

Zrinka Kolaković  
Institut für Slawistik  
Alpen-Adria Universität Klagenfurt  
Zrinka.Kolakovic@aau.at

Edyta Jurkiewicz-Rohrbacher  
Institut für Slavistik  
Universität Regensburg  
Edyta.Jurkiewicz-Rohrbacher@ur.de

### **BCS pronominal clitic variants *ju* and *je* in standard and colloquial varieties**

Studies of Bosnian, Croatian and Serbian (BCS) clitics (CLs) are mainly characterized by either formal-theoretical or normativist descriptions of their morphology and syntax. At the same time, studies on variation in language use based on big data are extremely rare. However, careful and contrastive analysis of BCS grammar books and language usage guides reveals various differences in the usage of CLs not only between standard BCS varieties, but also within their colloquial idioms. This paper deals specifically with variation in the usage of the third person singular feminine accusative CLs *ju* and *je* ‘her’.

Although the usage of the pronominal CL *ju* is still widespread not only in Croatian but also in Bosnian and Serbian dialects (cf. Lisac 2003; Peco 2007; Okuka 2008), standard Bosnian and Serbian limit it to contexts of suppletion triggered by its combination with the verbal CL *je* ‘is’ and its negated form *nije* (cf. Ridjanović 2012: 434; Piper & Klajn 2014: 97).

(1a) \**On je je video.*

(1b) *On ju je video.* (Piper & Klajn 2014: 97)

‘He saw her.’

(2) *Nije ju video.* (Piper & Klajn 2014: 97)

‘He did not see her.’

In contrast, many Croatian grammarians and authors of language usage guides allow a much wider usage of the CL *ju* and some put it on the same footing as its variant *je* or even prefer it over the pronominal CL *je*. This last case may be seen in examples (3a-b). Ham et al. (2014: 74) give normative preference to the pronominal CL *ju* and consider the usage of the CL *je* in the same context to be unacceptable in the standard Croatian language.

(3a) *Vidim ju.* (Ham et al. 2014: 74)

(3b) \**Vidim je.* (Ham et al. 2014: 74)

‘I see her.’

Piper & Klajn (2014: 97) present the exact opposite position with respect to the state of the art in standard Serbian. Moreover, they explicitly state that using the CL *ju* beyond contexts of suppletion is to be considered a foreign construction in the standard Serbian language (Piper & Klajn 2014: 97).

Considering the state of the art presented in the literature, we address the following research questions:

RQ1: Does the distribution of pronominal CL variants *ju* and *je* in corpora representing BCS standard varieties confirm the expected differences between these language varieties?

RQ2: Does the distribution of pronominal CL variants *ju* and *je* in corpora representing colloquial BCS varieties show any similarities to their distribution in standard varieties?

RQ3: Does the distribution of pronominal CL variants *ju* and *je* within one BCS variety show different patterns in the standard and colloquial registers?

To test the range of this variation we turned not only to traditionally compiled reference corpora with language material which should reflect standard BCS varieties (Santos 1998; Utvić 2011; Čermák & Rosen 2012; Brozović Rončević et al. 2018), but also to {bs,hr,sr}WaC, three massive web corpora (Ljubešić & Klubička 2014), and extracted random samples of size 400 per pronominal CL (*ju* vs *je*) for each variety. In the next step we manually annotated words preceding and following the pronominal CLs *ju* and *je* to determine the impact of context on CL choice. The full list of variables is presented in Table 1 below.

Table 1: List of variables

Class	Variable	Levels
Dependent	1. CL form	<i>ju</i> <i>je</i>
Independent	1. <i>je</i> verb form following the pronominal CL	yes no
	2. word ending with (j)u before the pronominal CL	yes no
	3. word ending with (j)e before the pronominal CL	yes no
	4. BCS variety	Bosnian Croatian Serbian
	5. register	colloquial standard

An overview of preliminary data (1813 sentences in total) is presented in Table 2.

Table 2: Usage of the pronominal CL *ju* outside suppletion contexts in BCS

	standard	colloquial
Bosnian	3%	26%
Croatian	24%	52%
Serbian	0.5%	6%

The data not only confirm the expected differences between the three standards, but also reveal differences between standard and colloquial idioms of the varieties. We expect more interesting insights from logistic regression analysis after the annotation process is completed.

#### References:

- Brozović Rončević, Dunja; Ćavar, Damir; Ćavar, Małgorzata; Stojanov, Tomislav; Štrkalj Despot, Kristina; Ljubešić, Nikola; Erjavec, Tomaž. 2018. Croatian language corpus Rizmica 0.1. *Slovenian language resource repository CLARIN.SI*. <http://hdl.handle.net/11356/1180>.
- Čermák, František; Rosen, Alexandr. 2012. The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics* 17/3. 411–427.
- Ham, Sanda; Mlikota, Jadranka; Baraban, Borko; Orlić, Alen. 2014. *Hrvatski jezični savjeti*, Školska knjiga. Zagreb.
- Lisac, Josip. 2003. *Hrvatska dijalektologija 1. Hrvatski dijalekti i govori štokavskoga narječja i hrvatski govori torlačkoga narječja*. Golden marketing – Tehnička knjiga. Zagreb.
- Ljubešić, Nikola; Klubička, Filip. 2014. {bs,hr,sr}WaC — Web corpora of Bosnian, Croatian and Serbian. *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*. Ur. Bildhauer, Felix; Schäfer, Roland. Association for Computational Linguistics. Gothenburg. 29–35. 10.3115/v1/W14-04.
- Okuka, Miloš. 2008. *Srpski dijalekti*. SKD Prosvjeta. Zagreb.
- Peco, Asim. 2007. *Izabrana djela I–VI. Knjiga I: govori istočne i centralne Hercegovine*. Bosansko filološko društvo. Sarajevo.
- Piper, Predrag; Klajn, Ivan. 2014. *Normativna gramatika srpskog jezika*. Matica srpska. Novi Sad.
- Ridjanović, Midhat. 2012. *Bosnian for foreigners with a comprehensive grammar*. Rabic Publishing Company. Sarajevo.
- Santos, Diana. 1998. Providing access to language resources through the World Wide Web: the Oslo Corpus of Bosnian Texts. *Proceedings of The First International Conference on Language Resources and Evaluation*. ELRA. 475–481.
- Utvić, Miloš. 2011. Annotating the corpus of contemporary Serbian. *INFOtheca: Journal of Informatics & Librarianship*. 12/2. 36–47.