

A thick black L-shaped frame surrounds the text. The top-left corner is a horizontal bar extending to the right, then a vertical bar extending downwards. The bottom-right corner is a horizontal bar extending to the left, then a vertical bar extending upwards.

# THE SOUTH SLAVIC DIALECT CONTINUUM

Political divisions driving linguistic change?

Joseph Rhyne (jtr92@cornell.edu)



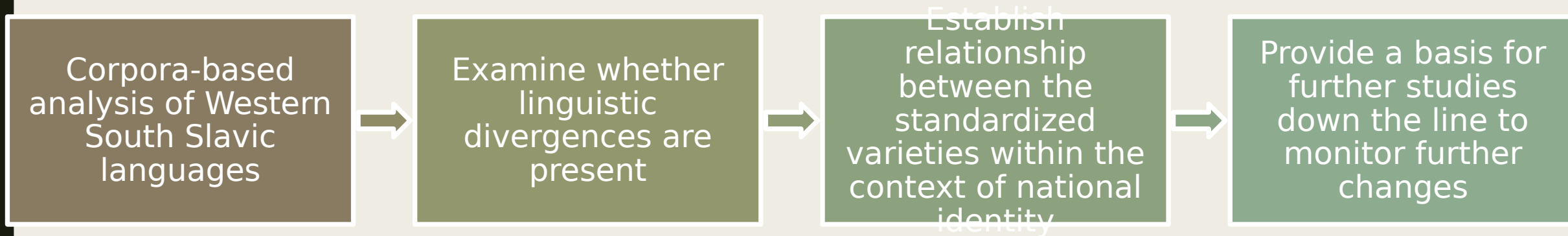
# How do we define languages?

- Mutual intelligibility?
- Shared Standard?
- Sociolinguistic factors?
- I-language vs. E-language (Chomsky 1986)

# Language and Identity

- Crystal (1987): language is “the systematic, conventional use of sounds, signs, or written symbols in a human society for communication and self-expression.”
- External factors play as much a role as linguistic ones (Alexander 2006, Langston & Peti-Stantic 2003)
- Language exists both within the mind of individual speakers and within a speech community
- One cannot exist without the other
- Language reinforces identity and vice versa

# Current Study



# Political History



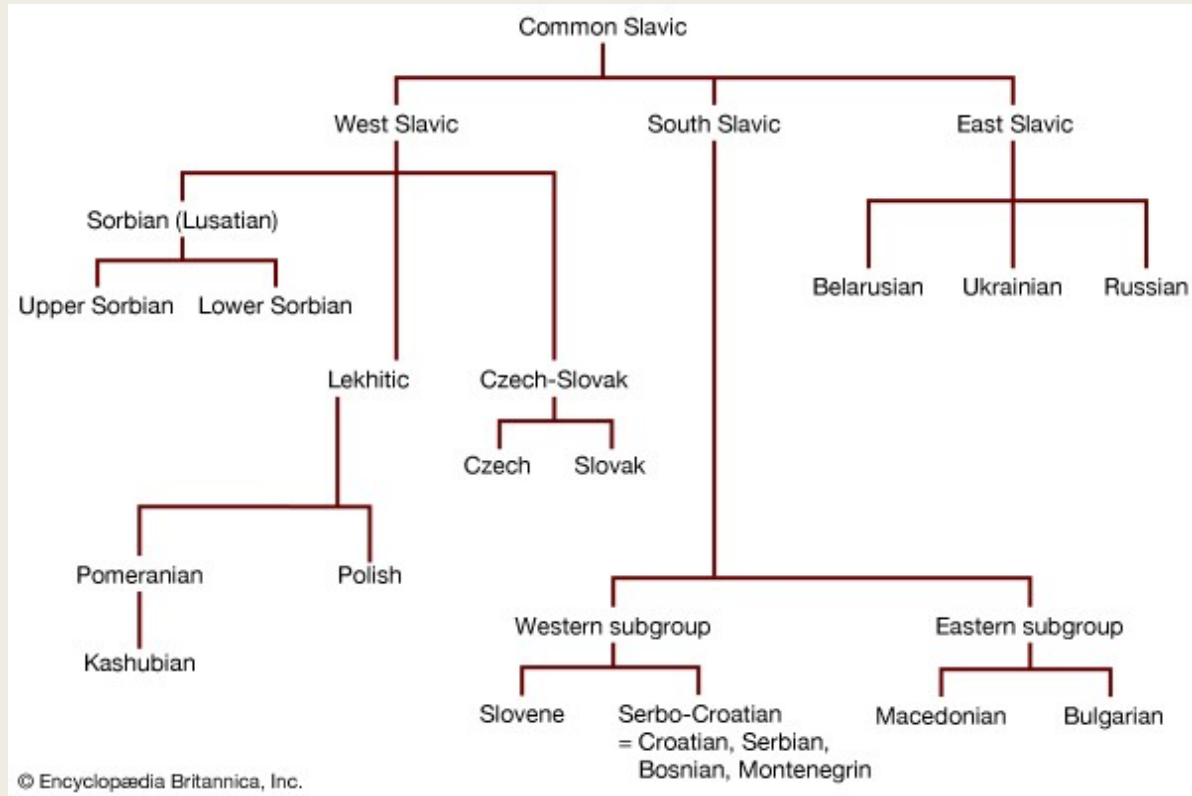
- The area is a complex of religious and political tensions
  - *Manifested in linguistic attitudes today*
- Orthodox church, Catholic church, Islam
- Area dominated by Ottoman and Austro-Hungarian Empires in the 19<sup>th</sup> c.
- Eventual unification under Yugoslavia

