



HAL
open science

Predicting stress assignment in lexical blends: the case of English and Serbian

Vincent Renner, Gordana Lalić-Krstin

► **To cite this version:**

Vincent Renner, Gordana Lalić-Krstin. Predicting stress assignment in lexical blends: the case of English and Serbian. ELLSIIR International Conference, 2009, Belgrade, Serbia. pp.Pages 265-273. hal-00712647

HAL Id: hal-00712647

<https://hal.science/hal-00712647>

Submitted on 27 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vincent Renner & Gordana Lalić-Krstin

University of Lyon – University of Novi Sad

**PREDICTING STRESS ASSIGNMENT IN LEXICAL BLENDS:
THE CASE OF ENGLISH AND SERBIAN**

“Despite the exceptions, it seems better to attempt to produce some stress rules (even if they are rather crude and inaccurate) than to claim that there is no rule or regularity in English word stress.” Peter Roach

Abstract

This article tackles the question of stress assignment for lexical blends. The analysis is based on the study of the lexical stress pattern of 268 English and 253 Serbian blends. The combination of the stress pattern homology rule and the last stressed nucleus rule leads to correct overall predictions in 81% of cases in English and 96% of cases in Serbian, and up to respectively 89% and 98% for the prototypical (and largest) formal subtype of lexical blend, coined by hind clipping of the first source lexeme and fore clipping of the second one.

1. Introduction

This research aims to provide empirical insight into the question of stress assignment for lexical blends. The topic has gone largely unaddressed in the blending literature and the few claims that have been made have limitations and lack a sufficient quantitative basis. Our objective is to formulate a freely applicable prediction rule which will yield the best possible results on a large sample of English blends, and then to turn to Serbian to assess its usefulness beyond English.

2. Defining lexical blending

The definitions of lexical blending encountered in the literature vary to quite a considerable extent. López Rúa (2004), Ronneberger-Sibold (2006), Lehrer (2007) and Brdar-Szabó and Brdar (2008) consider that the clipping of at least one source lexeme is sufficient to determine membership, and they include complex lexemes such as *cyborg* (< *cybernetic* + *organism*) and *contrail* (< *condensation* + *trail*) in the category, but other linguists object to this minimal definition:

- some add a formal criterion for membership: for Bauer and Huddleston (2002) and Bauer (2006), a complex lexeme is unambiguously a blend only if its two source lexemes have been clipped; for Bat-El (2006), a complex lexeme is a blend only if its two source lexemes have been clipped or (in the case where only one of the source lexemes is clipped) if the right source lexeme has been fore-clipped, as in *skinoe* (< *ski* + *canoe*);
- Grésillon (1984) and Kubozono (1990), followed by Plag (2003), who distinguishes between "*blends*" and "*proper blends*", add a semantic criterion, considering that only those complex lexemes whose truncated source lexemes are in a coordinate relation (i.e. are not in a modifier-head relation) are rightful blends (items such as *motel* (< *motor* + *hotel*) and *sitcom* (< *situation* + *comedy*) are excluded from the category);
- Grésillon (1984), followed by Bankov et al. (1989), adds a morphophonological criterion, considering that only those complex lexemes exhibiting segmental overlap (e.g. *motel* (< *mot* + *otel* < *motor* + *hotel*), *gazunder* (< *gazu* + *under* < *gazump* + *under*)) are rightful blends.

In our opinion, these divergent views are not irreconcilable. One simply needs to adopt a prototype-based definition of lexical blending and consider that the truncation of the two source lexemes, semantic coordination, and segmental overlap are not defining characteristics, but typicality features of the concept. This approach has the advantage of accommodating those complex lexemes which were not formed by internal clipping (i.e. by back-clipping the left source lexeme

and/or fore-clipping the right source lexeme), but display segmental overlap, a characteristic which is specific to lexical blends:

- *botulin + toxin > bot + tox > Botox*
- *modulator + demodulator > mod + dem > modem*
- *teleprinter + exchange > tele + ex > telex.*

Complex lexemes formed by double back-clipping will therefore be dealt with in the study.

