Ján Mačutek, Emmerich Kelih

Free or not so free? On stress position in Russian, Slovene, and Ukrainian

Abstract: Stress position in three languages with free stress is scrutinized. Although there are no deterministic rules for stress in such languages, some statistical tendencies are clearly observable. Simple mathematical models for the mean stress position and for the relative mean stress position are suggested. As a "by-product," we show that word length distributions differ among different parts of speech.

Keywords: free stress, stress position, word length, Slavic languages

1 Introduction

This contribution provides some quantitative insight into the stress position in three Slavic languages (Russian, Slovene, and Ukrainian). Stress in these languages can be placed on any syllable and is therefore often described as free or unpredictable (Hyman 1977). While such a statement is true in the sense that there are no deterministic rules for stress position, the stress position displays some statistical tendencies.

In the first step, we summarize some results from older quantitative stress studies in Russian (cf. the overview by Kempgen 1995: 32–35, who re-analyzed data by Moiseev 1976). The data indicate that stress in Russian occurs predominantly in the second half of words and that it gravitates towards the middle of the word. Thus, based on these findings, a tentative assumption about a general "tendency towards the centre" can be made. In the second step, we will present new results from two other Slavic languages with free stress, namely Slovene and Ukrainian. In particular, we will present data on the stress position (also distinguishing different parts of speech). Some (preliminary) mathematical models, which predict the stress position as a function of word length, will be developed.

Acknowledgment: J. Mačutek was supported by the grant VEGA 2/0096/21.

Ján Mačutek, Mathematical Institute, Slovak Academy of Sciences and Constantine the Philosopher University in Nitra, e-mail: jmacutek@yahoo.com
Emmerich Kelih, University of Vienna, e-mail: emmerich.kelih@univie.ac.at

https://doi.org/10.1515/9783110763560-008

Finally, our findings allow us to formulate some preliminary rules for free stress (for the time being valid for the analyzed Slavic languages): it is free, but not entirely random (the distribution of stressed syllables is not uniform, but certain medium positions are preferred), and it is predictable (albeit not deterministically, but only stochastically).

2 Linguistic stress

Stress is an important suprasegmental feature of sounds, syllables, words, multiword expressions, and sentences. Hence there are multiple possibilities of a quantitative analysis of stress and stress position, a domain which is notoriously less researched in quantitative linguistics. Therefore, our contribution is focused on word accent in Slavic languages only, where already some quantitative studies and data are available. Word stress is usually related to a stressed syllable, being characterised by pitch change, greater duration, and a greater intensity of a vowel (Clark & Yallop 1995: 328ff).

In phonetics, different kinds of accents are distinguished, e.g., a dynamic one, usually named as stress, and a "musical" pitch accent. Another distinction concerns the position of the accent within words, which can be either fixed or free. Fixed stress is positionally restricted, as, for instance, in Czech, Latvian, Finnish, Hungarian, etc. A more fine-grained distinction can be made when one distinguishes whether stress is, e.g., on the first (initial) syllable, the last (final) syllable, or the penultimate syllable. Hyman (1977) and Gordon (2002) provide empirical data on stress position in the languages of the world. Based on Hyman's (1977) analysis of over 300 languages, it appears that approximately 37% of them have initial stress, 31% final stress, 25% penultimate stress, and the remaining few have either peninitial or antepenultimate stress. According to Hyman (1977), only these five stress systems are attested in world fixed stress languages.

In languages with free stress (like e.g., Russian, Bulgarian, Spanish, Italian, English, etc.), the freedom of stress is utilized for marking differences in the meaning (e.g., in German *umf**a**hren* 'to drive round' vs. ***u**mfahren* 'to knock someone over') or for marking different parts of speech (e.g., in English the noun *pr**o**gress* vs. the verb *progr**e**ss*). Free-stress languages seem to be less present, since according to the analysis by Hyman (1977), 138 out of 444 languages (31%) are of this type. Stress in these languages can occur practically anywhere; however, from a morphological perspective, particular rules and tendencies of placement of the stress can be obtained as, for instance, a tendency towards stress placement on the morphological root or on prefixes and suffixes.

From a statistical point of view, one has to ask about stochastic tendencies of the stress position, i.e., are there some preferred patterns to be found in the stress placement? It seems unlikely that the stress placement in so-called free stress languages is indeed absolutely random[1] (in the sense of the uniform distribution, i.e., with each syllable being equally probable to carry stress).

