Original research article
UDK 811.163.41`373.611
DOI 10.21618/fil2532326t
COBISS.RS-ID 143855361

Gorica R. Tomić[1]
University of Kragujevac
Faculty of Education in Užice

POSITIONING SOURCE WORDS IN TWO-MEMBER SERBIAN LEXICAL BLENDS: AN ANALYSIS OF SOURCE WORDS' SYLLABLE NUMBER AND FREQUENCY

Abstract: *Starting from the research results for intentional English blends which demonstrate that their first source word tends to have fewer syllables (Kelly, 1998; Renner, 2014 for coordinate blends; Gries, 2004b; 2004c; 2012 for (non-)coordinate blends) as well as greater frequency than their second source word (Kelly, 1998 for coordinate blends; Gries, 2004c; 2012 for (non-)coordinate blends), this paper aims to investigate whether the factors of syllable number and word frequency influence the positions of two source words in intentional paradigmatic and non-paradigmatic Serbian blends. To this end, a sample of 358 (non-)paradigmatic blends was drawn from the existing collections of blends in Serbian. The results of the statistical analysis show that the positioning of two source words in both paradigmatic and non-paradigmatic blends is significantly influenced by the source words' number of syllables, but not their frequency.*

Key words: *Serbian blends, source words, positions, syllable number, frequency*

1. Introduction

In present-day Serbian, blends are indeed the most authentic, as well as the wittiest lexical formations (Prćić, 2018:p.85). Following Ronneberger-Sibold's (2006:p.157) definition of lexical blends, they are defined here as intentional "creation[s] of new word[s] out of two (or rarely more) previously existing ones in a way which differs from the rules or patterns of regular compounding".[2] Some of

[1] tomic@pfu.kg.ac.rs

[2] According to Ronneberger-Sibold (2006: p.157), "[t]his definition calls for a clear delimitation of blending, first, from different kinds of non-deliberate operations [e.g., error blends], second, from

the newest products of lexical blending in Serbian include creations such as *grižoto* 'pea risotto' ← gr(ašak) 'pea' × rižoto 'risotto', *knjidžamice* 'mini pyjamas for books' ← knji(ge) 'books' × (p)idžamice 'mini pyjamas' and *rokesto* 'Serbian radish leaf pesto' ← ro(t)k(vica) 'radish' × (p)esto 'pesto sauce'.[3] Lexical blending, as a non-rule governed word-formation process (Gries, 2012:p.145), is, according to some authors (e.g., Bauer, 2012:pp.11–12), probably best understood as a prototypical category "whose description centers around a number of defeasible constraints" (Renner et al., 2012:p.4), which include but are not limited to phonological constraints, word frequency, prototypicality of source words (SWs), semantic constraints, pragmatics, ease of pronunciation.[4] Because of their "fascinating structure" (Bauer, 2012:p.11), the structural aspects of blends, including Serbian blends, have been, unsurprisingly, most widely studied (see, e.g., Bugarski, 2001; 2019; 2021; Halupka-Rešetar & Lalić-Krstin, 2009; Lalić-Krstin & Halupka-Rešetar, 2007; Tomić, 2019; 2022a; 2022b; 2023; Tomić & Danilović Jeremić, 2020). Thanks to these research studies, we know that blends in Serbian typically involve two source words, that they can be formed by a wide variety of blending techniques (which usually involve overlaps between the source words), that they are mostly nouns, that they combine various syntactic categories as well as non-adapted foreign words, and that they are normally grammatically right-headed. Although the factors of length (in terms of syllables), word frequency, as well as prototypicality of source words are expected to hold in blends from other languages (besides English) because of the "general principles of economy, Zipf's Law, and the like" (Bauer, 2012:p.13), Halupka-Rešetar & Lalić-Krstin (2012) found that the factor of prototypicality does not determine the order of source words in experimentally-elicited coordinate blends in Serbian, whose members belong to the semantic categories of *fruit* and *vegetables*. Notwithstanding the above list, there are still some formal aspects of Serbian blends we have no knowledge of, such as the influence of the factors of source words' length (in terms of syllables) and frequency on their positions in the

regular word formation, in particular compounding, and third, from other kinds of word creation outside regular grammar". For more about the definition of lexical blending, see Renner (2023: pp. 2–4).

[3] The '×' symbolizes the process of lexical blending. Segments of the source words which are not contained in the blend are bracketed. Segment overlapping is indicated by underlining. All three examples are taken from my personal collection of new blends in Serbian.

[4] Interestingly, more recent studies on blends (primarily in English) seem to be far more optimistic about the predictability of blend structure (see, e.g., Gries, 2004a; 2004c; 2006; Mattiello, 2013) than those from earlier times (see, e.g., Bauer, 1983: pp.232, 235, who classified blends under "unpredictable formations", arguing that it is not known what the constraints are "beyond pronounceability and spellability"; cf. also Cannon, 1986: p.744).

blend. It is exactly this question the present research aims to answer by analyzing a sample of intentional (non-)paradigmatic blends in Serbian. Accordingly, the aim of this paper is to establish whether the number of syllables and the frequency of source words influence their positions in intentional (non-)paradigmatic Serbian blends, and whether there are any differences between the two morphosemantic types of blends with respect to the influence of these two factors.

The paper is organized into five sections. The introduction is followed by a review of relevant research studies (Section 2). Section 3 provides more details about the data, i.e., a sample of Serbian blends, as well as the methods used in data analysis. In Section 4, the research results are presented and discussed in light of the other relevant findings. Finally, Section 5 concludes the paper by summarizing the research results and discussing a few implications for future research into the structure of (Serbian) lexical blends.

2. Literature review

In this section, I will give an overview of the studies whose research focus involved analyzing the number of syllables as well as the frequency of source words in two-member blends with the aim of finding out whether the two factors have an effect on the source word ordering. To the best of my knowledge, such studies have been conducted only for English blends so far. This is not surprising because blends in English are by far the most extensively and most systematically explored products of lexical blending. English blends are also most numerous, which is not insignificant for studies such as the ones overviewed here or presented in this paper, since the number of available examples of blends can have a considerable impact on the feasibility of the study (cf. Halupka-Rešetar & Lalić-Krstin, 2012: p.477), as well as on the validity and generalizability of its results. For my research, particularly relevant are the papers by Kelly (1998), Gries (2004b; 2004c; 2012) and Renner (2014). In reviewing these papers, I will focus on the research results that are most pertinent to the analyses to be performed in this paper.

