Serbo-Croatian Lexicon: Inherited, borrowed, and engineered

Danko.Sipka@asu.edu (Arizona State University)
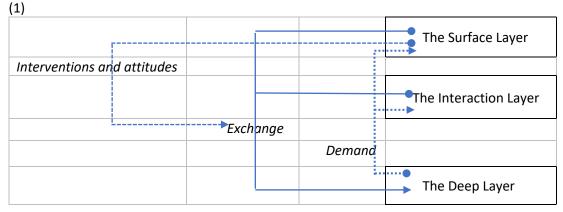
1. Introduction

A thought experiment: The year is 2050, our world comes to its end, all people are gone and all of their artifacts. The only surviving artifact is a section of a library in Belgrade, Serbia, with various dictionaries and encyclopedias – most of them Serbo-Croatian, monolingual and bilingual. Three years later, aliens recolonize the Earth. What would they be able to tell about
Serbo-Croatian and its cultures based on the data from the dictionaries?

Indo-European > Proto-Slavic > South-Slavic > Serbo-Croatian (Serbian, Croatian, Bosnian, *Montenegrin*) – Kordić (2010), Brozović and Ivić (1988), Browne (1993), Sussex and Cubberly (2006), Gutschmidt et al. (2014), Greenberg (2004).

2. Three lexical layers

    a. Russian *синий* 'deep blue'*: голубой* 'grayish blue' versus Serbo-Croatian *плав* 'blue',
    b. Russian *подушка* 'pillow' versus Serbo-Croatian *јастук* 'pillow',
    c. Polish *tłok* 'piston' versus Serbo-Croatian *клип* 'piston'.

    a. The deep layer – the lexical expression of cultural identity that features the highest degree of stability, which does not seem to be directly influenced by historical or political developments, or the will of the community of speakers.
    b. The interactional layer – the lexical expression of cultural identity, which is a result of direct or indirect intercultural communication.
    c. The surface layer – the lexical expression of cultural identity created by conscious intervention of linguistic and political elites within the community of speakers.
(1)

Three Lexical Layers: Interaction

3. Inherited lexicon

Proto-Slavic (East, West, and South Slavic)

(2)

Category	Number	Percent
Preservation in all three groups	471	27.7
Change in all three groups	436	25.6
Loss in all three groups	20	1.2
A combination of loss, change and/preservation	776	45.5
Total	1703	100.0

(3)

	East Slavic		West Slavic		South Slavic	
	Number	Percent	Number	Percent	**Number**	**Percent**
Loss	175	10.3	204	12.0	**168**	**9.9**
Change	707	41.5	707	41.5	**873**	**51.3**
Preservation	821	48.2	792	46.5	**662**	**38.9**
Total	1703	100.0	1703	100.0	**1703**	**100.0**

Loss – *volsti* 'rule' is not attested in South Slavic languages
Change – see narrowing, transfer, widening below
Preservation – *vòńa; *vòńь* 'smell' yields *woń* 'smell' in Polish

In 1:1 comparison, the South Slavic group is equidistant from the other two groups, which are closer to each other than to the South Slavic group.

(4)

	East		West		South	
Narrowing	357	52%	367	50%	**478**	**47%**
Transfer	277	40%	316	43%	**452**	**45%**
Widening	55	8%	51	7%	**82**	**8%**
Total	689	100%	734	100%	**1012**	**100%**

narrowing (specification) - Czech *jahoda* 'strawberry' from *àgoda* 'berry'
transfer – Serbo-Croatian *slatina* 'mineral spring' from *soltina* 'salt marsh'
widening (generalization) - Slovak *slatina* 'marsh' from *soltina* 'salt marsh'

(5)

Category	Frequency	Percent	Example (Serbo-Croatian – English)
Social Interaction	172	27.3	Prijatelj = *friend, follower, lover, son's (or daughter's) father-in-law*
Institutions	65	10.3	Pošta = post office, mail
Measures	159	25.2	Pair = pair, couple, match (to someone), partner (in school)
Logical Operators	58	9.2	A = and, but, while
Body Parts	27	4.3	Ruka = hand, arm
Other	149	23.7	
Total	630	100	