3 Stress in Slavic languages

Slavic languages form an important group of the Indo-European language family. Members of this group comprise both fixed stress and free stress languages. Regarding stress, there is an interesting internal differentiation within Slavic languages (cf. Krüger 2009). On the one hand, all contemporary West Slavic languages have fixed stress (Polish has penultimate stress, while Czech, Slovak, and Sorbian have initial stress). On the other hand, within the South Slavic languages, only Macedonian is characterized to have a fixed stress system (antepenultimate; cf. however, Kempgen 2008, who observes a switch to the initial stress placement). All other South Slavic languages (Bosnian, Bulgarian, Croatian, Serbian, and Slovene) have free stress. The same holds for all contemporary East Slavic languages (Belarusian, Russian, and Ukrainian).

As far it is known from historical and comparative linguistics, Proto-Slavic was a language with free stress (cf. Baerman 1999). Thus, seen from a historical perspective, the introduction of fixed stress systems seems to be an innovation among Slavic languages.

4 Language material and data

4.1 Russian

The Russian stress system has been analysed in much detail (cf. Lehfeldt 2012), also by means of quantitative analyses. Regarding its quantitative properties, in

1 In some languages, there are several linguistics factors which seem to determine the stress placement, even if the language is considered to have "free" stress position. Therefore, languages with a predictable free stress are sometimes distinguished. Phonotactic restrictions (syllable weight, long vs. short vowels in case of quantity-sensitive languages), parts of speech, conjugation and declination patterns, intonation, and pragmatic-contextual realizations of speech are, among others, discussed as relevant factors.

particular, Kempgen (1995: 32–35) has to be mentioned due to his summary of the most important empirical observations of the accent position. In Russian, as can be expected for a free stress language, practically any syllable of the word can be stressed, but every single word form is characterized by one particular stress position (there are, of course, a few examples where a variation in stress can be observed). However, it has been recognized in the past by many researchers, as reported by Kempgen (1995: 32), that there is an obvious tendency towards the centre, i.e., towards the syllable in the middle of a particular word.

This observation has been further specified in two aspects. First, stress occurs but rarely in the peripheries of words (i.e. in the initial and in the final position), which is a natural consequence of the preference towards the centre. Next, the second half of the word is clearly preferred. What this means in the empirical dimension can be seen in Tab. 1, where data on stress positions in 49,483 Russian word forms taken from an orthoepic Russian dictionary are presented (cf. Moiseev 1976 for the original counts and further details).

Tab. 1: Stress position in Russian (WL – word length, SP – stress position) according to Kempgen (1995: 34), based on Moiseev (1976).

SP	1	2	3	4	5	6	7	8
WL								
2	3834	5183						
3	3138	8224	6189					
4	447	5241	6139	1440				
5	50	340	3685	1970	324			
6	2	59	332	1592	388	55		
7		1	24	199	405	66	6	
8			1	7	50	63	3	
9					3	7	7	
10					1	1	2	4
11								1

A closer look at the data reveals a remarkable mechanism in stress placement in Russian. As already pointed out by Kempgen (1995: 33), one can easily observe that the peak of every single frequency distribution systematically moves with the increasing word length from the left edge to the right edge of the word form. This relation between word length and stress position can be, according to Kempgen (1995: 33), represented by a simple linear model, where the average position of stress can be predicted based on word length – the longer the word form is, the more to the right (counted from the beginning of word forms) stress

tends to be placed. In any case, the data give strong evidence that the stress position in Russian is a statistical mass phenomenon, which is predictable based on word form length.

4.2 Slovene

Slovene is a representative of South Slavic languages. For Slavic linguistics, Slovene is of particular interest because, in the contemporary language, both a pitch accent and a stress accent can be observed. This fact reflects a rather complex situation in the various dialects of the language. Moreover, contrary to Russian, Slovene is also quantity-sensitive, but the length of vowels is inherently connected with stress (cf. Priestly 1993: 390).