# Political History



- Yugoslavia collapsed and disintegrated in the 90's
- Military conflicts, civil war, genocide
- Each new nation sought to reinforce their newfound national identities
  - *Goes back to European nationalism of the 19<sup>th</sup> and early 20<sup>th</sup> c.*
  - *One nation, one language: Language as a tool for shaping these identities*
- Conflict and controversy remains

# SLAVIC LANGUAGES



# The intersection of politics and language

- Separate literary traditions before 1850s
- Serbian and Croatian linguistic unification to combat Turkish and Austro-Hungarian influence
  - *Standardization of the language varieties*
- Unification under Yugoslavia → Unification of a “Serbo-Croatian” language
  - *Novi Sad Agreement: creation of standard Serbo-Croatian*
  - *Attempt to build a single ethnic, linguistic and national identity for Yugoslavia*



# WESTERN SOUTH SLAVIC DIALECTS



Distribution of central South Slavic dialects before 16th century migrations:

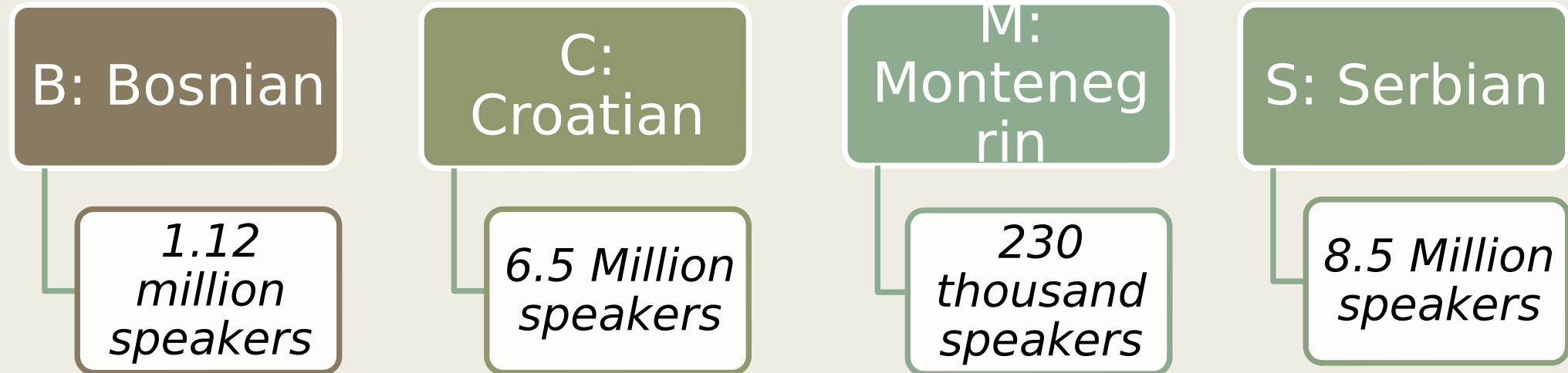
- - Chakavian
- - Kaykavian
- - Western Shtokavian
- - Eastern Shtokavian
- - Torlakian

Note: countries and cities presented in this map are those from the early 21st century.