The results of Kelly's (1998) pioneering study reveal (by use of inferential statistics) that the order of two source words in English coordinate blends in which the source words are paradigmatically related (e.g., *smog* ← sm<u>o</u>(ke) × (f)<u>o</u>g or *smaze* ← sm(oke) × (h)aze) is largely predictable on the basis of certain linguistic and cognitive factors such as source words' number of syllables and frequency. Specifically, using a sample of 320 two-member English blends (extracted from Pound (1914) and the *Oxford English Dictionary*), he (Kelly, 1998: p.582) showed

that there is a strong tendency towards placing a shorter (in terms of syllables) as well as a more frequent word first.

Kelly's results were fully supported by the results of Gries's (2004b; 2004c; 2012) studies, which were based on much larger samples of English blends. Note, however, that, unlike Kelly (1998), Gries (2004b; 2004c; 2012) included both coordinate and non-coordinate blends in his analyses. Although Gries (2004b; 2004c; 2012) analyzed both types of blends in his studies, he, unfortunately, did not discuss if there were any significant differences between coordinate and non-coordinate blend structures in terms of the syllable number and frequency effects on the ordering of two source words.

In his first study, Gries (2004b) used a sample of 1,028 two-member English blends to analyze, among other things, the number of syllables in their source words in relation to their positions in the blend. He (Gries, 2004b: p.421) found that more than half of the examples (50.77%) had a shorter source word appear earlier in the blend. In 24.90% of the blends, the source words were equally long, while 24.3% of them had a longer source word appear earlier in the blend.

Using a slightly smaller sample of 988 intentional English blends, Gries (2004c:pp.205–206) demonstrated, through the use of statistical tests, that their first source word (SW1) is significantly shorter as well as more frequent than their second source word (SW2). He further compared those results for intentional blends with the results of the statistical analysis of the source words in non-deliberate or speech-error blends in order to determine the degree of similarity between intentional and speech-error blends with regard to the two factors. The results of those comparisons showed that the two types of blends behave quite differently (Gries, 2004c:p.206).

Gries (2012) further confirmed the above findings by statistically analyzing another large sample of English blends (2,329 examples). Specifically, he (Gries, 2012:pp.149,158) found that "in intentional blends sw1 is significantly more frequent than sw2" and that "sw2 turns out to be highly significantly longer than sw1, namely about half a syllable" (Gries, 2012:p.157). Gries (2012) also compared those results with the results of the analysis of the average number of syllables and frequency of source words of authentic error and induced error blends. He (Gries, 2012:p.149) concluded that while the number of syllables and frequency of the source words of both kinds of error blends do not differ significantly from each other, the source words of intentional blends "behave differently from the source words of errors", both in terms of their average number of syllable and frequency.

Last but not least important is the paper by Renner (2014) in which the author statistically tested (using a Pearson's chi-square goodness of fit test) the significance of a number of (phonological, lexical, semantic and pragmatic) constraints on the element ordering in English coordinate lexical items including binomials, compounds and intentional blends. In the case of blends, he (Renner, 2014) analyzed a small sample of 142 blends and found that syllable number and initial consonant complexity are the only two statistically significant factors as far as the source word order is concerned, although the former (i.e., syllable number) was not independent of the frequency factor. As to lexical frequency of blend's elements, he found that the left/right element distribution in coordinate blends is not statistically significant. However, it must be noted that, as regards the frequency of the elements, Renner (2014) included only those lemmatized word forms whose frequency in the British National Corpus was minimum 5.

3. Materials and methods

Contrary to some of the studies reviewed above, which analyzed exclusively coordinate (or paradigmatic) blends in which the source words are related both syntactically and semantically, my sample includes paradigmatic as well as non-paradigmatic (or, in Dressler's (2000:p.5) and Mattiello's (2013:pp.123–125) terms, syntagmatic) blends. "Syntactically, [the source words of coordinate blends] are paradigmatically equivalent, i.e., belong to the same syntactic category, and both share their syntactic class with the final blend. Semantically, they are generally co-hyponyms of a superordinate term" (Mattiello, 2013:p.125). In syntagmatic blends, on the other hand, one source word usually functions as a syntactic and semantic head, while the other acts as its modifier, although there are examples of syntagmatic blends whose members "exhibit an exocentric relationship" (Mattiello, 2013:p.124; cf. also Beliaeva, 2014:pp.75,188). Here, for the blends whose source words are not paradigmatically equivalent and therefore cannot be "expanded into coordinate phrases" (Gries, 2004c:p.204), the term non-paradigmatic is used, since not all of them fulfil the criteria for syntagmatic blends as defined in Mattiello (2013: p.124). Namely, there are examples of Serbian blends in my sample where one source word is an argument of the other (e.g., *iskoronimo* 'eradicate coronavirus' ← isk<u>or(</u>e)<u>n</u>imo 'eradicate' × <u>koron</u>(a) 'coronavirus') (hence the term *argumental blends* (Bauer et al. (2013: pp.483–484)). According to Bauer et al. (2013:p.484), argumental blends can be either subject-referencing (e.g., *vilozofirati* 'philosophize in a fairy's manner' ← v<u>il(</u>a) 'fairy' × (f)<u>il</u>ozofirati 'philosophize') or object-refer-

encing (e.g., *iskoronimo*). Following Renner (2014:p.448; cf. also Lalić-Krstin, 2010:p.109), the paradigmatic or coordinate relationship between the source words of Serbian blends was established by using the available explanations of blends in the sources, which included markers such as "cross of A and B", "mixture of A and B", "combining A and B" or "functioning as A and B".