3. English blends

3.1. Methodology

The English pre-study data came from a personal collection of 400+ lexical blends which were culled from various dictionaries and glossaries over a number of years. First, blends such as *brunch*, *frings*, *skort*, *smog*, or *Spork* were discarded as stress assignment in monosyllabic lexemes is a non-issue. Second, five dictionaries – *The American Heritage Dictionary of the English Language*, 4th ed., *The Cambridge English Pronunciation Dictionary*, 17th ed., *The Longman Pronouncing Dictionary*, 3rd ed., *The Random House Unabridged Dictionary*, and *The Oxford English Dictionary online* – were consulted for stress pattern information and we only retained the 282 blends which were listed in at least two dictionaries, so as to gauge to what extent phonolexicographers agree or disagree on stress assignment. Disagreement happened to be a minor issue, as only 14 blends out of 282 (or 5 percent) have a disputable stress pattern. There are two cases of disagreement – two blends have two listed stress patterns in most dictionaries (*anthraquinone*, /0001/ and /0010/, and *desipramine*, /0100/ and /0010/), and twelve more have different stress patterns, depending on which dictionary is consulted (e.g. *vocoder*, /100/ and /010/). The 14 blends were excluded from the final list of blends under scrutiny, which contains 268 items.

3.2. Data analysis

The 50 blends resulting from double back-clipping need to be isolated into a separate class as their behavior is atypical and 98 percent regular: with the exception of *ceramal* (< *ceramic* + *alloy*), stress invariably falls on the initial syllable of the blend (e.g. *Amerind*, *hazmat*, *Interpol*, *satnav*).

To predict the stress assignment of the remaining 218 blends, two observations from the blending literature are used as a starting point. Quirk et al. (1985:1583) claim that a "*blend tends to have as a whole the prosodic shape of the untruncated end-part*". This statement having a limited predictive scope because it is by definition only applicable to those blends which have the same number of syllables as their right source lexeme, we suggest to extend the prediction rule to the relation holding between the blend and its left source lexeme. Out of 218 blends, 140 (or 64.2 percent) obey the stress pattern homology rule, which states that a blend will adopt the stress pattern of one of its source lexemes if the two are equisyllabic:

- *digitoxin* /0010/ < *digitalis* /0010/ + *toxin* /10/
- *dockominium* /00100/ < *dock* + *condominium* /00100/
- *silane* /10/ < *silicon* /100/ + *methane* /10/
- *witticism* /1000/ < *witty* /10/ + *criticism* /1000/

The second observation comes from Bat-El (1996:321), who claims that in Modern Hebrew blends, "*main stress [...] falls on the rightmost stressed syllable that surfaces*". To take into account the fact that the process that clips source lexemes into fracto-lexemes (also called lexical splinters) does not always preserve syllable boundaries (e.g. *elevator* + *aileron* > *elev* + *on* > *elevon*; *huge* + *monstrous* > *hu* + *mon* + *g* + *ous* > *humongous*), our prediction rule will consider the rightmost stressed source lexeme nucleus (and not syllable) that surfaces in the blend. This rule has a higher predictive potential than the previous one as it is applicable to any blend that has preserved at least one stressed nucleus in the blending process. Out of 218 blends, 160 (or 73.4 percent) follow the rightmost stressed nucleus rule:

- *altazimuth* < *altitude* + *azimuth*

- *fantabulous* < *fantastic* + *fabaulous*
- *humongous* < *hu* + *mon* + *g* + *ous* < *huge* + *monstrous*
- *planetesimal* < *planet* + *infinitesimal*

This second prediction rule yields markedly better results and resolves conflicts arising from the fact that a blend may be equisyllabic with its two source lexemes, but at the same time have source lexemes which have different stress patterns:¹