# SOUTH SLAVIC TODAY

# BCSM: The present situation



## Standard Languages

basis for  
Standard  
Croatian



basis for  
Standard  
Bosnian

basis for  
Standard  
Serbian (in all  
of Serbia)

transition from  
Shtokavian to  
Macedonian and  
Bulgarian

Note: countries and cities presented in this map are those from the early 21st century.

# BCMS: STANDARDS

# Origins of the Modern Languages?

- Source of debate
- Unified language of Serbo-Croatian disintegrated into successor languages in the 1990s as Yugoslavia disintegrated into successor states (Alexander 2006, Greenberg 2004)
- Modern languages are a continuation of the distinct languages that existed before the artificial unification under Serbo-Croatian (Langston and Peti-Stantic 2003)

# BCMS: linguistic differences

- Examples of differences taken from Langston & Peti-Stantic 2014, Popovic & Ljubescic 2015, Tiedemann & Ljubescic 2012, and Ford 2002
- By no means comprehensive
- Scripts:
  - *Bosnian and Montenegrin: both Cyrillic and Latin, with Latin being favored*
  - *Croatian: Latin only*
  - *Serbian: Cyrillic is the official one, but Latin is widely used in media and on the internet*

# BCMS: Linguistic Differences

## ■ Outcomes of *jat*:

- *Croatian and Bosnian: ije or je, e.g. dijete 'child,' htjeti 'to want'*
- *Serbian: e, e.g. dete, hteti*

## ■ Outcome of <h> /x/:

- *Croatian and Bosnian typically maintain it in all positions*
- *Serbian consistently replaces it with /v/ after /u/, and loses it elsewhere*
- *Cf. kuhati vs. kuvati 'to cook'*

## ■ Final /r/:

- *Serbian drops it while it is kept elsewhere*
- *Cf. jučer vs. juče 'yesterday'*

# BCMS: Linguistic Differences

## ■ Spreading of /x/:

- *Bosnian lahko vs. lako* 'easy'
- *kahva vs. kava, kafa* 'coffee'

## ■ Vowel alternations:

- *Croatian točno vs. Serbian tačno* 'accurate'

## ■ Relative/Interrogative Pronouns

### - 'what:'

- Serbian and Bosnian use *što* for the relative and *šta* for the Interrogative
- Croatian uses *što* for both

### - 'who:'

- Croatian uses *tko*
- Serbian and Bosnian use *ko*



# BCMS: Linguistic Differences

## ■ Borrowings:

### - *Proper Nouns:*

■ Transliterated In Serbian: *Vašington*

■ Original orthography in Bosnian and Croatian: *Washington*

### - *Borrowed nouns frequently adapted into different declensions:*

■ *minuta: Fem, II Decl in Croatian*

■ *minut: Masc, I Decl in Serbian*

# BCMS: Linguistic Differences

## ■ Lexicon:

### - *Month names:*

■ Serbian has the Latin-based: *januar, april*

■ Croatian has Slavic-based: *sijenčanj, travanj*

### - *Numerous differences in Lexical items*

■ Croat *kruh* vs. Serb *hleb* 'bread'

■ Croat *tvornica* vs. Serb *fabrika* 'factory'

■ Croat *gospodarstvo* vs. Serb *ekonomija* 'economy'

# BCMS: Linguistic Differences

## ■ Derivational morphology:

- *Often a difference in affix used to derive feminines, diminutives, agentives and other nominals*
- *Agentive:*
  - Croatian tends to use *-telj*
  - Bosnian tends to use *-lac*

## ■ Prepositions:

- *s 'with:'*
  - Appears as *sa* before certain environments in Croatian
  - *sa* generalized in Serbian
- *k 'to:'*
  - Appears as *ka* in certain environments in Croatian
  - *Ka* generalized in Serbian