According to our knowledge, Slovene has never been analysed statistically with respect to stress position. We analysed the stress position in Slovene on the basic vocabulary from a Slovene-German learner's dictionary (cf. Kelih & Vučajnk 2018). Moreover, we are in a position to distinguish different parts of speech. Our data basis is as follows: 1522 nouns, 273 adjectives, and 528 verbs. Lemmas are taken as basic units (i.e., in the case of nouns, word forms in nominative singular; adjectives in the short form of masculine nominative singular, as commonly given in dictionaries of Slovene; and verbs as infinitives). The lemma analysis thus excludes (and this is important to note) the possible change of the stress position in other cases, e.g., Nom. Sg. *m**o**st* ('bridge') is considered, but Gen. Sg. *most**u*** is not. One further characteristic of free stress in Slovene is a remarkable tendency towards variation. There are many word forms where an alternate stress positioning would be allowed, e.g., *z**i**dati* and *zid**a**ti* ('to build', 'to construct'), both common in the Slovene Standard language. However, for the sake of simplicity, for our analysis, we took the accent position, which is mentioned in the most important dictionary of contemporary Slovene in the first position (cf. SSKJ 2014 at fran.si). Frequencies of stress positions in Slovene can be found in Tab. 2.

It is obvious that stress tends to occur in the middle of words and that it mostly avoids word peripheries. However, as opposed to Russian, there is no clear preference for the second half of the word. In shorter nouns and adjectives (those consisting of two and three syllables), stress on the first half of words prevails, while in longer ones, the opposite is true. Verbs are an extra category, as the infinitive of Slovene verbs ends in the suffix *-ti* (there are a few exceptions ending in the suffix *-či*), which never carries stress. Regardless of verb length, the penultimate position (i.e., the placement of stress on the last possible syllable) is the most frequent.

Tab. 2: Stress position in Slovene (SP – stress position, WL – word length).

	nouns					adjectives				verbs				
SP	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**5**
WL														
1	171					34								
2	377	166				78	26			42				
3	95	333	55			15	67	15		97	145			
4	10	86	156	6		1	7	20	4	1	86	101		
5		9	23	26				3	3			21	22	
6				7	2								1	2

4.3 Ukrainian

Ukrainian belongs, alongside Russian, to the subgroup of Eastern Slavic languages (it is, therefore, interesting to check for similarities and differences between these two languages). The situation concerning quantitative-oriented investigations on stress position in Ukrainian is similar to that of Slovene – we are not aware of any published results (Łukaszewicz & Mołczanow 2018: 259 even write that Ukrainian "remains a *terra incognita* in phonological literature").

The Ukrainian data we present in Tab. 3 were created from the corpus of Ukrainian works of fiction (see http://www.mova.info/cfq1.aspx?fdid=hproz2018). We chose the same approach as for Slovene, i.e., we considered nouns, adjectives, and verbs separately, and we worked with word lemmas (nominative singular for nouns, masculine nominative singular for adjectives, infinitive for verbs). First, 1500 most frequent lemmas from each of the three parts of speech were taken from the corpus. Some of them were then disregarded, because either they were not included in the Ukrainian-Slovak dictionary by Popeľ (1960), which was used to determine stress positions, or stress position was ambiguous (the dictionary offers two possibilities for some words, without an indication which one is preferred, occurs more often, etc.). Thus, we analysed a sample consisting of 1221 nouns, 1122 adjectives, and 1369 verbs.

Once again, syllables in the centre of a word carry stress more often than those at word peripheries. As far as the influence of word length is concerned, the Ukrainian data display a tendency similar to those from Slovene – in two-syllable words, the beginning of the word is preferred; three-syllable ones do not provide a clear picture; while in longer words, the second half is stressed more often.

Tab. 3: Stress position in Ukrainian (SP – stress position, WL – word length).

	nouns					adjectives					verbs				
SP	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
WL															
1	208					7					1				
2	380	201				179	90				52	27			
3	56	225	62			61	426	44			128	310	35		
4		30	39	3		5	85	170	2		60	203	291	4	
5			10	3	2			22	28		5	63	117	45	
6				2					1	2		5	13	6	2
7													1	1	