Despite the claims encountered in the literature (cf. Bauer, 2012:p.12; Halupka-Rešetar & Lalić-Krstin, 2012:p.476) that the positions of SWs in (English and Serbian) syntagmatic blends are predetermined by "the rules of ordering constituents in phrases and compounds", Gries (2004a:p.649) argues that both these types should be included in the analysis. As aptly pointed out by Gries (2004a: pp.648–649), although "on the one hand, it intuitively seems to make sense to restrict the analysis of source-word frequencies to those blends that can be expanded into coordinate structures [...]. On the other hand, there is no reason not to include noncoordinate blends (such as Westralia [West Australia]) in the analysis as well to see whether frequency plays a role for all kinds of blends".

With the aim of establishing whether SW1 of paradigmatic and non-paradigmatic blends in Serbian tends to be shorter and more frequent than SW2, a sample of 224 blends from the collection of 400 Serbian blends given in Tomić (2023) was used. Considering the aim of the research, one of the methodological problems was the question which blends to sample. For example, hybrid English-Serbian and Serbian-English blends were not included, since the English words used in them are not adapted to the phonological and/or morphological system of the Serbian language (cf. Tomić, 2022b). That is, since the source words of such cross-linguistic blends belong to two different languages, they are not comparable as regards the frequency factor. The frequency of a non-adapted English word in a Serbian corpus would actually give false results about its frequency in the English language. On the other hand, the blends whose one source word is an adapted form of an English word are included in the sample (e.g., *krindžama* ← kr<u>indž</u> 'cringe'× (p) <u>idž</u>ama 'pyjamas') because, as such, they only exist in the given language (in this case, the Serbian language).

I did not include blends whose one member is a zero-syllable word (e.g., the onomatopoeic ideophone *zzz*, representing the sound of bees buzzing) either. Excluded were also (a) all graphic blends (Bugarski, 2019) (e.g., *SOler*, *PODstaknite*, *doČEPaj se*) where "a string of phonemes [or graphemes] in the longer base is identical with the entire second base" (Halupka-Rešetar & Lalić-Krstin, 2009:p.118), as well as (b) some examples of discontinuous blends (referred to as "sandwich blends" by Renner (2019)), in which one of the source words interrupts the continuity of the

other, while usually overlapping with it (e.g., *Miškomor*, *ledilo*).[5] Specifically, these two types of blends were excluded because it is extremely difficult to decidedly or unmistakably determine which of the two source words appears in first position, that is, which one of them is primary in the process of analysis (cf. Ronneberger-Sibold, 2006:p.170). For example, should we decompose the blend *Miškomor* into Mišk**o** 'a Serbian businessman's nickname' × mišomor 'rat poison' or mišomor × Mišk**o**? Similarly, is the example *PODstaknite* to be formally analyzed into **pod** 'floor' × **pod**staknite 'encourage' or **pod**staknite × **pod**? Considering that it is of fundamental importance for the analyses in this paper to be sure of the positions of the two source words, I decided not to include such blends in my sample, although one way of analyzing or reconstructing them is given in Tomić (2023, see also the appendices therein).

Lastly, I did not include few examples of (non-)paradigmatic blends where one or even both source words are homographs. The reason is that I was not able to retrieve the frequency for such source words from the corpus (see below). Put differently, it was not possible to disambiguate these forms with their specific meanings in the blend from the same forms (which may additionally belong to a different part of speech) with different meanings (e.g, *grickalice* 'snacks' vs. *grickalice* 'nail clippers', *kokice* 'popcorn' vs. *kokice* 'a hypocoristic term for a hen', *Citra* 'a type of beer' vs. *citra* 'a musical instrument', etc.). In those cases where the homographs belong to two different parts of speech (e.g., *Rasti* 'a name of a character' vs. *rasti* 'grow'), I tried using the POS tag, but it was not of much use, mainly because those words were also lumped together in one concordance or they were not correctly tagged.

All 224 blends were then divided into two subsamples, one consisting of paradigmatic (32 examples) and the other of non-paradigmatic blends (192 examples). To make the two subsamples more balanced, 129 paradigmatic examples from Bugarski (2019) and 5 paradigmatic examples from Bugarski (2021) were sampled (according to the above criteria) and added to 32 examples of paradigmatic blends taken from Tomić (2023).[6] From Bugarski's (2019; 2021) collections, I excluded examples of error blends, blends which are or may be interpreted as being imported from English into Serbian ready-made (e.g., *vežbigrice* 'exergames', *prineprijatelji* 'frenemies') (Bugarski, 2019:pp.74–75), blends which are recorded in the Croatian dictionary of lexical blends (e.g., *Karalanović*, *Mirovčak*) (Marković et al., 2016),

[5] As evinced by the examples listed here, graphic blends are "only revealed by the orthography of the blend" (Ronneberger-Sibold, 2006:p.167).

[6] Even a cursory glance at the existing collections of blends in Bugarski (2019) and Tomić (2023) suggests that the great majority of Serbian blends are of the non-paradigmatic type.

as well as the examples of blends which have exactly the same form as the existing Serbian words (e.g., *Dučić*).

