

# Remnants of Serbo-Croatian Lexis in Present-day Croatian

DARIO LEČIĆ

*The University of Sheffield, Department of Russian and Slavonic Studies,  
Jessop West, Sheffield, S3 7RA, United Kingdom, dario.lecic@gmail.com*

---

SCN VII/2 [2014], 49–66

---

Članek se ukvarja s procesom jezikovnega purizma in prisotnostjo srbizmov v hrvaščini. Skušamo ugotoviti, ali govorniki hrvaškega jezika poznajo in uporabljajo leksikalne elemente, za katere na splošno velja, da so bolj značilni za srbščino. V ta namen smo uporabili dva vira podatkov: korpus hrvaškega jezika in anketni vprašalnik. Rezultati kažejo, da se govorniki hrvaškega jezika v veliki meri zavedajo razlik med hrvaščino in srbščino. Večina govorcev sicer pozna pomen leksemov, ki so bili uporabljani v srbohrvaškem obdobju, a jih večinoma ne uporabljajo. Zaključimo lahko, da se proces čiščenja hrvaščine uspešno izvaja (četudi še zdaleč ni dokončan), pri čemer avtor nasprotuje upravičenosti tega postopka.

This article deals with the process of language purism and the presence of Serbisms in Croatian. We attempt to show whether the speakers of Croatian know and use specific lexical items that are generally considered to be more characteristic of Serbian usage. For this purpose we use two data sources: a corpus of Croatian and a questionnaire survey. The results show that the speakers of Croatian are aware of the differences between Croatian and Serbian to a large extent. Even though they mostly do not use them, most speakers do know the meaning of lexemes that were used in the Serbo-Croatian period. This leads to the conclusion that the process of “cleansing” Croatian is being successfully implemented (but is far from completion). However, we argue against the justifiability of this process.

**Ključne besede:** hrvaščina, srbščina, srbizmi, jezikovni purizem, besedje

**Key words:** Croatian, Serbian, Serbisms, language purism, lexis

---

## 1 Introduction

The most often discussed problem in present-day Croatian philology is the period of existence of the so-called “Serbo-Croatian” language from the end of the XIX century up to the year 1991. The definition of this language has been, and still is, the subject of many debates – whether it was one common language for several nations, as the Serbian side claims, or an artificial language that didn’t exist in practice, as the Croatian side claims. In close connection to this is the question of the present-day relationship between Serbian and Croatian. No unambiguous answer can be provided to these questions since they are burdened by non-linguistic arguments, primarily political ones.

The present-day language situation in Croatia is characterised by a certain rebellion against lexemes and derivational affixes that were in use during the period of existence of an official common language, which creates a (false) impression that they are in fact Serbisms. “The new axiom is: Croatian and Serbian are different languages; and if that difference is not visible, it should be made so”<sup>1</sup> (Kordić 2010: 47). “A hunting season for Serbisms” is open. All things that may be associated with the former state of Yugoslavia and its official Serbo-Croatian language are thrown out of the language. However, these differences are sometimes so absurd and overly pedantic that there is no sense in pointing them out. It is a classical example of nitpicking. Bešker (2007: 11) gives a few examples of such “philological pedantry”: “I simply refuse to accept that the abbreviation *dr* (without a period) is the Serbian way, whereas *dr.* with a period is Croatian or that in Croatian you should put a period after a Roman numeral (which the Romans never did) while in Serbian or Montenegrin you should not.” Furthermore, new, “proper” Croatian words are being invented, and a speaker is being publicly castigated if he uses the language in any way other than the prescribed one.<sup>2</sup> The most notable example of this re-invention of Croatian is the contest for the “best new Croatian word” designed by the journal *Jezik* (Language). One of the criteria for a suggestion to be considered for the award is that it “replaces unnecessary and unacceptable loan words” (Babić 1993: 29). Over 500 suggestions are received every year. However, most of the laureate words are short-lived; they get talked about in the media but they never actually become used. This shows that any abrupt innovation to language introduced solely for innovation’s sake (rather than out of an actual need) is likely to fail.

The author of this article is of the belief that this kind of practice is completely wrong. Croatian and Serbian are too similar languages for this fact to

<sup>1</sup> All translations of originally Croatian citations were done by the author himself.

<sup>2</sup> See discussion in Kapović (2011: 50) on the fate of words *drug*, *zdravo* and *radnik*, which became highly stigmatised due to their “communist connotations”. In the same article, the author also illustrates how a person’s use of language is used to predict their political affiliation – for instance, if you write *ne ću* ‘I won’t’ instead of the traditional *neću*, you are labelled a rightist, whereas if you write *Evropa* instead of *Europa*, you are assumed to be a leftist.

be ignored. However, through the years, they have also drifted sufficiently apart – a natural process for languages – and there is no need to break them apart even further by invoking nationalist feelings. Moreover, many lexemes that are being thrown out of the language are typical of the Štokavian dialect, which serves as the dialectal basis for both Croatian and Serbian, so there are no grounds to claim they are typical only of Serbian. Fortunately, in recent years this practice has been tamed. Katičić rejects the practice of some philologists who are obsessed with clearly distinguishing between Croatian and Serbian. He concludes that “this definition of the Croatian language as opposed to Serbian is the worst one; moreover, it is completely wrong. The Croatian language does not exist by being different from Serbian but by being what it is. We cannot kick the words *ruka*, *nos*, *voda* out of the language just because they are used in Serbian as well. If we carried through this process consistently, we would lose the Croatian language” (Greenberg 2004: 117). Pranjković (2008: 101) also wonders why people should waste time and energy to start using, instead of an existing word to which no plausible objection can be found, a word which is, by objective criteria of standardology, worse or less correct.