# BCMS: Linguistic Differences

## ■ Modal constructions:

- *Modal+da+Present used throughout Serbian:*
  - *Moram da radim* 'I have to work'
  - *Mogu da radim* 'I can work'
- *Modal+Infinitive used in Croatian:*
  - *Moram raditi*
  - *Mogu raditi*

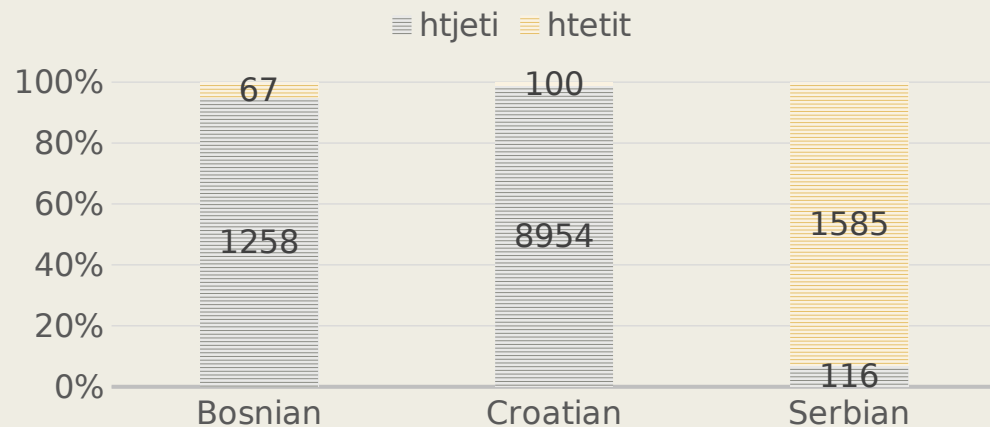
# Methodology: Corpora

- Three different Corpora (Ljubesic & Klubicka 2014):
  - *hrWaC--Croatian Web Corpus: 1.9 billion tokens*
  - *srWaC—Serbian Web Corpus: 894 million tokens*
  - *bsWac—Bosnian Web Corpus: 429 million tokens*
- All are annotated with lemma, morphosyntax, and dependency layers
- No major corpus available for Montenegrin

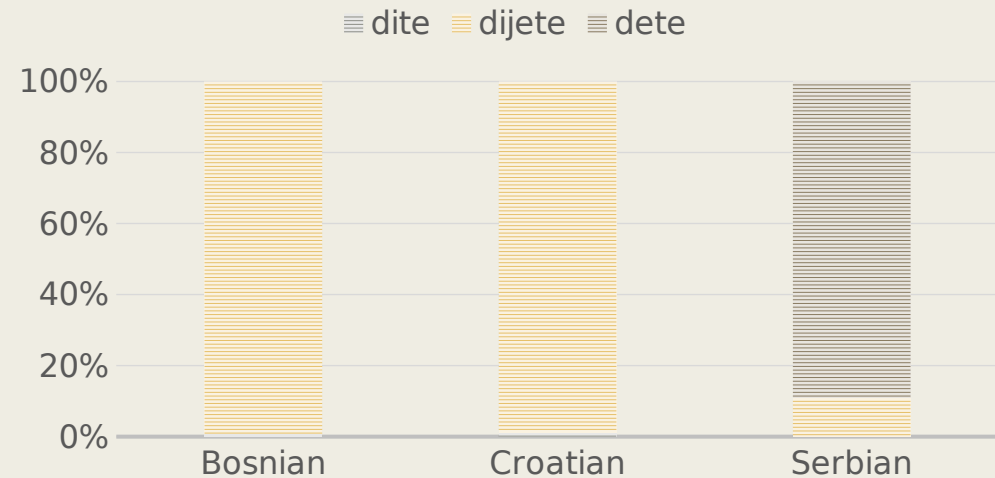
# Results: Outcomes of *jat*

	Bosnian	Croatian	Serbian
dite	371	2790	36
dijete	51104	254164	7096
dete	1	7	58339
htjeti	1258	8954	116
hteti	67	100	1585