In total, there were 358 blends in my sample, 166 paradigmatic (e.g., *Malimeta* 'raspberry and lime juice' ← mali(na) 'raspberry' × limeta 'lime', *Kerametal* 'name of a company selling products made of ceramics and metal' ← keram(ika) 'ceramics' × metal 'metal', *brundža* 'cranberry and orange tea' ← bru(snice) 'cranberries' × (po-mora)ndža 'orange', *čizmadrile* 'footwear that is a hybrid of boots and espadrilles' ← čizma 'boot' × (esp)adrile 'espadrille', *celeraba* 'a celery and kohlrabi hybrid' ← celer 'celery' × (k)eleraba 'kohlrabi', *izumetnik* 'a person who is both an inventor and an artist' ← izum(itelj) 'inventor' × umetnik 'artist', *medverica* 'a hybrid of bear and squirrel' ← medve(d) 'bear'× (ve)verica 'squirrel', *maždaja* 'a hybrid of cat and dragon' ← ma(čka) 'cat' × aždaja 'dragon', *duhovište* 'an imaginary creature which is both a ghost and a monster' ← duh 'ghost' × (čud)ovište 'monster') and 192 non-paradigmatic (e.g., *majmunitet* 'immunity of a monkey' ← majmun 'monkey' × (i)munitet 'immunity', *isvinjavam se* 'apologize, in a pig's manner' ← i(z)vinja-vam se 'apologize' × svinja 'pig', *kovidljamovka* 'a name of pear brandy during the Covid-19 pandemic' ← kovid 'Covid-19'× viljamovka 'pear brandy', *patlidžanstvene* 'magnificent because one of the ingredients is eggplant' ← patlidžan 'eggplant' × (ve)li(č)anstvene 'magnificent', *mumlafon* 'a microphone into which one mum-bles' ← mumla 'mumbles' × (mikro)fon 'microphone', *šumkula* 'a forest shark' ← šum(ska) 'forest' × (aj)kula 'shark', *rogulja* 'a horned eel' ← rog(ata) 'horned' × (je) gulja 'eel', *muzikaljenje* 'mastering musicianship skills' ← muzika 'music' × kaljenje 'mastering', *zbunjoblak* 'a confused cloud' ← zbunj(en) 'confused'× oblak 'cloud', *jastuknuti* 'succumb to hardships by going to sleep' ← jastuk 'pillow'× (u)stuknuti 'succumb') blends.[7]

[7] The remaining paradigmatic blends include: *veronautika*, *veromučitelj*, *Srboslavija*, *Srbogorci*, *Rusopejci*, *rigogrozno*, *restoteka*, *rekondestrukcija*, *razočajan*, *psihopaćenik*, *profetitelj*, *Mađarogorac*, *ljubičanstveno*, *kreaktivan*, *kerefetluci*, *katastrašno*, *Jugoslovenac*, *hrpski*, *Grbija*, *glupovito*, *fenofantastično*, *ekologično*, *družiciranje*, *doktator*, *bravugrozno*, *arhitekturisti*, *babćeha*, *Bajkomedija*, *birokratura*, *brodocikl*, *Cecerluša*, *ćurafa*, *dečkojčica*, *doktogirati*, *džulov*, *đuturativno*, *fenomentastično*, *gastronauti*, *gastronomad*, *geopard*, *gramovir*, *grožbuka*, *hameleon*, *horkestar*, *intelegantan*, *jučeras*, *kapuks*, *kešovina*, *kišnoća*, *kivizle*, *klinceze*, *Knjigračka*, *krajčetak*, *krmačka*, *kurveštija*, *lavgar*, *lubendinja*, *lutovati*, *Mačker*, *narkoholičar*, *njokačinke*, *obažljivo*, *odličanstveno*, *ogromantan*, *patkula*, *pivovica*, *pljugareta*, *Poezika*, *Radiovizija*, *Radiovizor*, *rečovit*, *Slobeliks*, *sociozofski*, *stripovetka*, *škozorište*, *tiglav*, *zabavantura*, *zekodlak*, *zimoća*, *Malinada*, *Medolada*, *Prasorog*, *Urmolada*, *budžep*, *Politikeni*, *ružičanstvena*, *čudečanstvena*, *đedaj*, *Gurmetnica*, *komuljare*, *Politikormer*, *smamačica*, *trolоš*, *antirafa*, *Biblioskop*, *BAŠTeAtar*, *jazopas*, *kamileoni*, *vevečka*, *žiramun*, *kolosjajno*, *zajefrkancija*, *uspostanoviti*, *tragonija*, *Srvatska*, *patridiotizam*, *premijersednik*, *autokritet*, *opsimista*, *Đitler*, *Crncuz*, *čarobični*, *dolodlazak*, *droksa*, *kenjoar*, *krebil*, *kretard*, *limunžada*, *Mačkonjić*, *Moštružnica*, *mozgonetke*, *patriopata*, *piski*, *pragnje*, *Hrbija*, *Hrbin*, *hrengleski*, *Srbazija*, *Srbat*, *svinjga*, *svrabac*, *zajebuniti se*, *ćelepinja*, *kikilada*,

The source words of all blends in my sample (i.e., 716 source words) were first analyzed for their individual number of syllables. The analysis of the *lemma* frequency of both source words of 358 blends was performed using the MaCoCu Serbian Web v1 (2021–2022) corpus in the Sketch Engine software. The MaCoCu Serbian Web v1 (2021–2022) corpus is the largest Serbian corpus available, boasting a database of 2.4 billion words. "The MaCoCu corpora were built by crawling the Internet top-level domain in 2021 and 2022, extending the crawl dynamically to other domains as well" (Sketch Engine). Following the principles of corpus linguistics (Brezina, 2018: p.46), both absolute (AF) and relative frequencies (RF) (per million tokens) for each of 716 lemmas were retrieved from the corpus.

I then counted the number of blends in which SW1 is shorter than SW2, and vice versa, as well as the number of blends in which SW1 is more frequent than SW2, and vice versa. I did this for paradigmatic and non-paradigmatic blends, respectively. All these data were subsequently used in statistical analysis. Similarly to Renner (2014), I also used a chi-square goodness of fit test (χ^2-test for goodness of fit) to establish whether the difference between the observed distributions in each blend type is statistically significant with regard to the influence of the two factors

