the base and the derivative contain complex stem spaces in their phonological representation. The verb *komponov(at')* has two stems: in inflection, Stem A is used, among others, for the construction of the infinitive and of the past tense (cf. *kompono-val*$_{\text{PAST.M.}}$), Stem B for the present indicative (cf. *komponuju*$_{\text{1.SG.PRES.IND.}}$). As far as the noun *komponovk(a)* is concerned, it contains one stem (Stem A') which is used for all the inflected forms except for the plural genitive, where the Stem B' is used. A third Stem (Stem C') is not used for inflected forms and only serves for derivation, as in the formation of the adjective *komponovočn(yj)*. We consider that these three stems are simultaneously formed by a morphological construction that has a complex formal representation, namely the deverbal suffix traditionally represented as *-k(a)*. Note that the identification of three different exponents is justified by the fact that the lexemes formed by this suffix display either vowel/Ø alternation and/or mutation in different contexts. As claimed above, in this case, as in the case of *ugol*, the variation between the stems cannot be accounted for phonologically, and they are thus declaratively listed in the stem space of the lexeme. Thus, the morphological construction specifies a base stem to be selected, and also a set of operations to be performed in order to obtain each stem in the derivative's stem space. In such a picture, stem selection constitutes a fundamental feature in the formal representation of a morphological construction, along with its phonological specification (the exponent). In addition to the selection of a stem from a pre-established set, the formal part of a construction may include more or less regular stem modification processes, which may be phonologically-driven, or determined by other factors (like in the case of *komponovk(a)*). Other examples, from our low-frequency database, are given in (15).

[11] The vowel alternation between /o/ and /a/observed in derivation is due to the phonological phenomenon of vowel reduction known as *akan'e*, and is not relevant here.

(15) a. *xajtek* → *xajte**čn**(yj)* ‘high-tech’
b. *piknik* → *pikni**čn**(yj)* ‘picnic’

These examples show that the final velar of a borrowed lexeme may undergo mutation as well. If, on the one side, it is difficult to claim that such lexemes as *xajtek* or *piknik* possess a complex stem space which includes a mutated stem, on the other side the modification in question clearly cannot be attributed to phonological factors (/kn/ is an admitted sequence in Russian). Rather, it should be considered as purely morphological, since, as observed above (section 4.3.) mutation appears to be obligatory for velar-final bases with the suffix *-n-*. If –by default– a stem ending in /k/ undergoes mutation, the emergence of this morpho-phonological modification may be overriden by other factors. Consider the examples in (16).

(16) a. *Gosbank* → *gosbank**ovsk**(ij)* ‘Gosbank [name of a bank]’
b. *Kubrik* → *kubrik**ovsk**(ij)* ‘Kubrick’
c. *Piknik* → *piknik**ovsk**(ij)* ‘Picnick [name of a rock band]’

These examples illustrate a case in which the morpho-phonological constraint which imposes the surfacing of a mutated stem with a deadjectival suffix is dominated by two stronger constraints. The first one is morpho-semantic: proper names (especially foreign ones) strongly favor *-Ovsk-* as a suffixal variant. The second one is a faithfulness constraint, that favors non alternating stems.[12]

6. Conclusions

In this article we presented the first results of a corpus-based investigation of denominal adjectives in contemporary Russian having a relational function, a type of derivation implying a wide range of different exponents, and thus representing a particularly interesting case study of morphological competition. As we have seen, three main suffixes may be considered as productive in Russian for the construction of denominal adjectives, -SK-, -N- and -OV-. While the latter only appears in derived lexemes under its bare form (*-Ov-*), the first two appear in a number of different variants, all of which add phonological material to the bare forms *-sk-* and *-n-*. The competition between all these forms was observed on data obtained from the Russian (National corpus of Russian language). In particular, we compared the properties of high-frequency denominal adjectives with those of very low-frequency ones (hapaxes). The latter may in fact be considered as more representative of the creative use of morphological processes, since they are less likely to have undergone lexicalization and thus to be formally and semantically opaque.

In our analysis we claimed that a Construction Morphology representation is particularly fit for the treatment of these data. It allows, in fact, to establish hierarchical links between general, underspecified constructions and constructions that are more specified as to their phonological representation. The selection of a specific construction is determined by a complex interaction of constraints. The observations we proposed show that, at least in some cases, phonological factors, such as base length, the final segment of the base and the presence of potential allomorphies in the base (namely vowel/Ø alternation and/or consonant mutation) are predictive of the specific variant selected in the derivation. However, a more thorough observation that

[12] A Web search (June 2019) showed that for the adjective constructed on the common noun *piknik* both variants (*pikni**čn**(yj)*/*piknik**ovsk**(ij)*) are attested, only the second is attested to express the meaning ‘referring to the rock band Piknik’.

takes into account a larger set of factors, including morphological (class membership, for instance) and semantic ones, will be necessary in order to draw a complete picture of the competition in question.