The aim of this research is to discover to what extent the aforementioned process of “the hunt for Serbisms” has succeeded in practice, i. e., to what extent Croatian philologists have managed to frighten speakers of Croatian and exorcise Serbisms out of them. We hope for the research to show that this is not the case – that Croatian speakers are aware of the similarities and common elements of Croatian and Serbian and their views on this matter are not easily influenced by the dominant language policy. For this purpose we conducted a corpus study and a questionnaire study in which we examined usage patterns of lexemes that are commonly labelled as Serbisms.

In the second section we present the problems of the present-day relationship between Croatian and Serbian. In section 3 we detail our research methodology. The results of the research are presented in section 4. Conclusions are given in section 5.

## 2 Croatian and Serbian: how many languages?<sup>3</sup>

Many authors, when asked the question whether Croatian and Serbian are one language or more, give an ambiguous answer: both yes and no, or to put it poetically, as you like it. They will say that this question can – indeed should – be tackled from two perspectives: the perspective of language as a system (linguis-

---

<sup>3</sup> Space limitations do not allow us to go into much detail as to the complex relationship of language and dialect. Therefore, this section only looks at how this problem is tackled in Croatian and Serbian linguistic circles. For a more comprehensive overview of the problem, the reader is advised to consult broader sociolinguistic literature, such as Florian Coulmas (ed.), 1998: *The Handbook of Sociolinguistics*, Blackwell Publishing or John Edwards, 2009: *Language and Identity: An Introduction*, Cambridge University Press.

tic perspective) and the perspective of language as a standard (sociolinguistic perspective). Genetic linguistics and sociolinguistics have different criteria for defining the objects they operate with, hence the discrepancy in terminology.

From the perspective of language as a system, the whole area of former Yugoslavia belongs to a single dialect continuum, which is often referred to in linguistic literature as the *Central South Slavonic diasystem*, in which four dialects can be differentiated: Štokavian, Kajkavian, Čakavian and Torlak. Kajkavian and Čakavian are spoken only by Croats, Torlak is spoken only by Serbs, while the Štokavian dialect is common to Croats, Serbs, Bosnians and Montenegrins. From the viewpoint of language as a standard, the Neo-Štokavian ijekavian dialect was chosen as the basis of both standard Croatian and standard Serbian in the 19<sup>th</sup> century.<sup>4</sup> Some authors claim that Croatian and Serbian are two different versions of Štokavian, meaning that they contain elements of different dialects as well. But the definition of standard already takes that into consideration. A standard language is a super-dialect, which has one dialect as its basis and which is then further expanded by elements of other dialects. However, Croatian philologists put too great an emphasis on this “cultural-civilization expansion”. A further claim is that the choice of the standard dialect was made in different time periods, under different circumstances, and with different motivations (Brozović 1978: 56). Oczkova (2010: 38) claims that “during the period of Serbo-Croatian the diasystem held the function of *langue*, whereas its concrete realizations, i. e., Standard Croatian and Standard Serbian, also known as variants, were equal to *parole*. In the end, the common diasystem brought the two peoples together, but was practically realized by two standards.” The fact that is often pointed out as a unique feature of Croatian is its three-dialect structure, i. e., the co-existence of Štokavian, Kajkavian and Čakavian dialects. Katičić believes that “the language of Croats is both Kajkavian and Čakavian, of which the Serbs have not even a trace” (Greenberg 2004: 120). To sum up, whereas Croatian and Serbian might as well be regarded as one language system, Standard Croatian should most certainly not be identified with Standard Serbian.

The relationship between Croatian and Serbian is certainly affected by other criteria as well, such as mutual intelligibility and speaker identification. Mutual intelligibility has for a long time been considered the main criterion for distinguishing between a language and a dialect, but has been abandoned in recent years, as it is almost impossible to unequivocally determine at which point intelligibility stops and unintelligibility starts. One thing is forgotten, though: “the ability of communicating with members of other nations is a virtue of a language, not its flaw” (Kordić 2010: 104). If we take languages to be purely social constructs, then speaker identification would have to be considered a primary factor in defining languages.<sup>5</sup> However, research in dialectology and con-

<sup>4</sup> However, even though ijekavian is still officially accepted as part of the standard, contemporary Standard Serbian is primarily ekavian nowadays.

<sup>5</sup> I thank an anonymous reviewer of SCN for this insight.

tact linguistics has focused on the distribution of linguistic features (phonemes, lexemes, etc.) between languages and dialects more than on anything else.

Kordić classifies Croatian and Serbian (or Serbo-Croatian) as a *polycentric standard language*. By definition, this represents one standard language with several centres, which is based on the same dialect. This is the definition which the author of this text also adheres to. The most quoted examples of such languages are Hindi/Urdu, Portuguese in Portugal and Brazil, American/British/Australian varieties of English, etc. Comparisons of the core lexicon have revealed a 100-percent correspondence of Croatian and Serbian, and their syllable structure is also identical (according to the theory of the Copenhagen School of Linguistics, syllable structure is considered to be the DNA profile of a certain language) (Kordić 2010: 97).

The greatest number of differences between Croatian and Serbian appears on the lexical level. It is this level which will be the focus of our research. Lexical differences result from several factors: adoption of elements of local speech (in Croatian, those would be Kajkavian, Čakavian and non-Neo-Štokavian), different reflexes of the Old Slavonic phoneme *yat* (ě), different tolerance of loanwords (Croatian *tvornica*, *stroj*, Serbian *fabrika*, *mašina*), different paths by which a word entered a language (e. g. *okean/ocean*; the former word entered Serbian directly from Greek, the latter entered Croatian via Latin). Silić (2008: 58) points out that all of these are differences on the level of the standard language. That is, “it often happens that what the system treats as the same thing, the standard treats as two different things. To the system *formulirati* and *formulisati*, *Bizant* and *Vizant*, *diplomacija* and *diplomatija*, *Atena* and *Atina*, *amen* and *amin*, *Uskrs* and *Vaskrs* are the same. To the standard, on the other hand, they are different things”. Furthermore, Brozović (2008: 34) claims that “what is identical in the standards of the Serbs and the Croats are all those things that were necessary for everyday communication up until the XVIII century. But all those things civilization brought in the XIX and XX centuries differ in these two languages.”