rurbani, *enigmature*, *simpavetinja*, *sočka*, *bananas*, *šampidžan*, *šljivoviski*, *hamburek*, *fentilator*, *aforizjava*, *Aliproroković*, *Alivučić*, *bajčetnički*, *ćirilatinica*, *Prodorović*, *Šulin*, *Toris*, *Tramputin*, *Vulinda*, *Vušelj*, *Soranija*, *Šormasić*, *Kristanija*. As for the non-paradigmatic blends, the remaining examples include: *Baštastično*, *Bibiriki*, *čokolivud*, *Dečarolija*, *FILUMni*, *Irkofas*, *kokotastični*, *Krofnorog*, *kvascinantno*, *metalosaurus*, *oplaneti se*, *papirosaurus*, *plastikosaurus*, *plazmastično*, *reciklosaurusi*, *sirane*, *Sovembar*, *staklosaurus*, *StaPARK*, *susamljen*, *Testival*, *tunapčić*, *Zdravoteka*, *Zvezdapol*, *Žabalesku*, *žiraćino*, *afoerizmi*, *amandatar*, *anaprednjakinja*, *anasteziolog*, *antiNovakser*, *AVaskeva*, *AVra*, *dekonstrakcija*, *dekonstraktivno*, *Dveromobil*, *ETNArizmi*, *Festosaurus*, *FIBuloze*, *Finovacije*, *fitneSNS*, *Fratrogasci*, *globiranje*, *Hepidemiolog*, *hepidemiologija*, *Kešformisanje*, *Koštulend*, *Kukupedija*, *lažovizija*, *Lilifašistanci*, *lupmolog*, *Mitrogol*, *napisotine*, *napredNJAKANJE*, *reanomobil*, *Stripformer*, *šoljoskop*, *štamposaurusi*, *Toksimoron*, *tviteroristkinja*, *Vakcionalizam*, *Vakcionalnost*, *Veberotine*, *Vesnotres*, *Vojskosaurus*, *Vučićanstvo*, *Brokoličinke*, *Blebećkovićka*, *BLJurednik*, *BLJuformer*, *botuljeni*, *Cecesija*, *fekaloid*, *glupičanstvo*, *Grokoko*, *HEPInfekcija*, *iskoronimo*, *klozetoid*, *Kovidovdan*, *krindžama*, *Lončmar*, *miničinke*, *neINFORMERisan*, *Nolesnici*, *nultikulturalnost*, *očajnormar*, *ovsičanstveno*, *preletosaurusi*, *psujeveran*, *rebusanje*, *Smećegraf*, *Sposaurusi*, *Spssaurusi*, *sveKRIVA*, *Svinjformer*, *svinjoid*, *svinjonavirus*, *Šljamformer*, *Tviterapeut*, *unazapađen*, *urbanosaurusi*, *Uvlačićević*, *zajebAVancija*, *Zliformer*, *zloslobođenje*, *Žarkonije*, *Cepelinkovac*, *Čvoroskop*, *Dedinjastija*, *dronolitičari*, *džambasador*, *Džebanat*, *entuzijaspazam*, *EUrolog*, *gizdavitelj*, *hEUmoglobin*, *kradministrativni*, *Krasnokalipsa*, *krekonomski*, *krekusionista*, *Krokiriki*, *lepotizam*, *Mozgosaurus*, *pčeličanstvo*, *poPLJiŽdesmo*, *priPLJiŽava se*, *prohibiskucija*, *razvaškarila*, *Rokvilj*, *sarmada*, *slobistkinja*, *stanoreksija*, *Stresolovka*, *tenkvivalentnost*, *tevektualci*, *tužibabuška*, *Večnuci*, *vucibaština*, *Šizlapomotiva*, *zgurabija*, *zurlanje*, *vilozofirati*, *Grokenrol*, *Dadapokalipsa*, *zaštipci*, *Zlikovnica*, *zMAjTEMATIKA*, *izštipci*, *istetrebljenje*, *joštipci*, *Kakanik*, *kakavizija*, *kakaceza*, *marštipci*, *puzalone*, *slovembar*, *Maksimirenje*, *Novakser*, *SNSelektor*, *SNSrbija*, *SNStadion*, *Srbrljanje*, *tabloidioti*, *turbanizam*, *Veronautičar*, *Gnjavnik*, *hranac*, *mančastično*, *Promikron*, *SPSoid*, *širokoko*, *Alogaritmije*, *kaparazit*, *kokainkasant*, *kondukterijer*, *Kornjačurke*, *marihuanalitičar*, *Noktobar*, *pečalbatros*, *vokabulimija*, *ponštipci*, *šazbuka*.

on the positions of their source words. However, as rightly observed by Kelly (1998: p.582), there is one methodological problem with the analysis of the influence of the factors of syllable number and word frequency on the element ordering. Namely, the problem arises due to the *tendency* (my emphasis) of shorter words to be more frequent (Zipf, 1935), which implies that the two factors under analysis are interdependent and that they in addition need to be controlled in some way if we aim to determine the influence of each factor separately, i.e., independently of each other. This is why not all 358 blends could be used in all parts of the analysis.

4. Results and discussion

The results of the analysis aimed at determining the number of syllables in SW1 and SW2 of 166 paradigmatic and 192 non-paradigmatic blends (as individual samples) show that there is a tendency towards positioning the shorter word first, both in the paradigmatic and the non-paradigmatic (Table 1) blends alike, and that this tendency appears to be somewhat stronger in non-paradigmatic blends. Specifically, out of 166 paradigmatic blends, SW1 of 70 examples has fewer syllables than SW2. In 47 paradigmatic blends, SW1 is longer than SW2. The source words of 49 paradigmatic blends are equisyllabic. Similar results are obtained for non-paradigmatic blends. Out of 192 non-paradigmatic blends, 129 examples have the shorter word appear in first position, 33 examples have the longer word appear in first position and 30 examples have equally long source words.

Table 1 Number of (non-)paradigmatic blends whose SW1 is shorter than SW2 and those whose SW1 is longer than SW2

	SW1 shorter than SW2	SW1 longer than SW2
Paradigmatic blends	70	47
Non-paradigmatic blends	129	33

To establish whether the difference in the observed distributions of paradigmatic and non-paradigmatic blends as regards the number of syllables in SW1 and SW2 is statistically significant (equisyllabic blends are excluded from this part of the statistical analysis), I performed a chi-square goodness of fit or χ^2-test in Excel (the alpha level was set at 0.05), which is used to "test the degree of fit between your observed and an expected distribution" (Gries, 2009:p.152). The expected values, which must be larger than or equal to 5, were computed based on a 50/50

probability. The null hypothesis (H_0) states that there is no difference between the two distributions. The alternative hypothesis (H_1) states that there is a significant difference between the two distributions, and that it is not due to chance or error sampling.