(6)

Category	Frequency	Percent	Example (English – Serbo-Croatian)
Rural World	227	40	Bed = krevet, postelja, ploha, ploča, podloga, ležaj, postolje, gredica, lijeha, temelj, podloga…
General Features	82	14.5	Open = otvoren, slobodan, nezauzet, nepokriven, otkrit, izložen, nesložen, nesavijen, probušen, porozan, isprekidan, slobodan, nevezan, blag, očit, jasan, javan, otvoren, iskren, darežljiv, gostoljubiv, otvoren, tekući, neodlučan, neriješen, pristupačan
Negative Features	30	5.3	Bad = loš, bezvrijedan, nevaljao, škodljiv, pokvaren, neugodan, odvratan, zao, opak, loša zdravlja, bolestan, bolan, nedovoljan, nedostatan, težak, jak, nemio, grdan
Movement	54	9.5	Road = sidrište, rada, cesta, drum, put, staza, tračnice, željeznička pruga, putovanje
Other	174	30.7	
Total	567	100	

4. Borrowed lexicon

Geographical contact (subsequent vertical contact, concurrent vertical contact, horizontal contact) vs. Cultural influence
Subsequent vertical contact – Latin and pre-Latin substratum (nature, animal husbandry, etc. *bosiljak* 'basil', *čuka* 'hillock')
Concurrent vertical contact
       Turkish, some ultimately Arabic or Persian (daily life and spirituality: *čarapa* 'sock', *sevdah* 'yearning for love'
       German (daily life, technology, adminstration: *kofer* 'suitcase', *šina* 'railroad rail')
          Dalmatian, Venetian (daily life, administration: *kapitan* 'captain', *štatut* 'statute')
Horizontal contact
       Hungarian (daily life *lopov* 'thief', *ašov* 'spade')
Cultural influences (spontaneous or by lexical engineering)
       Church Slavonic, Russian, Czech (mostly abstract, administrative, etc.: *prosvjeta* 'education', *činovnik* 'clerk', *časopis* 'journal')
       Graeco-Latin (directly or via German or French, mostly abstract and technological: *kultura* 'culture', *telefon* 'telephone')
       French (cuisine, art, etc. *buke* 'bouquet', *uvertira* 'ouverture')
       English (technology, popular culture: *tenk* 'tank', *džez* 'jazz')
       Italian (arts: *libreto* 'libretto', *aria* 'aria')

Generality index = metalanguage presence in the general lexicon minus average presence in the general lexicon
Presence index = percentage of the metalanguage/metafield in question minus average metalanguage/metafield percentage

(7)

Source	*Metalanguage Presence Index*	*Metalanguage Generality Index*
European – Classical	35.2	-8.9
European – Germanic	1.6	-8.5
European – Romance	-0.3	-2.8
Average for European Sources	*12.2*	*-6.7*
Near Eastern	-6.0	6.5
Other	-15.0	10.6
Slavic	-15.6	3.17

Metalanguage distribution, based on Klajn and Šipka (2006)

(8)

Field	*Metafield presence index*
Arts	-2.9
Belief systems	-3.8
Sciences – Natural	14.0
Sciences – Social	6.0
Sports	-4.9
Vocational – production	-5.1
Vocational – services	-3.1

Metafield distribution, based on Klajn and Šipka (2006)

5. Engineered lexicon

Lexical introduction vs. Lexical refereeing – purism as a merger of the two (Thomas 1988 and Langston and Peti-Stantic 2014)

Standard vs. dialects

Ethnic variants
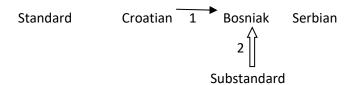
Normative and cryptonormative refereeing

(9)