5 Results

5.1 Stress position

For all three languages under study, it is obvious (see Tab. 4, which also contains a column with Ukrainian nouns and adjectives, i.e. verbs were excluded, the reason to be explained later in this section) that the mean stress position moves to the right as word length increases – mathematically speaking, if $MSP(x)$ is the mean stress position in x-syllabic words, $MSP(x)$ is an increasing function of the variable x. In addition, we also introduce the notion of the relative mean stress position in x-syllabic words,

$$RMSP(x) = \frac{MSP(x) - 1}{x - 1}$$

The relative mean stress position can be interpreted as the "centre of gravity" with respect to stress. For example, the stress in a three-syllabic word can be on either of the first, second, or third syllable. One can imagine it as a line segment with endpoints 1 and 3, i.e., as a line segment of length two. The formula above normalizes the segment so that its endpoints are 0 and 1, with possible positions of stress adjusted accordingly (i.e., in the case of a three-syllable word, possible stress positions are 0, 0.5, and 1, which correspond to the first, second, and third syllable, respectively). The values of the relative mean stress positions can be found also in Tab. 4. We note that in Tab. 4, we disregarded the single word of length 11 in Russian and two words of length 7 in Slovene, as we considered these samples too small to be used in our analyses.

Tab. 4: Dependence of mean stress position (MSP) and relative mean stress position (RMSP) in Russian, Slovene, and Ukrainian on word length (WL).

	Russian		Slovene		Ukrainian		Ukrainian (without verbs)	
	MSP	RMSP	MSP	RMSP	MSP	RMSP	MSP	RMSP
WL								
2	1.57	0.57	1.28	0.28	1.34	0.34	1.34	0.34
3	2.21	0.61	1.83	0.42	1.92	0.46	1.99	0.50
4	2.65	0.55	2.58	0.53	2.51	0.50	2.64	0.55
5	3.34	0.59	3.39	0.60	3.02	0.51	3.45	0.64
6	4.02	0.60	4.33	0.66	3.39	0.48	4.40	0.68
7	4.75	0.63						
8	5.48	0.64						

The mean stress position can be modelled by the linear function

$$MSP(x) = ax + b$$

which achieves an excellent fit for all three languages ($a = 0.65$, $b = 0.20$, $R^2 = 0.99$ for Russian; $a = 0.77$, $b = -0.38$, $R^2 = 0.99$ for Slovene; and $a = 0.52$, $b = 0.36$, $R^2 = 0.99$ for Ukrainian with all words taken into account).

The relative mean stress positions are increasing only for Slovene. However, for Russian, only three values are problematic (the sequence of relative mean stress positions is increasing for word lengths from 4 to 9). First, there is a decrease at the very end, but the sample contains only eight words of length 10 (see Tab. 1), hence the robustness of the value is questionable, and it can easily change if more words are added. Second, an increase can be observed at the beginning. We remind that the Russian sample differs from the other two in two aspects: (a) it is comprised of all word forms, not only lemmas, (b) there are no restrictions with respect to parts of speech, while only nouns, adjectives, and verbs were chosen for Slovene and Ukrainian (where frequently used items were analysed). For the time being, we do not know anything about the behaviour of stress in particular parts of speech in Russian, but as most synsemantic words are short, they are represented in higher proportion among shorter words. In addition, the stress in some Russian words is movable, e.g., *ruk**á*** ('hand') in Nom. Sg., but *r**ú**ku* in Acc. Sg, *r**ú**ki* in Nom. Pl. Kempgen (1995: 34) claims that a movable accent occurs much more often in very frequent words, but according to the famous Zipf's law of brevity (cf. Zipf 1949, or recently Casas et al. 2019), these words tend to be short. Consequently, some fluctuations in stress positions can be

caused by movable stress, and possibly also by different stress patterns in synsemantic words (which is an open question).

The only irregularity in Ukrainian can be observed at the end for words of length 6, and it disappears if verbs are disregarded. As opposed to Slovene, reflexive verbs in Ukrainian are written, according to the orthography of the language, together with the reflexive suffix *-sya*, which never carries stress,[2] e.g., бо**я**тися [boy**a**tisya] in Ukrainian vs. ***ba**ti se* (both: 'to be afraid') in Slovene. Thus, reflexive verbs are "artificially" made longer in Ukrainian, and stress can never fall on their last syllables. This fact could – albeit speculatively – explain the decrease in the relative mean stress position.