- *motel* < *mototor* + *hotel*
- *positron* < *positive* + *electron*
- *rockoon* < *rocket* + *balloon*
- *thyristor* < *thyratron* + *transistor*

The rightmost stressed nucleus rule has, however, its own weaknesses as the blending process does not always preserve stressed nuclei:

- *electret* /010/ < *electricity* + *magnet*
- *Japlish* /10/ < *Japanese* + *English*
- *Norplant* /10/ < *levonorgestrel* + *implant*

More generally, a number of blends do not obey the rightmost stressed nucleus rule, but follow the stress pattern homology rule:

- *acupressure* /1000/ < *acupuncture* /1000/ + *pressure*
- *Medicare* /100/ < *medical* /100/ + *care*
- *Populux* /100/ < *popular* /100/ + *deluxe*

It seems therefore advantageous to formulate a mixed rule, which combines the two aforementioned rules, in order to obtain the highest possible percentage of correct prediction:

- *if a blend does not have the same number of syllables as either of its source lexemes, apply the rightmost stressed nucleus rule;*
- *if it has the same number of syllables as only one of its source lexemes, apply the stress pattern homology rule;*
- *if it has the same number of syllables as its two source lexemes and the source lexemes have the same stress pattern, apply the stress pattern homology rule;*

¹ The rightmost stressed nucleus rule does not, however, account for the stress pattern of *camcorder* (< *camera* + *recorder*).

- *if it has the same number of syllables as its two source lexemes and the source lexemes do not have the same stress pattern, apply the rightmost stressed nucleus rule.*

Out of 218 blends, 177 (or 81.2 percent) obey the mixed rule, which is another marked prediction improvement. By giving precedence to the stress pattern homology rule, the mixed rule fails to correctly predict stress assignment in blends which obey the rightmost stressed nucleus rule but violate the stress pattern homology rule:

- *Amerasian < American + Asian*
- *Ebonics < ebony + phonics*
- *galumph < gallop + triumphant*
- *morphosyntax < morphology + syntax*

Because only four blends out of 218 are affected, this shortcoming is considered a minor one.

To reach a percentage of correct prediction higher than 81 percent, it appears necessary to distinguish between the different formal types of lexical blends. The mixed rule correctly predicts the stress pattern for:

- Type I: 60 percent² of all blends formed by back-clipping of the left source lexeme (e.g. *gazunder* < *gazump* + *under*);
- Type II: all blends³ formed by fore-clipping of the right source lexeme (e.g. *gaydar* < *gay* + *radar*);
- Type III: 88.9 percent⁴ of all blends formed by back-clipping of the left source lexeme and fore-clipping of the right source lexeme (e.g. *Oxbridge* < *Oxford* + *Cambridge*);
- Type IV: 71.4 percent⁵ of the blends which are formed by other types of clipping (e.g. *bulimarexia* < *bulimia* + *anorexia*; *Lastex* < *elastic* + *latex*).

² 33 items out of 55.

³ 45 items.

⁴ 80 items out of 90.

⁵ 20 items out of 28.

4. Serbian blends

4.1. Methodology

The Serbian data was taken mostly from Bugarski (2006) and Lalić-Krstin and Halupka-Rešetar (2007), with the addition of a small number of newly coined lexemes taken mostly from advertisements and the media. Blending is a relatively new lexeme-formation process in Serbian and as such, it has not been studied extensively. No blends have yet entered dictionaries, nor are there any studies of their pronunciation that we are aware of. That is why the stress patterns of Serbian blends had to be determined by eliciting help from native speakers of the language.