Group	Examples	Number	Percent
Ekavian (Serbian-Serbian disctinction)	*međususedski, mega-uspeh, proterivač, svetskost*	1880	4.19%
Croatian	*bojišnica, božićnica, uradak, zuboliječnik*	632	1.41%
Bosniak	*lahko, mehko, poselamiti, sadaka, šehid*	243	0.54%
Radical Croatian	*datkovni obradnik, munjkovni, objasnitba, rednik*	203	0.45%
Radical Bosniak	*dekika, greb, hudovica, krhat, zijehati*	68	0.15%
Serbian	*instalisati, naduvan, paraklis, pasuljast*	61	0.14%
Radical Serbian	*đeneral, sohran, toržestven, vaznositi*	20	0.04%
Marked		*3107*	*6.92%*
Unmarked		*41772*	*93.08%*
Total		*44879*	*100.00%*

Lexical changes of the 1990s

(10)
Choose if this label means that the word marked by it is non-standard

	yes	no
slang (шатровачки)	78.26% (54)	21.74% (15)
jargon (жаргонски)	68.12% (47)	31.88% (22)
regional (регионално)	57.97% (40)	42.03% (29)
vulgar (вулгарно)	56.52% (39)	43.48% (30)
old literary style (старокњижевно)	56.52% (39)	43.48% (30)
archaic (архаично)	55.88% (38)	44.12% (30)
folksy (народски)	54.41% (37)	45.59% (31)
school student speech (ђачки)	51.47% (35)	48.53% (33)
colloquial (разговорно)	49.28% (34)	50.72% (35)
familial/homely (фамилијарно)	47.06% (32)	52.94% (36)
pejorative (пејоративно)	45.59% (31)	54.41% (37)
derisive (подругљиво)	34.78% (24)	65.22% (45)
contemptuous (презриво)	30.43% (21)	69.57% (48)
ironic (иронично)	27.54% (19)	72.46% (50)
facetious (шаљиво)	27.94% (19)	72.06% (49)

Primary and secondary normative labels

(10)

Standard   Croatian 1 → Bosniak   Serbian

2 ⇑

Substandard

Examples

1. *slavenski, općina, definirati, listopad*
2. *truhao, mehak, babo, amidža, šehid*

(11)

(tokens)	Secular corpus	Islamic corpus
Total	997149	997861
Bosniak	754	17555
Percentage	0.076%	1.759%

R= 0.0883, p=0.00
Newly Introduced Bosniak words: Corpus distribution comparison

(12)

Total	2824	
Islamic	373	13.21%
General	177	6.27%

Newly introduced Bosniak words: Percentage used in the two corpora

1. References

Browne, Wayles. 1993. Serbo-Croat in Comrie, Bernard and Greville G. Corbett (eds.). *The Slavonic languages*. London and New York: Routledge. pp. 306-387.

Brozović, Dalibor and Pavle Ivić. 1988. *Jezik. Srpskohrvatski/hrvatskosrpski, hrvatski ili srpski*. Zagreb: Leksikografski zavod "Miroslav Krleža".

Greenberg, Robert. 2004. *Language and Identity in the Balkans Serbo-Croatian and Its Disintegration*, Oxford: OUP.

Gutschmidt, Karl et al. ed. 2014. *Die slavischen Sprachen/The Slavic Languages: An International Handbook of their Structure, their History and their Investigation*, Berlin. De Gruyter Mouton.

Klajn, Ivan and Milan Šipka. 2006. *Veliki rečnik stranih reči*. Novi Sad: Prometej.

Langston, Keith and Anita Peti-Stantić. 2014. *Language Planning and National Identity in Croatia*. Houndmills, New York: Palgrave Macmillan.

Kordić, Snježana. 2010. *Jezik i nacionalizam*. Zagreb: Durieux.

Sussex, Ronald and Paul Cubberly. 2006. *The Slavic Languages*. Cambridge: Cambridge University Press.

Thomas, George. 1988. *The Impact of the Illyrian Movement on the Croatian Lexicon*. München: Verlag Otto Sagner.