## Outcomes of *jat*



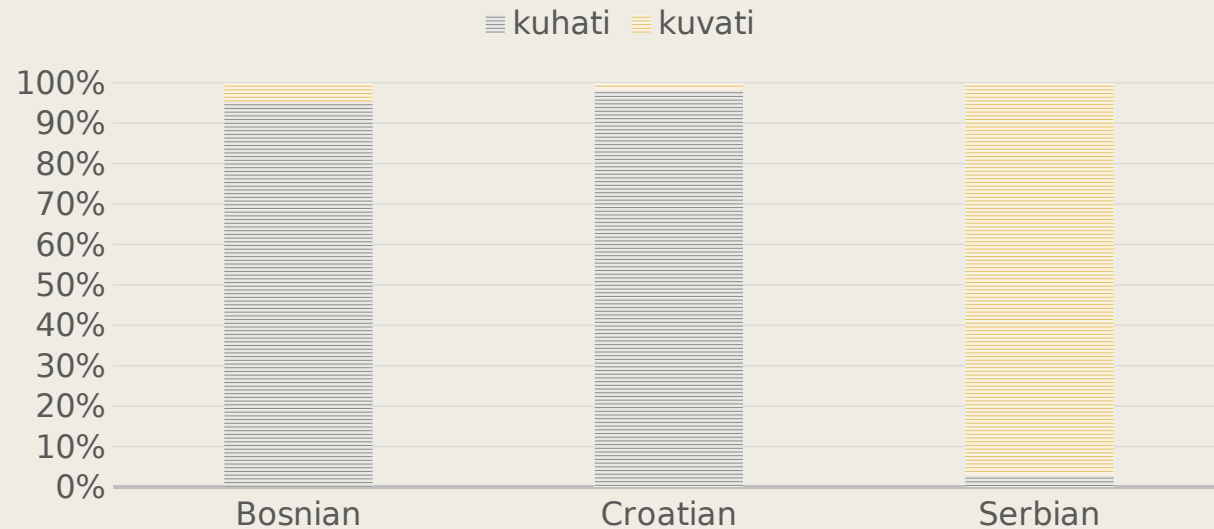
## Outcomes of *jat*



# Results: Outcomes of /x/

	Bosnian	Croatian	Serbian
kuhati	10042	63604	484
kuvati	486	1057	15441

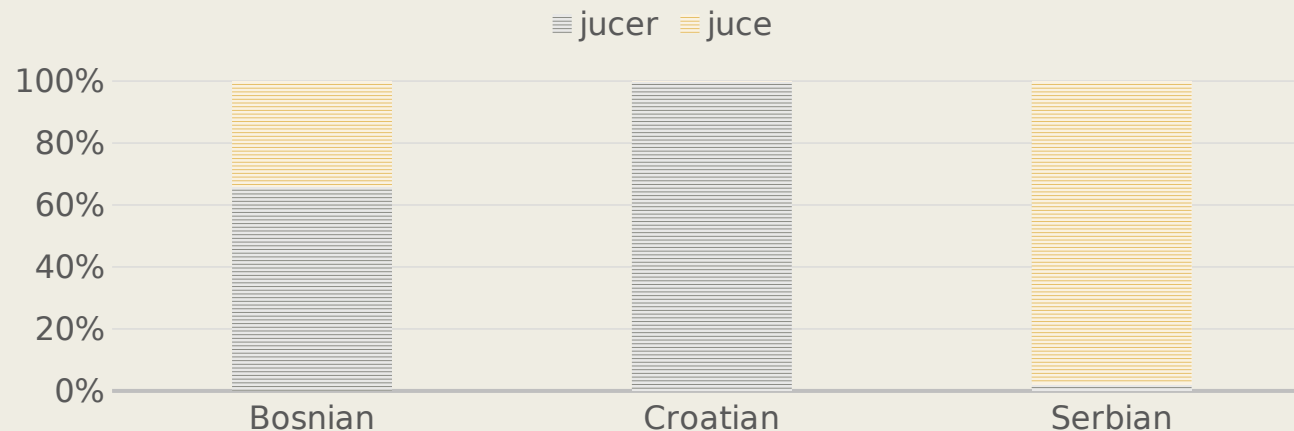
## Outcomes of /x/



# Results: Final /r/

	Bosnian	Croatian	Serbian
jučer	31361	199823	1371
juče	16264	6882	61870
večer	5440	58042	569
veče	4514	4650	18549
također	118374	625038	6889
takođe	44773	14723	285608