General stress rules of Serbian say that lexemes can be stressed on any syllable but the ult (with the exception of some loanwords). This means that besides monosyllabic lexemes, disyllables had to be excluded from the analysis, because they are automatically stressed on the first syllable. Further restrictions govern the type of tone individual syllables can have: monosyllables can only have a falling tone and polysyllabic lexemes stressed on a medial syllable can only have a rising tone, but neither of these restrictions relates to the position of stress as such. We also had to exclude blends that are homophonous with one of their source lexemes, i.e. which are recognizable as blends only in their written form. This seems to be a very popular model as there are quite a few of these graphic blends: *rePRESSija* (< *represija* + *press*), *SPOmenar* (< (*SPO* + *spomenar*)), *nagRADIO* (< *nagradio* + *radio*). Blends that are unpronounceable (e.g. *čuDSSni* (< *čudesni* + *DSS*)) were dismissed too, and so were those whose pronunciation could not be determined with certainty, either because informants disagreed or because they were uncertain of how a blend should be pronounced (e.g. *Azbukvar* (< *azbuka* + *bukvar*), which was stressed on the first and the second syllable). Finally, there were a number of Serbian blends where one of the source lexemes (and it was always the first source lexeme) was not completely retrievable. In such cases, what is usually retrievable is the root morpheme(s) whereas derivational and/or inflectional morphemes are not. Although this does not generally affect the

meaning of the blend, it can have a bearing on the number of syllables of the source lexeme (e.g. *krenzije* < *kreten* (2 syllables) / *krenski* (3 syllables) + *prezije*). Where this was the case, the blend was not included in the data to be analyzed. This reduced the list to 253 items.

4.2. Data analysis

The application of the aforementioned stress assignment rules to Serbian blends produces results markedly similar to those obtained for English. Out of 253 blends, 201 (or 79.4 percent) follow the stress pattern homology rule:

- *lovčanik* /010/ < *lova* /10/ + *novčanik* /010/
- *bleferendum* /0010/ < *blef* + *referendum* /0010/
- *kulturajci* /0010/ < *kultura* /010/ + *policajci* /0010/
- *mafijavelizam* /000010/ < *mafija* /100/ + *makijavelizam* /000010/.

As with English, the rightmost stressed nucleus rule proves to be a better predictor, with 93.3 percent of correct stress assignment (236 items out of 253):

- *Anarheologija* < *anarhija* + *arheologija*
- *Frikombinovati* < *Frikom* + *kombinovati*
- *Kablendar* < *kablovska* + *kalendar*
- *profaraon* < *prof* + *araon*

However, as mentioned above, this rule poses a problem in cases where neither stressed nucleus is preserved. Besides, it does not work when in conflict with the general stress assignment rule which does not allow the stress to fall on the ult. In the few cases where the application of the rightmost stressed nucleus rule would result in the complex lexeme being stressed on its last syllable, the rule is blocked by the general rule:

- *Mačker* (not *Mačker*) < *mačka* + *ker*
- *profesmor* (not *profesmor*) < *profesor* + *smor*
- *SeneGOL* (not *SeneGOL*) < *Senegal* + *gol*
- *Karlohag* (not *Karlohag*) < *Karlobag* + *Hag*

Incidentally perhaps, in all these cases, the blend is equisyllabic with one of the source lexemes, which makes it eligible for the stress pattern homology rule.

The prediction can be improved if the two rules are combined into the mixed rule outlined above for English. As many as 245 blends out of 253 (or 96.8 percent) obey the mixed rule, with the following breakdown into formal types:

- Type I: all blends⁶ formed by back-clipping of the left source lexeme (e.g. *žvazbuka* < *žvaka* + *azbuka*);
- Type II: 96.6 percent⁷ of all blends formed by fore-clipping of the right source lexeme (e.g. *kršilak* < *krš* + *kadilak*);
- Type III: 97.8 percent⁸ of all blends formed by back-clipping of the left source lexeme and fore-clipping of the right source lexeme (e.g. *lovčanik* < *lova* + *novčanik*);
- Type IV: 90 percent⁹ of the blends which are formed by other types of clipping (e.g. *profaraon* < *profa* + *faraon*; *inter-net-vju* < *internet* + *intervju*).