Since linguistics still has not managed to find a criterion for effectively separating language from dialect, the question of these two languages cannot be solved in the near future. Perhaps the best answer to the question asked in the title of this section was provided by the great Croatian writer Miroslav Krleža (which many have held against him): “Croatian and Serbian are one language, which the Croats call Croatian and the Serbs call Serbian” (Kordić 2010: 38).

### 3 Methodology of research

As has already been said, this research aims to explore to what extent speakers of Croatian know and use lexical units that Croatian reference works consider to be Serbian. Since these units were, during the period of Serbo-Croatian, used

more or less equally in the ‘eastern’ and ‘western’ variants of the language,<sup>6</sup> whereas now it is recommended (even prescribed) that they are not to be used in Croatian, this research can also be interpreted as an evaluation of the success of the process of language purification, or “the hunt for Serbisms”, as we called it in section 1.

As the main source of lexical information we consulted the 2004 edition of *The Great Dictionary of the Croatian Language* by Vladimir Anić (shortened to *Anić* henceforth). Since there are several orthographic manuals, and even more so-called “language advisers”, coexisting in Croatia, each of them with their own philosophy and principles, we decided on the work that is still considered the key work of Croatian lexicography.<sup>7</sup> The list of potential candidates was retrieved from the 1993 edition of the *Dictionary of Differences between Serbian and Croatian* by Vladimir Brodnjak. However, since there are many issues with such dictionaries (e. g., it is unclear where the data comes from, evaluations of what counts as Serbian are fairly subjective, etc.), we decided to double-check the information found in Brodnjak in a more objective source. So, only words that carry the label *srp.* (Serbism) in *Anić* were taken into consideration in the final stage of our research. For instance, words such as *vazduh* ‘air’, *ostrvo* ‘island’, *sprat* ‘floor’, *ker* ‘dog’, which do not appear in *Anić*, were not taken into consideration. One thing should be noted here. The very fact that some Serbisms were included in *Anić* shows that they are units of the Croatian language (regarded here in its broadest sense as the totality of idiolects spoken in Croatia and Serbia at any point in time) just like any other and this status, at least from the perspective of language as a system, should therefore not be disputed. However, in most cases they are not considered standard forms and usually, alongside the label *srp.*, they carry another label, such as *reg.* (word characteristic of a particular region), *žarg.* (jargon word) or *razg.* (word characteristic of conversational style).

After producing a final list of Serbisms, their frequency was first determined in the Croatian Language Corpus (HJK).<sup>8</sup> This was followed by a questionnaire-based survey. The majority of words were selected for the questionnaire based on their frequency in the Corpus ( $f > 50$ ), but some less frequent ones were selected specifically by the author (being words that he uses himself). The

<sup>6</sup> For instance, a simple corpus search would show that one of the lexemes used in the study, *hartija*, was used more or less interchangeably with *papir* by Croatian authors of that period, such as Miroslav Krleža, Dragutin Tadijanović, etc.

<sup>7</sup> However, *Anić* has not been received without criticism either; see for instance Bašić (2005). The author criticises it for “ignoring well-established norms of standardology and functionalist styles of Croatian”.

<sup>8</sup> *Hrvatski jezični korpus* (HJK) is an unbalanced 100-million token corpus of Croatian developed at the Institute of Croatian Language and Linguistics. It contains texts from various periods of Croatian literature (28%) as well as newspaper texts from 2000–2006 and specialised writing (72%). For more information on it, see Čavar and Brozović Rončević, 2012. The corpus is accessible for free at <http://riznica.ihj.hr/philologic/Cijeli.whizbang.form.en.html>.

questionnaire examined several aspects of language, primarily language use and language attitudes. The questionnaire was distributed using the SurveyMonkey service in spring 2011. Respondents were recruited via the author's personal contacts on social networks and were not paid for their participation.

The questionnaire consisted of three parts. In the first part 20 pairs of synonymic words were provided, i. e., two (sometimes three) words of the same meaning, but with one word of the pair being a standard Croatian word, the other one being labelled a Serbism. The respondents were asked to indicate how often they used a particular word in everyday conversation on a scale of 1 to 5 (1 – I never use it, 2 – I rarely use it, 3 – I sometimes use it, 4 – I often use it, 5 – I always use it). The author is aware that this is not the most reliable method since it is fairly subjective and artificial; however, there was no time or opportunity for individual interviews, which would enable a more reliable insight into actual language usage. At the end of this part, the author left some space for the respondents' comments. They were given guidelines for what these comments should focus on: whether they use a particular word only in some situations, what the reason may be for using or not using a particular word is, whether there are any other words they use to express the same meaning, whether they would like to see any of these words banned from the Croatian language, etc. Even though the author suggested these as possible comments, the respondents were free to write anything they wanted and the majority of respondents did use this opportunity. A selection of the comments can be found in the Appendix.

In the second part of the questionnaire, the respondents were given 20 different Serbisms in isolation. They were asked to answer 3 questions for each word:

- Do you know the meaning of this word? (If they did, they were supposed to write it in order to see whether there are any discrepancies between the meanings found in dictionaries and speakers' meanings)
- How do you know its meaning? (The author suggested possible answers including, but not restricted to, "Somebody told me", "I heard it on TV", "My parents talk like that", etc.)
- Indicate how often you use this word on a scale of 1 to 5.

The third part of the research examined the respondents' attitudes to various language issues. The respondents were given 12 statements, 5 of them being statements about general attitudes towards loanwords and foreign languages and 7 of them statements about the attitude towards the Serbian language and Serbisms. They had to indicate to what extent they agreed with a particular statement on a scale of 1 to 5 (1 – I completely disagree, 2 – I mostly disagree, 3 – Neither agree nor disagree, 4 – I mostly agree, 5 – I completely agree).