References

Aronoff, M. 2016. Competition and the lexicon. In A. Elia, C. Iacobini & M. Voghera (Eds), *Livelli di Analisi e fenomeni di interfaccia. Atti del XLVII congresso internazionale della Società di linguistica Italiana*. Roma: Bulzoni, 39-52.

Blevins, J. P. 2006. Word-based morphology. *Journal of Linguistics* 42(3): 531-573.

Bonami, O. & G. Boyé. 2003. Supplétion et classes flexionnelles dans la conjugaison du français. *Langages* 152 : 102-126.

Bonami, O., Boyé G. & F. Kerleroux. 2009. L'allomorphie radicale et la relation dérivation-construction. In B. Fradin, F. Kerleroux & M. Plénat (Eds), *Aperçus de morphologie du français*. Saint-Denis: Presses Universitaires de Vincennes, 103-125.

Bonami, O. & J. Strnadová. 2018. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2): 167-197.

Bonami, O. & J. Thuilier. 2018. A statistical approach to rivalry in lexeme formation: French *-iser* and *-ifier*. *Word Structure* 11(2): 4-41.

Booij, G. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Booij, G. & J. Audring. 2018. Partial motivation, multiple motivation: The role of output schemas in morphology. In G. Booij (Ed.), *The Construction of Words. Advances in Construction Morphology*. Berlin: Springer, 59-80.

Corbett, G. 1987. The morphology/syntax interface: Evidence from possessive adjectives in Slavonic. *Language* 63(2): 299-345.

Dressler, W.U. 1987. Subtraction in a polycentristic theory of Natural Morphology. In E. Gussmann (Ed.), *Rules and the Lexicon*. Lublin: Redakcja Wydawnictw Katolickiege Uniwersytetu Lubielskego, 67-78.

Dubkova, O. 2004. *Прилагательные с заимствованной основой в русском языке*. PhD. Thesis. MPGU.

Hathout, N. 2009. *Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie*, Habilitation à diriger des recherches. Université de Toulouse II – le Mirail.

Hénault, C. & S. Sakhno. 2015. Чем *супермаркет-н-ый* лучше *супермаркет-ск-ого*? Словообразовательная синонимия в русских адъективных неологизмах по данным Интернета. In B. Tošovic & A. Wonisch (Eds), *Wortbildung und Internet*. Graz: Institut für Slavistik.

Kapatsinski, V. 2010. Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology* 1(2): 361-393.

Kustova, G. 2018. Прилагательные. In *Материалы к корпусной грамматике русского языка. Вып. 3. Части речи и лексико-грамматические классы*. Sankt Peterburg: Nestor-Istorija, 40-107.

Lignon, S. 2013. *-iser* and *-ifier* suffixation in French: Verifying data to 'verize' hypotheses. In N. Hathout, F. Montermini & J. Tseng (Eds), *Morphology in Toulouse. Selected prodeedings of Décembrettes 7*. Munich: Lincom Europa, 119-132.

Lindsay, M. & M. Aronoff. 2013. Natural selection in self-organizing morphological systems. In N. Hathout, F. Montermini & J. Tseng (Eds), *Morphology in Toulouse. Selected prodeedings of Décembrettes 7*. Munich: Lincom Europa, 133-153.

Manova, S. 2011. *Understanding Morphological Rules. With special empashis on conversion and subtraction in Bulgarian, Russian and Serbo-Croatian*. Berlin: Springer.

Montermini, F. 2018. Les affixes dérivationnels ont-ils des allomorphes ? pour une modélisation de la variation des exposants dans une morphologie à contraintes. In O. Bonami, G. Boyé, G. Dal, H. Giraudo & F. Namer (Eds), *The lexeme in theoretical and descriptive morphology*. Berlin: Language Science Press, 423-465.

Roché, M. 2010. Base, thème, radical. *Recherches linguistiques de Vincennes* 39: 95-134.

Roché, M. & M. Plénat. 2014. Le jeu des contraintes dans la sélection du thème présuffixal. In F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberger, J. Meinschäfer and S. Prévost (Eds), *Actes du 4e Congrès Mondial de Linguistique Française*. Paris: Institut de Linguistique Française, 1863-1878.

Schalchli, G. & G. Boyé. 2018. Paradigms and syncretism in derivation: The case of ethnics in French. *Lingue e linguaggio*, XVII(2): 197-215.

Sims, A. 2017 Slavic morphology: Recent approaches to classic problems, illustrated with Russian. *Journal of Slavic Linguistics* 25(2): 489-542.

Townsend, C. 1975. *Russian Word-formation*. Columbus (OH): Slavica Publishers.

Švedova, N. 1980. *Русская грамматика, volume 1*. Moskva: Izdatel'stvo Nauka.

Yanushevskaya, I. & D. Bunčić. 2015. Russian. *Journal of the International Phonetic Association* 45(2): 221-228.

Zemskaja, E. 2015. *Язык как деятельность. Морфема, слово, речь*. Moskva: Flinta.