If the Ukrainian words are re-analysed with verbs omitted (the data can be found in Tab. 4), and if we consider only Russian words of length from 4 to 9, the fit of the linear model to the mean stress positions remains excellent, and parameter values in the model differ less among the three languages ($a = 0.67$, $b = -0.01$, $R^2 = 0.99$ for Russian; $a = 0.76$, $b = 0.23$, $R^2 = 0.99$ for Ukrainian). The relative mean stress positions in the re-analysed data can be modelled by the function

$$RMSP(x) = 1 - cx^d$$

with $c = 0.83$, $d = -0.41$, $R^2 = 0.99$ for Russian; $c = 1.15$, $d = -0.66$, $R^2 = 0.99$ for Slovene; and $c = 1.03$, $d = -0.64$, $R^2 = 0.95$ for Ukrainian. The limit of the function is 1 for increasing x, which means that it respects the upper theoretical limit of the relative mean stress position. The values of the parameters c and d are strongly correlated (with the Pearson correlation coefficient –0.95), which indicates that a model with only one parameter (which would be easier to interpret) might be sufficient.

The overall relative mean stress position evaluated from all data, which are presented in Tabs. 1–3, is 0.58 for Russian, 0.41 for Slovene, and 0.44 for Ukrainian.[3] The relative mean stress with the value of 0.5 would mean that, on average, exactly the middle of a word is stressed; one can see that in Russian, the second half of words is preferred, with the opposite tendency in Slovene and Ukrainian. It remains an open question whether this difference is caused by different sample methodologies (all word forms vs. word lemmas; basic

2 The same orthographic principle is followed in Russian.

3 If only words with length from 4 to 9 are considered in Russian, and if verbs are omitted in Ukrainian, the overall relative mean stress positions do not change much – they attain values 0.57 for Russian and 0.45 for Ukrainian.

vocabulary vs. the most frequent words vs. no such restrictions; all parts of speech vs. nouns, adjectives, and verbs).

5.2 Word length in different parts of speech

Our research brings also an interesting "by-product" – for Slovene and Ukrainian. It is possible to test whether word length distribution differs among nouns, adjectives, and verbs.[4] Data are presented in Tab. 5. We remind that the data represent the length of word types (as they were taken from dictionaries) as opposed to tokens (which is a more usual approach in word length studies, cf. Grzybek 2006).

Tab. 5: Word length distribution for nouns, adjectives, and verbs in Slovene and Ukrainian.

	word length	1	2	3	4	5	6	7
Slovene	nouns	171	543	483	258	58	9	
	adjectives	34	104	97	32	6	0	
	verbs	0	42	242	198	43	6	
Ukrainian	nouns	208	581	343	72	25	2	0
	adjectives	7	269	531	262	56	3	0
	verbs	1	79	473	558	230	26	2

The chi-square test reveals that if word length distributions of all three parts of speech are tested, the null hypothesis (which says that the proportions do not differ among parts of speech) is rejected for both languages (p-value less than 0.01 in both cases). However, the difference between nouns and adjectives (i.e., without verbs) remains significant for Ukrainian, but not for Slovene (with a p-value equal to 0.11)

6 Conclusion

It is a well-known fact that even in languages with fixed stress, not all words "behave" as they should according to their typological affiliation. According to

4 We do not have this possibility for Russian, as data which we took from Kempgen (1995) (originally from Moiseev 1976) do not distinguish among parts of speech.

Hyman (1977: 56), for instance, in languages that are supposed to have final stress, the final syllable indeed carries stress only in 90% of the word forms. This observation leads to the question of where, in fact, to draw the line between languages with fixed and free stress. The behaviour of stress is, in principle, stochastic.

A similar problem appears, as we have shown, in languages with free stress. The freedom of stress position is by no means unlimited, but particular patterns and tendencies can be observed. Thus, one has to agree with Hyman (1977: 56), who asks: "How do we decide what is 'dominant'?" There is no trivial answer, but simple counting already provides remarkable hints for further discussion.

The data and analyses presented above give us a possibility to summarize and generalize tendencies in stress positions up to a certain extent (we are aware of the fact that the analysis of three languages from one language family is of limited relevance). There seems to be quite a systematic trend in languages with free stress – the stress prefers positions in the middle of words and moves towards the end of the word with the increasing word length. This trend must be verified in several typologically diverse languages.