Concerning the respondent structure, our target group was a younger generation of Croatian speakers. By younger generation, we imply people who did not come into direct contact with the Serbo-Croatian language, primarily through formal education. For that reason we took the year 1985 as the bottom borderline for the year of birth of the respondents. By the time respondents born



later than this year started school and received formal education in their native language, Serbo-Croatian and former Yugoslavia were a part of history and the aforementioned process of the separation of Serbian and Croatian invoking nationalist arguments was at work. In school, they were instructed in Standard Croatian, so the only possible contact with Serbo-Croatian was indirect. By way of comparison, it should be mentioned that the Croatian National Corpus (HNK), which is another 100-million token corpus of the language, takes the year 1990 as the bottom limit of its texts. The reason for this, according to one of its authors, Marko Tadić, is that “we all intuitively feel that from then on we could use Croatian more freely, more spontaneously, or put poetically, we could finally live and breathe it to the fullest” (Tadić 1998: 338).

#### 4 Results

The number of respondents who completed the questionnaire was 40 (12 male, 28 female). The oldest respondent was born in 1985, the youngest in 1993. When it comes to their regional diversity, 18 respondents come from Štokavian-speaking central Croatia (excluding the capital city of Zagreb), 12 come from the Kajkavian-speaking Zagreb and the surrounding region of Hrvatsko zagorje, 5 come from the easternmost region of Slavonija (Štokavian-speaking) and 5 come from Dalmatia (3 Štokavian-speaking and 2 Čakavian-speaking). Since the respondent sample is quite small and unequally distributed, especially for the last two regions, we cannot make any direct comparisons or analyses involving the various regions. We will now proceed to a discussion of the results of the individual parts of the questionnaire.

##### 4.1 Results of the 1<sup>st</sup> part

The results of the first part are presented in Table 1 below. Words in the table are sorted in such a way that the synonyms share one cell, the words on top being standard Croatian words, and the ones on the bottom (also italicised) those that *Anić* marks as Serbisms. In each field of the table we give the number of respondents who indicated a particular value of 1 to 5 for a particular word ( $N = 40$ ). In the last column, we give the raw frequencies of each lexeme in HJK. For the latter member of the pair we give two frequency values – frequency in the whole corpus and frequency in texts published after 1990 (given in brackets).<sup>9</sup> A comparison of these two numbers will enable us to see whether

<sup>9</sup> Since this is not one of the annotated sub-corpora of HJK, but was rather manually delimited, we have no information about its size. However, since the majority of the post-1990 texts are newspaper articles and the sub-corpus of “specialised texts”, which includes newspaper writings, makes up for the 72% of the whole corpus, we can conclude that the post-1990 sub-corpus makes up well over 50% of the whole HJK corpus.



there has been a noticeable decrease in usage of the lexemes in question following Croatian independence.

Lexeme	Gloss	never	rarely	sometimes	often	always	Mean value	Corpus freq.
bačva <i>bure</i>	barrel	0 14	4 12	18 12	10 2	8 0	3.55 2.05	748 188 (76)
Dobar tek! <i>Prijatno!</i>	Bon appetit!	0 32	0 5	1 3	6 0	33 0	4.80 1.28	22 1 (1)
puhati <i>duvati</i>	to blow	0 30	1 6	2 4	15 0	22 0	4.45 1.35	1100 226 (9)
izvanredan <i>vanredan</i>	extra-ordinary	1 8	6 18	10 6	12 6	11 2	3.65 2.40	6812 476 (18)
Oprostite! <i>Izvini!</i>	Excuse me!	0 21	1 13	2 5	6 1	31 0	4.68 1.65	2150 47 (7)
vrećica <i>kesa</i>	plastic bag	0 26	1 6	4 2	13 5	22 1	4.40 1.73	1256 464 (72)
kralješnica kraljeznica <i>kičma</i>	spine	22 1 0	11 6 9	3 11 13	2 11 11	2 11 7	1.78 3.63 3.40	154 502 238 (171)
kotač <i>točak</i>	wheel	0 27	2 3	4 5	11 3	23 2	4.38 1.75	1354 218 (26)
milijun <i>milion</i>	million	0 20	1 10	2 4	12 5	25 1	4.53 1.93	77593 58 (5)
natjecati se <i>takmičiti se</i>	to compete	0 22	0 12	3 5	11 1	26 0	4.58 1.63	17533 65 (51)
znanost <i>nauka</i>	science	0 10	0 16	4 11	14 3	22 0	4.45 2.18	15202 816 (411)
vezica žniranac <i>pertla</i>	shoelace	18 9 15	8 7 4	9 8 3	3 5 9	2 11 9	2.08 3.05 2.83	94 6 0
tržnica <i>pijaca</i>	market	3 24	4 8	10 2	10 4	13 2	3.65 1.80	3343 84 (12)
Pomakni se! <i>Pomjeri se!</i>	Move over!	0 32	0 6	4 1	7 1	29 0	4.63 1.28	1471 25 (6)
popis <i>spisak</i>	list	0 9	0 9	5 12	13 8	22 2	4.43 2.63	10062 190 (155)
stupanj <i>stepen</i>	degree	1 37	0 3	2 0	9 0	28 0	4.58 1.08	4950 147 (3)
želudac <i>stomak</i>	stomach	0 17	1 9	3 5	14 6	22 3	4.43 2.23	1462 41 (14)
sudjelovati <i>učestvovati</i>	to participate	0 21	0 17	0 2	9 0	31 0	4.78 1.53	20824 144 (43)
uputa <i>uputstvo</i>	instruction	1 9	3 12	3 11	14 3	19 5	4.18 2.58	1931 106 (84)
uvjet <i>uslov</i>	condition	0 37	2 2	1 1	12 0	32 0	4.68 1.10	27095 127 (25)