For paradigmatic blends, the test ($\chi^2(1, 117) = 4.52$) gives a *p*-value of 0.033, which means that the result is significant at $p<0.05$ and that the null hypothesis should be rejected. In other words, the difference between the observed SW1<SW2 and SW1>SW2 (in terms of the syllable number) distributions in non-equisyllabic paradigmatic Serbian blends is statistically significant. For non-paradigmatic blends, the *p*-value is < 0.00001 ($\chi^2(1, 162) = 56.89$). Accordingly, the null hypothesis is rejected for non-paradigmatic blends as well. Since this result is also significant at $p<0.05$, we can conclude that the difference between the observed SW1<SW2 and SW1>SW2 (in terms of the syllable number) distributions in non-equisyllabic non-paradigmatic Serbian blends is statistically significant. Put differently, there is sufficient evidence to suggest that there is a statistically significant difference between the observed distributions of SW1<SW2 and SW1>SW2 blends (in terms of the two source words' number of syllables). The results for Serbian paradigmatic blends are in accordance with Renner's (2014) findings for English coordinate blends.

However, as previously mentioned, we also have to check whether the above differences remain significant independently of lexical frequency, that is, once the factor of word frequency is neutralized. Namely, if we neutralize the frequency factor in the above examples, i.e., if we include in the analysis only those examples of non-equisyllabic paradigmatic and non-paradigmatic blends whose SW1 is less frequent than SW2 (54 paradigmatic and 74 non-paradigmatic blends) (Table 2), the difference between the SW1<SW2 and SW1>SW2 (in terms of syllables) distributions is, interestingly, still statistically significant for non-paradigmatic blends, since the test ($\chi^2(1, 74) = 7.78$) gives a *p*-value of 0.005 (which is significant at $p<0.05$), but not for paradigmatic blends because the test (($\chi^2(1, 54) = 1.85$) gives a *p*-value of 0.17 (not significant at $p<0.05$). The result for paradigmatic blends in Serbian is also in line with Renner's (2014) findings for a somewhat smaller sample of English coordinate blends. This suggests that the syllable number factor, which predicts that SW1 will be shorter than SW2, is not independent of that of frequency in paradigmatic Serbian blends.

Table 2 Number of (non-)paradigmatic blends whose SW1 is shorter and less frequent than SW2 and those whose SW1 is longer and less frequent than SW2

	SW1 shorter than SW2 SW1 less frequent than SW2	SW1 longer than SW2 SW1 less frequent than SW2
Paradigmatic blends	22	32
Non-paradigmatic blends	49	25

The number of paradigmatic blends in which SW1 is more frequent than SW2 (93 examples) outnumbers blends whose SW2 is more frequent than SW1 (73 examples) (Table 3). Similarly, the number of non-paradigmatic blends in which SW1 is more frequent than SW2 (103 examples) outnumbers blends whose SW2 is more frequent than SW1 (88 examples) (Table 3). One non-paradigmatic blend has source words of equal relative frequency, hence the total number of 191.[8] Regarding the influence of the factor of lexical frequency which specifies that the more frequent of the two words will appear in first position (cf. Kelly, 1998; Renner, 2014), the difference between the SW1>SW2 and SW1<SW2 (in terms of their relative frequencies) distributions is found to be statistically significant neither for paradigmatic nor non-paradigmatic blends in Serbian, since the test ($\chi^2(1, 166) = 2.41$) gives a p-value of 0.12 for paradigmatic blends and a p-value of 0.28 ($\chi^2(1, 191) = 1.18$) for non-paradigmatic blends. In other words, both these results are not significant at $p<0.05$, which implies that we fail to reject the null hypothesis in this case. The results obtained herein are also in line with Renner's (2014) findings for a sample of 129 coordinate English blends.

Table 3 Number of (non-)paradigmatic blends whose SW1 is more frequent than SW2 and those whose SW1 is less frequent than SW2

	SW1 more frequent than SW2	SW1 less frequent than SW2
Paradigmatic blends	93	73
Non-paradigmatic blends	103	88

[8] Even though the combination of two source words having the same (relative) frequency is extremely uncommon (cf. Kelly, 1998:p.582), there is one non-paradigmatic blend *zgurabija* 'qurabiye prepared by a hunchbacked woman' ← zgur(ena) 'hunchbacked' (3 syllables) × gurabija 'qurabiye' (4 syllables), in which the two source words occur with the same frequency (RF=0.01) in the corpus utilized for the purposes of this paper.

After neutralizing the influence of the syllable number factor, that is, after excluding all non-equisyllabic paradigmatic and non-paradigmatic blends from the above subsamples (Table 4) and therefore from further statistical analysis, the test results reveal that neither of the differences between the distributions as regards SW1's and SW2' relative frequency is statistically significant at $p<0.05$. Specifically, the test ($\chi^2(1, 49) = 1.65$) gives a p-value of 0.20 for paradigmatic blends and a p-value of 0.72 for non-paradigmatic blends ($\chi^2(1, 30) = 0.13$). Consequently, we fail to reject the null hypothesis. There is, however, no sufficient evidence to conclude that the alternative hypothesis is true.

Table 4 Number of equisyllabic (non-)paradigmatic blends whose SW1 is more frequent than SW2 and those whose SW1 is less frequent than SW2

	SW1 more frequent than SW2	SW1 less frequent than SW2
Paradigmatic blends	29	20
Non-paradigmatic blends	16	14

This piece of evidence suggests that the influence of the frequency factor on the positioning of two source words in paradigmatic and non-paradigmatic Serbian blends is not statistically significant. It is noteworthy that the above results for Serbian paradigmatic blends are compatible with Renner's (2014) findings for the same number of English coordinate blends (49 examples) as well.

5. Conclusions

In this paper, I aimed to investigate whether the factors of syllable number and lexical frequency influence the positions of two source words in intentional Serbian blends, both paradigmatic and non-paradigmatic, as has been, for instance, demonstrated for intentional English blends in some studies. On the basis of the results obtained by means of inferential statistics, several conclusions can be drawn. First, the factor of syllable number is found to be statistically significant in determining the positions of two source words both in paradigmatic and non-paradigmatic Serbian blends, in that the shorter source word appears earlier in the blend than the longer source word. Second, although syllable number significantly influences the source word positioning in both these types of blends, it must be noted that it is not independent of the word frequency factor in paradigmatic Serbian blends. Third,

relative frequency of source words, in that the more frequent source word appears in first position, is found to be statistically significant neither for paradigmatic nor for non-paradigmatic blends in Serbian. Despite the fact that the subsamples of Serbian blends in certain parts of the statistical analysis were rather small, it is safe to conclude that, in addition to the ease of blend pronounceability and, possibly, phonemic or graphemic similarity of source words to each other and to the blend itself, the factor of syllable number plays a significant role in determining the source word position in intentional two-member Serbian blends. Perhaps more importantly, these findings also point to the fact that the positions of source words in Serbian blends are not entirely unpredictable and that there is actually "some method to blends' madness".