5. Conclusion

Although both the stress pattern homology rule and the rightmost stressed nucleus rule yield quite successful predictions regarding the position of stress in English and Serbian lexical blends, the combination of these two rules into a newly formulated mixed rule produces a marked improvement which is summarized in Table 1.

⁶ 44 items.

⁷ 86 items out of 89; the only exceptions are *stradunanje* (< *Stradun* + *stradanje*), *vikendvizija* (< *vikend* + *televizija*), and *Festovizija* (< *Fest* + *televizija*), provided that in the latter two cases the right source lexeme is pronounced *televizija* (the prescribed pronunciation), and not *televizija* (the more frequent pronunciation).

⁸ 88 items out of 90; the only exceptions are *Montevizija* (< *Montenegro* + *televizija*) and *demokratūra* (< *demokratija* + *diktatūra*).

⁹ 27 items out of 30; the only exceptions are *dafimentalna* (< *Dafiment* + *mentalna*), *Balcancan* (< *Balkan* + *cancan*), and *inter-net-vju* (< *internet* + *intervju*).

Table 1. Percentages of correct prediction of stress assignment for lexical blends

	English	Serbian
Stress pattern homology rule	64.2	79.4
Rightmost stressed nucleus rule	73.4	93.3
Mixed rule	81.2	96.8
Mixed rule for Type I blends	60	100
Mixed rule for Type II blends	100	96.6
Mixed rule for Type III blends	88.9	97.8
Mixed rule for Type IV blends	71.4	90

The results in the two languages are relatively robust, which hints at the possible cross-linguistic value of the mixed prediction rule.

References

- Bankov, D. et al. (1989). Contribution à l'étude des mots-valises ou "mixonymes". *Supostavitelno Ezikoznanie*, 14 (5), 8-11.
- Bat-El, O. (1996). Selecting the Best of the Worst: The Grammar of Hebrew Blends. *Phonology*, 13 (3), 283-328.
- Bat-El, O. (2006). Blend. In: K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd ed., Vol. 2, Amsterdam: Elsevier, 66-70.
- Bauer, L. (2006). Compounds and Minor Word-Formation Types. In: B. Aarts and A. McMahon (eds.), *The Handbook of English Linguistics*, Oxford: Blackwell, 483-506.
- Bauer, L. and R. Huddleston (2002). Lexical Word-Formation. In: R. Huddleston and G. K. Pullum (eds.), *The Cambridge Grammar of the English Language*, Cambridge: Cambridge University Press, 1621-1721.
- Brdar-Szabó, R. and M. Brdar (2008). On the Marginality of Lexical Blending. *Jezikoslovlje*, 9 (1/2), 171-194.
- Bugarski, R. (2006). *Žargon: lingvistička studija*. 2nd ed. Beograd: Biblioteka XX vek / Krug.

- Grésillon, A. (1984). *La Règle et le monstre : le mot-valise. Interrogations sur la langue, à partir d'un corpus de Heinrich Heine*. Tübingen: Niemeyer.
- Kubozono, H. (1990). Phonological Constraints on Blending in English as a Case for Phonology-Morphology Interface. *Yearbook of Morphology*, 3, 1-20.
- Lalić-Krstin, G. and S. Halupka-Rešetar (2007). Nešto novo o novim slivenicama u srpskom jeziku. *Svet Reči*, 23/24, 26-30.
- Lehrer, A. (2007). Blendalicious. In: J. Munat (ed.), *Lexical Creativity, Texts, and Contexts*, Amsterdam: Benjamins, 115-133.
- López Rúa, P. (2004). The Categorical Continuum of English Blends. *English Studies*, 85 (1), 63-76.
- Plag, I. (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.
- Quirk, R. et al. (1985). *A Comprehensive Grammar of the English Language*. Harlow: Longman.
- Ronneberger-Sibold, E. (2006). Lexical Blends: Functionally Tuning the Transparency of Complex Words. *Folia Linguistica*, 40 (1-2), 156-181.