Final /r/ loss

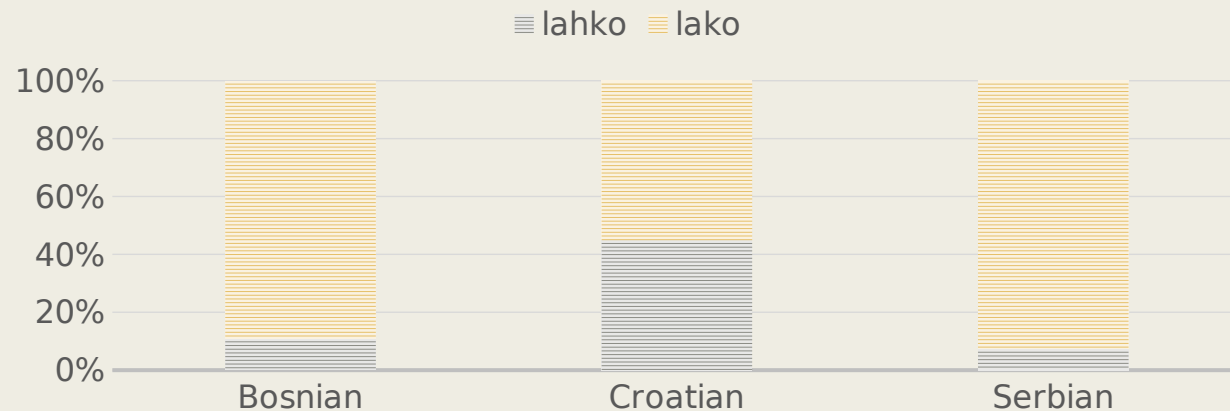




# Results: Spreading of /x/

	Bosnian	Croatian	Serbian
lahko	4169	2016	180
lako	33857	2472	2270

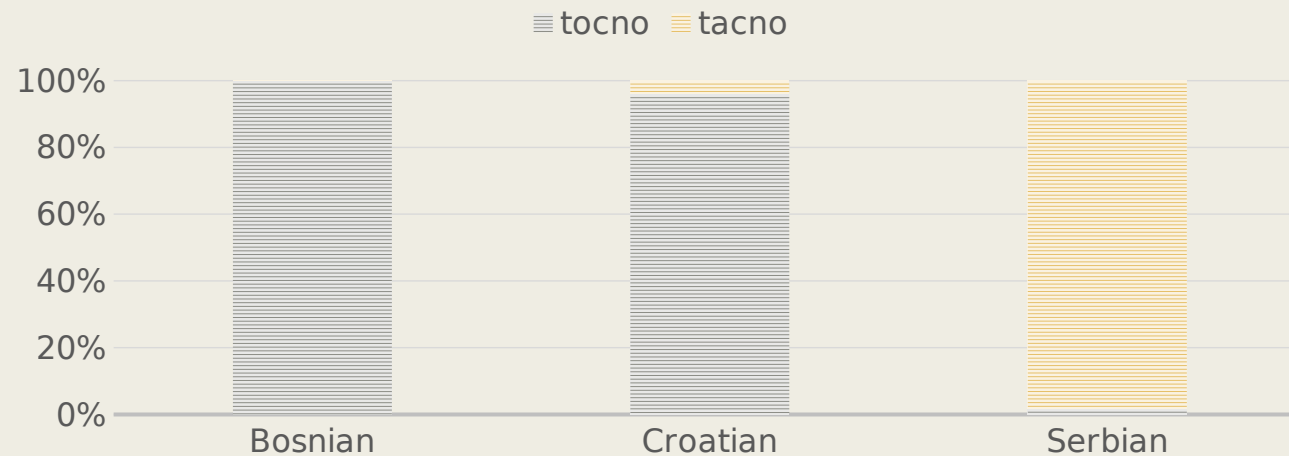
## Spreading of /x/



# Results: Vowel Alternations

	Bosnian	Croatian	Serbian
točno	10255	22441 8	1418
tačno	28798	9534	83141

Vowel Alternations



# Results: Relative Pronouns

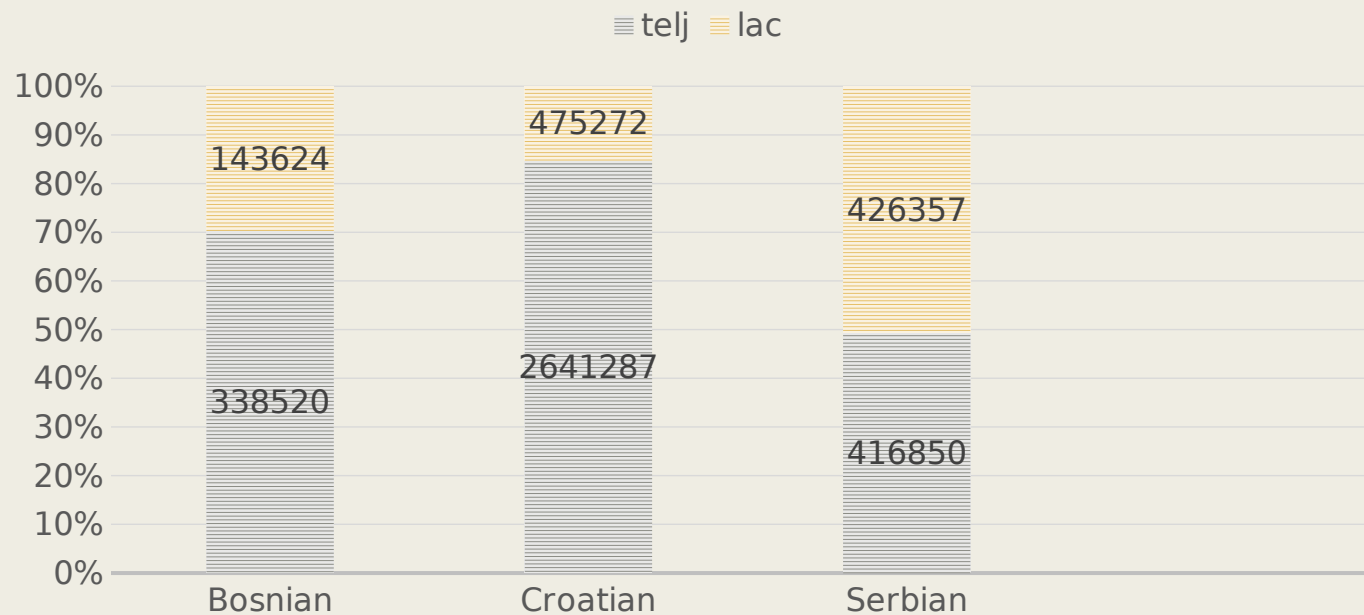
	<b>Bosnia n</b>	<b>Croatia n</b>	<b>Serbian</b>
<b>št a</b>	165670	340387	546631
<b>št o</b>	739876	4167894	1380353
<b>tk o</b>	29428	696362	5202
<b>ko</b>	723	1247	383468

# Results: Lexicon

	<b>Bosnia n</b>	<b>Croati an</b>	<b>Serbia n</b>
<b>sijecanj</b>	7312	12576 1	619
<b>januar</b>	22752	3699	71127
<b>travani</b>	8271	13564 6	875
	<b>Bosni an</b>	<b>Croati an</b>	<b>Serbia n</b>
<b>fabrika</b>	15228	4039	63064
<b>tvornica</b>	7929	86658	989