At first glance it can be seen that there is a fairly distinguishable difference in use between the two synonyms, namely that the respondents use standard Croatian words quite often (have a mean value between 4 and 5), whereas their Serbian equivalents are used relatively rarely (between 1 and 2). However, there are a few exceptions. The questionnaire contained two cases where three words were in a synonymic relationship (instead of the usual two). It was those triplets that gave the most interesting results. When it comes to the triplet *kičma* – *kralješnica* – *kralježnica*, according to *Anić*, *kralješnica* is a standard Croatian word, *kralježnica* is typical of conversational style, while *kičma* is a Serbism. However, both corpus data and respondents' answers show that the standard Croatian word is the one least used, the informal word is the most frequent, while the Serbism is somewhere in between. Similar results can be seen for the triplet *pertla* – *vezica* – *žniranac*, where *vezica* is a standard Croatian word, *žniranac* is stylistically labelled and *pertla* is a Serbism. However, unlike the Corpus, some respondents do use *pertla* on a regular basis. Several respondents mentioned a potential semantic differentiation between *vanredan* (used to refer to part-time students) and *izvanredan* (used to describe something extraordinary), but this difference was not mentioned in any of the reference works.

If we take a look at the corpus frequencies, we can see that there is a drastic drop in use of 'Serbisms' in texts written after 1990 for the majority of lexemes. There are some lexemes which do not seem to follow this pattern, but this is just an apparition. For instance, the word *nauka* appears in post-1990 sub-corpus 411 times. However, the majority of these tokens are actually used when referring to the Serbian Academy of Science (*Srpska akademija nauka i umetnosti*). If we disregarded those, the total number of tokens of *nauka* would be less than 100. The word *bure* appears in a popular phrase "*bure baruta*" 'powder keg', which is used to refer to a tight and potentially dangerous situation and was also used in the title of a popular play. The majority of post-1990 tokens are used in one of these two contexts. Interestingly enough, none of the respondents mentioned this phrase as the context where they would use *bure*. On its own, *bure* appears only 10 times in the post-1990 sub-corpus. Croatian purists often quote the word *stepen* as an example of a typical Serbism that should not be a part of Croatian in any case. It seems their attacks were fruitful as both the respondents and the Corpus do not use it anymore (the 3 tokens from HJK were all used by a single author). On the other hand it seems the words *uputstvo*, *spisak*, *takmičiti se* and *kičma* were less affected by such puristic charges.

#### 4.2 Results of the 2<sup>nd</sup> part

Results of this part are given in Table 2 below. The third column tells us how many respondents indicated they knew the meaning of the word (Y) as opposed to how many did not (N). Mean use in the fourth column was calculated using the same scale as in Part 1. We also give raw frequencies of the individual lexemes from HJK, both the whole corpus and post-1990 sub-corpus (in brackets). If a

respondent gave a meaning which did not correspond to the one found in Anić or any other reference work, it was placed in the N category. For instance, several respondents thought *talas* meant ‘wind’, *komšija* meant ‘friend’, *kajsija* meant ‘peach’, or *dušek* meant ‘pillow’. There were around 20 of such cases in total.

Lexeme	Gloss	Knowing the meaning	Mean use	Corpus frequency
<i>aždaja</i>	dragon	29 Y, 11 N	2.10	64 (14)
<i>banja</i>	bath, spa	28 Y, 12 N	1.48	6 (3)
<i>cicija</i>	a stingy person	23 Y, 17 N	1.75	14 (9)
<i>drum</i>	dirt road	22 Y, 18 N	1.25	353 (64)
<i>dušek</i>	mattress	14 Y, 26 N	1.25	60 (1)
<i>džigerica</i>	liver	26 Y, 14 N	2.03	45 (8)
<i>hartija</i>	paper	7 Y, 33 N	1.05	207 (11)
<i>hljeb</i>	bread	40 Y	1.30	577 (16)
<i>kajsija</i>	apricot	24 Y, 16 N	2.05	30 (4)
<i>kašika</i>	spoon	38 Y, 2 N	1.53	88 (16)
<i>kec</i>	1) ace in cards 2) negative mark	33 Y, 7 N	2.08	13 (2)
<i>komšija</i>	neighbour	33 Y, 7 N	1.80	164 (48)
<i>korpa</i>	basket	30 Y, 10 N	1.85	132 (17)
<i>kvasiti</i>	to soak	36 Y, 4 N	2.03	62 (7)
<i>patika</i>	shoe	40 Y	2.38	15 (14)
<i>peškir</i>	towel	39 Y, 1 N	1.50	41 (1)
<i>promaja</i>	air draught	33 Y, 7 N	1.83	19 (5)
<i>supa</i>	soup	39 Y, 1 N	1.28	2 (1)
<i>talas</i>	wave	25 Y, 15 N	1.28	390 (5)
<i>zvaničan</i>	official	21 Y, 19 N	1.48	41 (16)

We shall analyse a couple of individual examples from this part. The word *drum* is mostly used in two traditional proverbs: *Što na umu, to na drumu*<sup>10</sup> and *Baba šumom, dida drumom*.<sup>11</sup> That is the main reason for its relatively high frequency in the corpus, even in the post-1990 sub-corpus. Most respondents commented that they came across the meaning of this word from one of these proverbs. *Hljeb* is familiar to all our respondents. However, it is interesting to note how Anić’s Dictionary labels this word. *Hljeb* can have two meanings: 1) bread in general, or 2) a round loaf of bread. But the label *srp.* appears only in connection to the

<sup>10</sup> Used when someone reveals their true intentions and opinions, often unintentionally.