In light of the above conclusions, it would certainly be interesting to compare results obtained in this paper with the results of some future research studies on paradigmatic and non-paradigmatic blends from other Slavic languages in which blending has also been productive. As to Serbian blends alone, further research studies could, for example, examine whether the shorter source word actually contributes more of itself (in terms of syllables or graphemes (phonemes)) to the resultant blend, as has already been suggested for intentional English blends. Another possible avenue for research could involve investigating whether Serbian blenders tend to position the source word with positive connotation first, since we still have no knowledge of the influence of this semantic variable on the structure of intentional Serbian blends, specifically their source words' positions.

Sources

1. Bugarski, R. (2019) *Srpske slivenice: monografija sa rečnikom*. Novi Sad, Akademska knjiga.
2. Bugarski, R. (2021) *Saga o ćirilici: ogledi o jeziku i nacionalizmu*. Beograd, Biblioteka XX vek.
3. MaCoCu Serbian Web v1 (2021–2022) Sketch Engine. Accessed January 2024. https://auth.sketchengine.eu/#login.
4. Tomić, G. (2023) *Leksičke slivenice u savremenom engleskom i srpskom jeziku: strukturni, sadržinski i sociolingvistički aspekti*. Doktorska disertacija, Univerzitet u Kragujevcu.

References

1. Bauer, L. (1983) *English word-formation*. Cambridge, Cambridge University Press.
2. Bauer, L. (2012) Blends: Core and periphery. In: V. Renner, Maniez, F. & Arnaud, P. (eds.) *Cross-Disciplinary Perspectives on Lexical Blending*. Berlin, Boston, De Gruyter Mouton. 11–22. https://doi.org/10.1515/9783110289572.11

3. Bauer, L., Rochelle L., & Plag. I. (2013) *The Oxford reference guide to English morphology*. Oxford, Oxford University Press.
4. Beliaeva, N. (2014) *Unpacking contemporary English blends: Morphological structure, meaning, processing*. PhD Diss. Victoria University of Wellington. https://doi.org/10.26686/wgtn.17008108
5. Brezina, V. (2018) *Statistics in corpus linguistics: A practical guide*. Cambridge, Cambridge University Press.
6. Bugarski, R. (2001) Dve reči u jednoj: leksičke skrivalice. *Jezik danas*. 5 (13), 1–5.
7. Bugarski, R. (2019) *Srpske slivenice: monografija sa rečnikom*. Novi Sad, Akademska knjiga.
8. Bugarski, R. (2021) Saga o ćirilici: ogledi o jeziku i nacionalizmu. Beograd, Biblioteka XX vek.
9. Cannon, G. (1986) Blends in English Word Formation. *Linguistics*. 24 (4), 725–753. https://doi.org/10.1515/ling.1986.24.4.725
10. Dressler, W. U. (2000) Extragrammatical vs. Marginal Morphology. In: U. Doleschal & A. M. Thornton (eds.) *Extragrammatical and Marginal Morphology*. München, Lincom Europa, pp. 1–10.
11. Gries, S. Th. (2004a) Shouldnt It Be Breakfunch? A Quantitative Analysis of Blend Structure in English. *Linguistics.* 42 (2), 639–667. https://doi.org/10.1515/ling.2004.021
12. Gries, S. Th. (2004b) Isn't That Fantabulous? How Similarity Motivates Intentional Morphological Blends in English. In: M. Achard & S. Kemmer (eds.) *Language, Culture, and Mind*. Stanford, CA, CSLI Publications, pp. 415–428.
13. Gries, S. Th. (2004c) Some Characteristics of English Morphological Blends. In: M. Adronis, E. Debernport, A. Pycha, & K. Yoshimure (eds.) *Papers from the 38th Regional Meeting of the Chicago Linguistic Society*. Vol.II. The Panels, pp. 201–216.
14. Gries, S. Th. (2006) Cognitive Determinants of Subtractive Word Formation: A Corpus-Based Perspective. *Cognitive Linguistics*. 17 (4), 535–558. https://doi.org/10.1515/cog.2006.017
15. Gries, S. Th. (2009) *Statistics for Linguistics with R: A Practical Introduction.* Berlin, New York, Mouton de Gruyter.
16. Gries, Stefan Th. (2012) Quantitative Corpus Data on Blend Formation: Psycho- and Cognitive-Linguistic Perspectives. In: V. Renner, Maniez, F. & Arnaud, P. (eds.) *Cross-Disciplinary Perspectives on Lexical Blending*. Berlin, Boston, De Gruyter Mouton, pp. 145–168. https://doi.org/10.1515/9783110289572.145.
17. Halupka Rešetar, S, & Lalić Krstin, G. (2009) New blends in Serbian: typological and headedness-related issues. *Godišnjak Filozofskog fakulteta u Novom Sadu*. 34 (1), 115–24. http://digitalna.ff.uns.ac.rs/sadrzaj/2009/0374-0730
18. Halupka-Rešetar, S. & Lalić-Krstin, G. (2012) Prototypicality as governing factor in constituent ordering in Serbian blends. In: Snežana Gudurić (ed.) *Jezici i kulture u vremenu i prostoru: tematski zbornik* I. Novi Sad, Filozofski fakultet, pp. 475–486.