# Results: Derivational Morphology

e	Bosnian	Croatian	Serbian
telj	338520	264128	416850
lac	143624	475272	426357

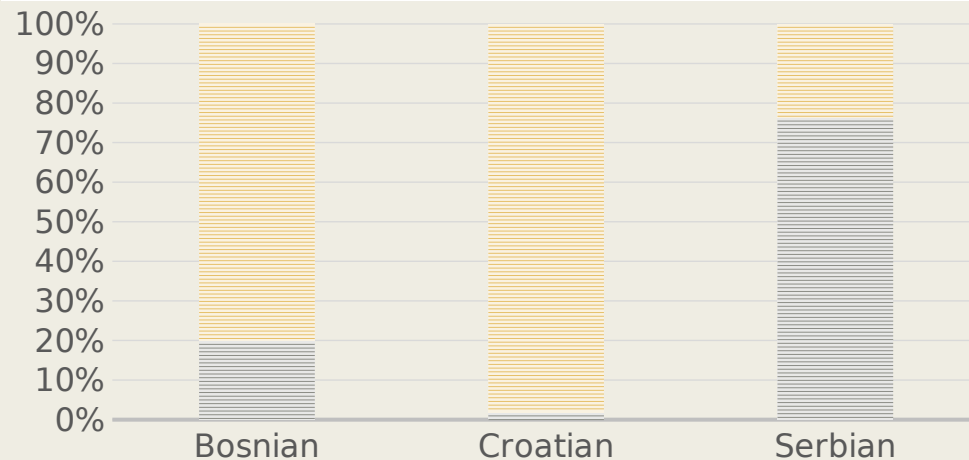


# Results: Prepositions

	<b>Bosnia n</b>	<b>Croatia n</b>	<b>Serbia n</b>
<b>s</b>	842658	690244 2	601111
<b>s</b>	130745	357254	346893
<b>a</b>	1	4	8
<b>k</b>	12793	111202	23194
<b>k</b>			
<b>a</b>	37072	105261	128033

# Results: Modal Constructions

	Bosnian	Croatian	Serbian
<b>moram raditi</b>	2070	12509	1134
<b>moram da radim</b>	467	260	3575
<b>mogu raditi</b>	3925	26753	2723
<b>mogu da radim</b>	1028	465	8892



# Conclusions

- Lots of variation, even within the standardized varieties
- Still some very apparent divergences in overall tendencies
- National identities still being formed
  - *Independence of BCS only within the last three decades*
- With more time, we could see more differences



# Further Study

- This examined the “standardized” versions of BCS
  - *Compare the actual spoken varieties*
- Integrate Montenegrin data
- Explore the linguistic situation of Kosovo
  - *Albanian, Macedonian influence*
- Examine the linguistic structures and literary traditions of BCS before the 19<sup>th</sup> century
- We know the “official” attitudes towards the languages, but how do the actual speakers feel?

# Works Cited

- Alexander, R. 2006. *Bosnian, Croatian, Serbian: A grammar with Sociolinguistic Commentary*. Madison, WI: The University of Wisconsin Press
- Chomsky 1986
- Crystal 1987
- Ford, C. 2002. Language Planning in Bosnia and Herzegovina: The 1998 Bihać Symposium. *SEEJ*, Vol.46, No.2.
- Greenberg, R. 1999. In the aftermath of Yugoslavia's collapse: The politics of language death and language birth. *International Politics*, 36(2), 141-158.
- Greenberg, R. 2004. *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*. Oxford University Press
- Greenberg, R. 2004. *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*. Oxford University Press.
- Klaus 1967
- Ford, C. (2002). Language Planning in Bosnia and Herzegovina, :The 1998 Bihać Symposium. *SEEJ*, Vol.46, No.2.
- Kordić, S. (2008). Jezični Nacionalizam. (review of Stjepko Težak"s „Hrvatski naš (ne)podobni"), *Slovo* 4(17)
- Langston & Peti-Stantic 2003. A language academy by any other name(s): the case of Croatia
- Langston, K. and A. Peti-Stantic. 2014. *Language Planning and National Identity in Croatia*. New York: Palgrave Macmillan.
- Ljubesic, N. and F. Klubicka. 2014. {bs,hr,sr}WaC - Web corpora of Bosnian, Croatian and Serbian. In F. Bildhauer & R. Schafer (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9) @ EACL 2014*. 29-35.
- Popovic & Ljubesic 2015. Tiedemann & Ljubesic 2012
- Ljubojević, A. (2009). Tomorrow People, Where is Your Past? Mapping of post-war national identities in new Yugoslav states.
- Anderson, B. (1983). *The Imagined Community*. NY: Verso