Our investigations also suggest several directions for future research. First, data from this paper were obtained by two different approaches (all word forms in Russian vs. word lemmas in Slovene and Ukrainian). Both approaches are relevant, but they are not directly comparable. All three languages under study are fusional, and different cases are marked by suffixes which make word forms longer. Moreover, there is also the phenomenon of the moving stress, which has an impact on stress positions in word forms, but cannot be taken into account if we restrict ourselves to lemmas. Second, stress position can be studied not only in dictionaries (i.e., on word types), but also in texts on word tokens (where proportions of shorter words will be higher than in dictionaries). Finally, it seems that word length behaviour (we remind that word length is one of the main factors which determine stress position) differs among different parts of speech, which is a topic that deserves systematic research, again on the levels of both types and tokens.

References

Baerman, Matthew. 1999. *The Evolution of Fixed Stress in Slavic*. München: Lincom.

Casas, Bernardino, Antoni Hernández-Fernández, Neus Català, Ramon Ferrer-i-Cancho & Jaume Baixeries. 2019. Polysemy and brevity versus frequency in language. *Computer Speech & Language* 58. 19–50.

Cubberley, Paul V. 1980. The Suprasegmental Features in Slavonic Phonetic Typology. (Bibliotheca Slavonica 20). Amsterdam: Hakkert.

Clark, John & Colin Yallop. 1995. *An Introduction to Phonetics and Phonology*. Oxford: Blackwell.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20(3). 491–552.

Grzybek, Peter. 2006. History and methodology of word length studies. In Peter Grzybek (ed.), *Contributions to the Science of Text and Language. Word Length Studies and Related Issues*, 15–90. Dordrecht: Springer.

Hyman, Larry M. 1977. On the nature of linguistic stress. In Larry M. Hyman (ed.), *Studies in Stress and Accent. Southern California Occasional Papers in Linguistics 4*, 37–82. Los Angeles: University of Southern California.

Kelih, Emmerich & Tatjana Vučajnk. 2018. Slovensko-nemški tematski slovar: osnovno in razširjeno besedišče. 4500 gesel, frazemov in stavčnih primerov. Grund- und Aufbauwortschatz Slowenisch-Deutsch. 4500 Lemmata, Phrasen und Satzbeispiele. Klagenfurt/Celovec-Ljubljana/Laibach-Wien/Dunaj: Hermagoras/Mohorjeva.

Kempgen, Sebastian. 1995. *Russische Sprachstatistik*. München: Otto Sagner.

Kempgen, Sebastian. 2008. Das Makedonische – auf dem Weg zur Anfangsbetonung? In Gabriel Altmann, Iryna Zadorozhna & Yuliya Matskulyak (eds.), *Problems of General, Germanic and Slavic Linguistics. Papers for the 70th Anniversary of Professor V.V. Levickij*, 311–318. Chernivtsi: Knihi-XXI.

Krüger, Kersten. 2009. Freier Akzent (Flexion). In Tilman Berger, Karl Gutschmidt, Sebastian Kempgen & Peter Kosta (eds.), *The Slavic Languages. An International Handbook of Their History, Their Structure and Their Investigation. Volume 1*, 86–100. Berlin/New York: de Gruyter.

Lehfeldt, Werner. 2012. Akzent und Betonung im Russischen, 2., verbesserte und erweiterte Auflage. München: Sagner.

Łukaszewicz, Beata & Janina Mołczanow. 2018. Leftward and rightward stress iteration in Ukrainian. In Bartłomiej Czaplicki, Beata Łukaszewicz & Monika Opalińska (eds.), *Phonology, Fieldwork and Generalizations*, 259–275. Berlin: Peter Lang.

Moiseev, Aleksandr. 1976. Mesto slovesnogo udarenija v sovremennom russkom literaturnom jazyke. *Studia Rossica Posnaniensia* 7. 77–87.

Popeľ, Ivan. 1960. *Ukrajinsko-slovenský slovník* [Ukrainian-Slovak dictionary]. Bratislava: Slovenské pedagogické nakladateľstvo.

Priestly, Tom M. S. 1993. Slovene. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic Languages*, 388–451. London: Routledge.

SSKJ. 2014. *Slovar Slovenskega Knjižnega Jezika. Prva knjiga A-Pa & Druga knjiga Pe-Ž* [Dictionary of Slovene literary language. The first volume A-Pa & The second volume Pe-Ž], 2nd edn. Ljubljana: Cankarjeva založba.

Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge (MA): Addison-Wesley.