<sup>11</sup> When two people have diametrically opposite opinions and are unlikely to ever agree on something.

first meaning. In meaning 2 *hljeb* is treated as an unmarked term. Next, let us focus on the word *hartija*. It is the least familiar word for the respondents (only seven respondents knew its meaning), which is quite surprising if we consider that many great Croatian writers, like Miroslav Krleža, Tin Ujević, etc., used that word in their works. It appears that the respondents are not so familiar with classical works of Croatian literature. Such examples of Serbisms appearing in older literary works are numerous. However, in recent years new editions of these works started appearing, but in these editions certain ‘inappropriate’ words (*talas* being another example) were replaced by a ‘better Croatian word’. For instance, Gluhak (2004) compared the original editions of the works of Marija Jurić Zagorka with the more recent editions and discovered numerous ‘corrections’ (*Evropa* substituted by *Europa*, *Kći krvnika* by *Krvnikova kći*, etc.). This is an example of a completely inappropriate intervention as it decomposes the integrity of the original work. In fact, Brozović (1978: 76) goes so far to say that “the contemporary audience does not even know what Šenoa and Kovačić actually wrote like” because their works have been adapted to the present-day standard.

When it comes to the question of how the respondents know the meaning of a particular word, the answers were diverse. The most common source mentioned was older members of the family who still used the word in everyday conversation (e.g. *supa*, *kajsija*). Some of the other common sources were films (e.g. Serbian film *Komšije*), songs, old recipes (for *kašika* and *džigerica*), folklore, etc. The word *patika* is especially common in Slavonija and Dalmatia, hence its relatively high mean value.

#### 4.3 Results of the 3<sup>rd</sup> part

The statements that the respondents evaluated are given in Table 3 below. As has already been said, the first five statements reflect the respondents’ general attitudes towards the standard language and borrowings, whereas statements 6–12 reflect their attitudes towards Serbian and Serbisms. In the actual questionnaire the statements appeared in a randomised order. The scale used was a 5-point Likert scale where 1 means ‘I completely disagree’ and 5 ‘I completely agree’.

Statements	1	2	3	4	5	Mean value
S1: <i>I often communicate with speakers of other languages.</i>	1	3	7	14	15	3.98
S2: <i>Foreign words pollute the Croatian language.</i>	7	11	16	5	1	2.55
S3: <i>All foreign words should be thrown out of the Croatian language.</i>	15	12	11	2	0	2.00
S4: <i>Loanwords from some languages are more acceptable to me than loanwords from some other language.</i>	9	6	9	11	5	2.93

Statements	1	2	3	4	5	Mean value
S5: <i>It is important to me that I speak my mother tongue in a proper, standard way.</i>	0	1	4	23	12	4.15
S6: <i>Croatian and Serbian are fairly similar languages.</i>	0	3	8	16	12	3.95
S7: <i>I often come in contact with the Serbian language.</i>	2	12	6	14	6	3.25
S8: <i>I do not object to using Serbisms in Croatian.</i>	2	9	13	14	2	3.13
S9: <i>I could easily communicate in the Serbian language.</i>	0	1	2	20	17	4.33
S10: <i>I do not consider Serbian a foreign language.</i>	2	7	14	10	6	3.28
S11: <i>All words of Serbian origin should be thrown out of the Croatian language.</i>	10	10	15	5	0	2.38
S12: <i>Croatian and Serbian are the same language.</i>	24	12	3	1	0	1.53

The best analysis of this part can be made by comparing the first five, ‘general’ statements with their counterpart statements for Serbian. Let us compare, for instance, statements 3, 4 and 11. The respondents show a fair amount of tolerance towards foreign words (i. e., they mostly disagree with them being thrown out of the language); however, their tolerance seems to decrease when Serbisms are concerned. Also, as Statement 4 shows, they tend to give different status to borrowings from different languages. This is something Greenberg (2004: 123–124) noticed as well. When reflecting on *Hrvatski Jezični Savjetnik* (*Croatian Language Advisor*), he states that its authors

have displayed tolerance towards borrowings from lending languages of nations for which the Croats have felt cultural affinity. Such tolerance is greater especially concerning borrowings from French, Italian, and Hungarian /.../ Conversely, the Croats have been intolerant of borrowings from lending languages of peoples they have considered to be culturally alien (such as Turkish or Russian) /.../ While they do not explicitly state this, the authors have rejected the Russian forms, since these are the very forms that have been adopted by the Serbs. The Croats, therefore, have rejected both Orthodox Slavic and oriental/Islamic elements from their language, thereby underscoring the place of the Croatian language within a Central European context, with the hope of giving their national image to a more European identity.

As this lengthy quote shows, in the late-1990s the dominant language policy was that everything that had come from the east was unacceptable and everything that had come from the west was tolerated. However, this dichotomy is no longer visible as the attention has in recent years turned to English as the topic of puristic debates. Now that the Croatian linguists have (more or less successfully) cleaned up the language of Serbisms, Anglicisms seem to be the next in line.

When it comes to Statements 9 and 10, there was quite a vivid polemic in Croatian media when one public figure put Serbian in the field “Foreign lan-

guages” of her CV.<sup>12</sup> By using this logic, most people from the ex-Yugoslavia region would be multilingual. Težak (1999: 161) claims that Croatian and Serbian are related languages. “From the viewpoint of intelligibility, it is not easy to proclaim Croatian and Serbian two foreign languages. However, from the viewpoint of familiarity with a language, it would be wrong to say that a Croat wishing to speak Serbian should not learn Standard Serbian.” It appears that some of our respondents would not agree with Težak.

When it comes to Statement 12, *Croatian and Serbian are the same language*, the author may have made a mistake in formulating this statement. We have already said that some authors would claim that the most realistic genetic relationship of Croatian and Serbian is two variants of the same language but none of them would presuppose a one-to-one correspondence between the two. Perhaps the answers to this question would have been different if we had put “one language” instead of “the same language”.