19. Kelly, M. H. (1998) To 'Brunch' or to 'Brench': Some Aspects of Blend Structure. *Linguistics*. 36 (3), 579–90. https://doi.org/10.1515/ling.1998.36.3.579
20. Lalić-Krstin, G. & Halupka-Rešetar, S. (2007) Nešto novo o novim slivenicama u srpskom jeziku. *Svet reči: srednjoškolski časopis za srpski jezik i književnost.* 11 (23/24), 26–30.
21. Lalić-Krstin, G. (2010) *Strukturni i sadržinski aspekti novih slivenica u engleskom jeziku: kognitivnolingvistički pristup*. Doktorska disertacija. Univerzitet u Novom Sadu.
22. Marković, I., Klindić, I. & Borković, I. (2016) *Hrvatski rječnik stopljenica*. https://stilistika.org/hrvatski-rjecnik-stopljenica.
23. Mattiello, E. (2013) *Extra-grammatical morphology in English: Abbreviations, blends, reduplicatives, and related phenomena*. Berlin, De Gruyter Mouton.
24. Prćić, T. (2018) *Ka savremenim srpskim rečnicima*. 1., Elektronsko izd. Novi Sad, Filozofski fakultet. http://digitalna.ff.uns.ac.rs/sadrzaj/2018/978-86-6065-454-2
25. Renner, V. (2014) A Study of Element Ordering in English Coordinate Lexical Items. *English Studies.* 95 (4), 441–58. https://doi.org/10.1080/0013838x.2014.908039.
26. Renner, V. (2019) French and English Lexical Blends in Contrast. *Languages in Contrast*. 19 (1), 27–47. https://doi.org/10.1075/lic.16020.ren.
27. Renner, V. (2023) Blending. In: P. Ackema, S. Bendjaballah, E. Bonet, et al. (eds.) *The Wiley Blackwell Companion to Morphology*. Oxford, Wiley-Blackwell, pp. 1–19. https://doi.org/10.1002/9781119693604.morphcom009.
28. Renner, V., Maniez, F. & Arnaud, P. J. (2012) Introduction: A Bird's-Eye View of Lexical Blending. In: V. Renner, Maniez, F. & Arnaud, P. (eds.) *Cross-Disciplinary Perspectives on Lexical Blending*. Berlin, Boston, De Gruyter Mouton, pp. 1–10. https://doi.org/10.1515/9783110289572.1.
29. Ronneberger-Sibold, E. (2006) Lexical Blends: Functionally Tuning the Transparency of Complex Words. *Folia Linguistica*. 40 (1–2), 155–181.
30. Tomić, G. & Danilović Jeremić, J. (2020) Ekspresivnost u nazivima hrane i pića: analiza. U: M. Kovačević & J. Petković (ur). *Srpski jezik, književnost, umetnost: zbornik radova sa XIV međunarodnog naučnog skupa održanog na Filološko-umetničkom fakultetu u Kragujevcu (25–27. X 2019). Knj. 1, Ekspresivnost u srpskom jeziku*, Kragujevac: Filološko-umetnički fakultet, pp. 289–302. https://drive.google.com/file/d/1osu8zIkj6XFiN84fzPR_5avo4fnXrHTX/view.
31. Tomić, G. (2019) Tvorbeno-semantička analiza novih slivenica u srpskom jeziku. *Zbornik radova Filozofskog fakulteta u Prištini*. 49 (2), 61–83. https://scindeks-clanci.ceon.rs/data/pdf/0354-7795/2019/0354-77951902061T.pdf.
32. Tomić, G. (2022a) Lexical Blends in Serbian: An Analysis of Morphosemantic Transparency. *Zbornik radova Filozofskog fakulteta u Prištini*. 52 (3), 13–37. https://scindeks-clanci.ceon.rs/data/pdf/0354-3293/2022/0354-32932203013T.pdf.
33. Tomić, G. (2022b) Hybrid English-Serbian and Serbian-English Lexical Blends in Serbian. *Philologia Mediana*. 14, 409–28. http://www.philologiamediana.com/index.php/phm/article/view/750/655.

34. Tomić, G. (2023) *Leksičke slivenice u savremenom engleskom i srpskom jeziku: strukturni, sadržinski i sociolingvistički aspekti*. Doktorska disertacija, Univerzitet u Kragujevcu.
35. Zipf, G. K. (1935) *The psycho-biology of language*. Cambridge, Massachusetts, The M.I.T. Press.

Gorica R. Tomić
Univerzitet u Kragujevcu
Pedagoški fakultet u Užicu

POZICIONIRANJE MOTIVNIH REČI U DVOČLANIM LEKSIČKIM SLIVENICAMA U SRPSKOM JEZIKU: ANALIZA BROJA SLOGOVA I FREKVENTNOSTI MOTIVNIH REČI

ЛЕКТОРИСАНО

Rezime

Ako se imaju na umu rezultati istraživanja (Gries, 2004b; 2004c; 2012; Kelly, 1998; Renner, 2014) svesno i namerno tvorenih slivenica u engleskom jeziku koji ukazuju da na redosled motivnih reči u (ne)koordinativnim slivenicama utiču broj slogova i frekventnost, kao i to da u literaturi o slivenicama u srpskom jeziku nedostaju istraživanja o tome koji sve faktori utiču na strukturu, tačnije redosled motivnih reči u slivenicama, cilj ovog rada jeste da istraži da li faktori kao što su broj slogova i relativna frekventnost motivnih reči imaju uticaja na njihov redosled u paradigmatskim i neparadigmatskim slivenicama u srpskom jeziku, kao i da li među ovim dvama morfosemantičkim tipovima u tom pogledu postoje razlike. U tu svrhu je iz postojećih izvora srpskih slivenica uzorkovano 358 primera paradigmatskog i neparadigmatskog tipa, s tim da nisu sve one bile deo svake statističke analize koja je izvršena u radu. Rezultati ovog istraživanja pokazuju da je u srpskom jeziku uticaj broja slogova motivnih reči na njihov redosled u paradigmatskim i neparadigmatskim slivenicama statistički značajan, ali ne i frekventnost motivnih reči.

▸ ***Ključne reči*:** srpske slivenice, motivne reči, pozicija, broj slogova, frekventnost.

Preuzeto: 1. 2. 2025.
Korekcije: 20. 10. 2025.
Prihvaćeno: 9. 11. 2025.