## 5 Conclusion

It was said that this work partially wanted to show how successful the process of “the hunt for Serbisms” is. Our respondents mostly do not use Serbisms (with a few exceptions) even though they know their meanings and are aware of their origin. Why they do not use them, whether it is for some personal reason or because they were taught, even forced that way, can partially be predicted from their comments (see Appendix), but for a more accurate conclusion more detailed research is needed. The corpus research has also shown that the great majority of Serbisms stopped being used as frequently after 1990. All of this seems to lead to the conclusion that Croatian language purism has produced the intended results. However, as was said before, we doubt whether a puristic approach is in fact good. We do not intend to put forward claims as drastic as Kordić (2010: 52), who says that “every attempt of language cleansing must be wrong because no language was ever clean in the first place.” Some degree of purism is necessary in order for there to be a codified standard;<sup>13</sup> however, we first need to unequivocally determine what belongs to the standard, what belongs to different dialects, sociolects and what belongs to another language.

Croatian and Serbian have, since their beginnings, functioned as separate standard idioms. The differences between them have, and always will be, present so that this view can be clearly defended. It is inadmissible that these differences are disregarded, as was the case in the past (in the form of the 1850 *Vienna Agreement* and the 1954 *Novi Sad Agreement*, see discussion in Peti-Stantić 2008), but at the same time it is also inadmissible that the similarities are disregarded and the differences augmented in a standardologically unacceptable way.

---

<sup>12</sup> See for instance a column by Igor Mandić in *Jutarnji list* from 21<sup>st</sup> and 28<sup>th</sup> January, 2010.

<sup>13</sup> I thank an anonymous reviewer of SCN for this insight.

We realise that both the respondent sample of this research and the material covered were relatively small. But even such a small sample opened up some interesting questions, which should be developed further in subsequent studies (especially the question of treating loanwords from different languages differently). We believe that the only way to go forward in clarifying the complex relationship between Croatian and Serbian is through argumentation based on empirical facts such as these rather than on the authority of the linguist or speculation.

#### APPENDIX: SELECTED RESPONDENTS' COMMENTS

In this Appendix we present some of the comments left by the respondents after filling out part 1 of the questionnaire. We selected the most interesting and illustrative ones. The information given in brackets are their year of birth, gender, place where they were born (or currently living), and education status (SSS means finished secondary school, student means currently studying and VSS means finished university).

**#1 (1992, M, Zagreb, student):** I use *prijatno* only with people I know, but I find the word funny; same for *Izvini*. I don't like *kralješnica* and *kralježnica* because I am irritated by words that have two different versions, so I end up using *kičma*. When it comes to *izvan-* and *van-*, I know that one of them is a Croatian prefix and the other Serbian, but I still use *vanzemaljci* and *vanbračan*, etc.

**#2 (1991, F, Kutina (born in Sarajevo), student):** *Bačva* sounds weird, *bure* sounds right. *Prijatno* is used in Bosnia, so people often look at you funny when you use it. I use *izvanredan* when I want to say that something is excellent, above average; I use *vanredan* when referring to part-time students. I use *kičma* when referring to the lower back, *kralježnica* when referring to the whole back; I never heard the variant with /š/. *Milijun* – I might use it subconsciously, I always use *milion*. *Spisak*, *stepen*, *stomak* are foreign words so I don't use them.

**#3 (1989, F, Sisak, student):** I use *vanredan* because it is shorter, but they are all the same to me. I use *pertla* because it is interesting to pronounce 3 consonants in succession.

**#4 (1986, F, Zagreb (born in Čapljina), student):** I never use *Izvini*, our 5<sup>th</sup> grade teacher of Croatian managed to scare that word out of us. I sometimes use *točak* for fun. *Vanredan* – sometimes when I don't feel like pronouncing *izvanredan*.

**#5 (1991, F, Vinkovci (living in Zagreb), student):** I use *izvanredan* often, but only in the meaning of 'excellent' rather than as the opposite of 'full-time student'.

**#6 (1990, F, Šibenik (living in Zagreb), student):** *Prijatno* – I only hear that from older people who want to sound more polite than they really are. *Duvati*, *pomjeri se*, *stepen* – I use them only if I am making fun or singing some Serbian rap.

**#7 (1984, F, Zagreb (born in Donja Stubica), VŠS):** *Uslov* sounds strict, I wouldn't really use it in everyday conversation. I've heard *točak*, but I like *kotač* better.

**#8 (1985, M, Zagreb, student):** I find the word *kesa* interesting because, at least to me, it has a slightly different meaning. I use *vreća* for a bag of potatoes, rubbish, etc., but



*kesa* is small and reminds me of medieval times, as if it should contain gold coins or other valuables. The words I use quite often – *pertla*, *kičma* or *spisak* – don't feel like they come from Serbian, but rather from some other language.

**#12 (1986, M, Sisak, VSS):** Some of these words became so common in street talk that we don't really notice the difference when we talk. I encounter most of these words when talking to my grandmother, so they all sound normal to me. I don't use *vezica* because it sounds weird; *pertla* is how the folk speak.

**#15 (1988, M, Šibenik (living in Sisak), SSS):** I think I used some of these words when I was little, but my parents or some other people corrected me and now I use the 'correct' words.

**#16: (1986, M, Cres (born in Belgrade), student):** From the three options *pertla*, *žniranač*, *vezica*, I use neither; I use *špigeta*. I don't like our word *kralješnica*, I prefer the variant with /ž/.

**#17 (1987, F, Petrinja (born in Virovitica), student):** Some of these words I don't use by accident but on purpose, for fun. But I am afraid they might catch on to me as I always found Serbian extremely cute.

**#20 (1987, F, Zagreb, student):** I don't use some of these words because they are not local (Kajkavian) to me (*duvati*, *kesa*, *točak*, *stomak*).

**#23 (1984, F, Slavonski Brod, SSS):** *pijaca/tržnica* – I use whichever one comes to mind first; I don't use *prijatno* and *duvati* because I know for a fact that they are not Croatian words.

**#24 (1987, M, Zagreb, student):** *Bačva/bure* – in my head they differ in shape and function (*bure* would be for alcoholic drinks, *bačva* for all other liquids). I use *duvati* only to refer to smoking grass. I use *nauka* only for pejorative meaning (Nije ti to baš nauka!) [equivalent to English *It's not exactly nuclear science*; author's note]. *Spisak* would be something negative, like a 'hit list'. I sometimes use *milion* to emphasise something.

**#25 (1989, F, Split (living in Zagreb), student):** *Kesa* is a completely normal word for me, but when I am in Zagreb I tend to use *vrećica* so I would be understood. My dad sometimes uses *takmičiti se*, but it sounds so archaic and it hurts my ears. When my stomach hurts, I will use *boli me stomak*; I will use *želudac* only when talking about the concrete organ or when expressing disgust (*okrenuo mi se želudac*) [lit. 'my stomach turned over', author's note]. I stopped using *prijatno* ever since my secondary school teacher shouted at me for using it. I hear *kičma* more often than I would use it, but it doesn't sound archaic or Serbian.

**#26 (1990, F, Daruvar (living in Zagreb), student):** I find *kralješnica* totally moronic. I use *duvati* and *kesa* for a joke.

**#27 (1991, F, Petrinja (born in Bugojno), student):** I've heard of most of these words and it's not unusual to hear them in everyday conversation nor do they evoke any negative emotions for me.

**#28 (1993, Ž, Petrinja (born in Zagreb), pupil):** I don't pay attention which of the two terms *Pomakni se/Pomjeri se* I use, but I noticed that people react differently to them; when I say *Pomjeri se*, they find it strange.

**#30 (1985, F, Sisak, VSS):** Words like *prijatno*, *duvati*, *kesa*, *točak*, *stepen*, etc., are typical Serbisms for me; I don't like them and I wasn't raised in a surrounding where they

would be used. I find it weird when I hear one of them in Croatia. Mostly older people use them. However, I can hear some other ones (*bure, kičma, million, takmičiti se, stomak*) not only from older people, but from young as well.

**#35 (1988, F, Zagreb, student):** I use *Izvini* relatively often as it is less serious than *Oprosti*. When I use Serbisms, I am aware that they are serbisms or perceived as such, so I use them either because I like them or spontaneously. But it's always interesting to see people's reaction, even a grimace.

**#36 (1986, F, Sisak (living in Zagreb), student):** *Prijatno* is such a cute word, but it can never come to my mind in conversation; *Kičma* is totally natural for me to use, although I would use *kralješnica/kralježnica* in official situations (I don't distinguish between the two).

## REFERENCES

- Vladimir ANIĆ, 2004: *Veliki rječnik hrvatskoga jezika*. Zagreb: Novi Liber.
- Stjepan BABIĆ, 1993: Najbolja i najgora riječ 1993. godine. *Jezik* 41/1, 29–31.
- Nataša BAŠIĆ, 2005: Jezična politika kao razgradnja hrvatskih jezičnih norma. *Jezik* 52/1, 7–21.
- Inoslav BEŠKER, 2007: *Filološke dvoumice*. Zagreb: Jesenski i Turk.
- Vladimir BRODNJAK, 1993: *Razlikovni rječnik srpskog i hrvatskog jezika*. Zagreb: Školske novine.
- Dalibor BROZOVIĆ, 1978: Hrvatski jezik, njegovo mjesto unutar južnoslavenskih i drugih slavenskih jezika, njegove povijesne mijene kao jezika hrvatske književnosti. *Hrvatska književnost u evropskom kontekstu*. Ed. Aleksandar Flaker and Krunoslav Pranjić. Zagreb: Liber 9–83.
- –, 2008: Jesu li Bečki i Novosadski ‘dogovori’ samo beznačajne epizode i činovi unitarističkog nasilja – ili jedine osnovne točke u hrvatskoj novoštokavskoj standardizaciji? *Identitet jezika jezikom izrečen*. Ed. Anita Peti-Stantić. Zagreb: Srednja Europa. 33–42.
- Damir ČAVAR and Dunja BROZOVIĆ RONČEVIĆ, 2012: Riznica: The Croatian Language Corpus. *Prace filologiczne LXIII*, 51–66.
- Alemko GLUHAK, 2004: Jezik vremena Marije Jurić Zagorkе. *Grička vještica, knjiga 7: Buntovnik na prijestolju*. Marija Jurić Zagorka (Ed. Miroslava Vučić). Zagreb: Školska knjiga. 664–685.
- Robert D. GREENBERG, 2004: *Language and Identity in the Balkans: Serbo-Croatian and its Disintegration*. New York: Oxford University Press.
- Mate KAPOVIĆ, 2011: Language, Ideology and Politics in Croatia. *Slavia Centralis IV/2*, 45–56.
- Snježana KORDIĆ, 2010: *Jezik i nacionalizam*. Zagreb: Durieux.
- Barbara OCZKOWA, 2010: *Hrvati i njihov jezik*. Zagreb: Školska knjiga.

Anita PETI-STANTIĆ (ed.), 2008: *Identitet jezika jezikom izrečen*. Zagreb: Srednja Europa.

Ivo PRANJKOVIĆ, 2008: *Sučeljavanja: polemički dueli oko hrvatskoga jezika i pravopisa*. Zagreb: Disput.

Josip SILIĆ, 2008: Neetnički i etnički identitet. *Identitet jezika jezikom izrečen*. Ed. Anita Peti-Stantić. Zagreb: Srednja Europa. 57–62.

Marko TADIĆ, 1998: Raspon, opseg i sastav korpusa suvremenoga hrvatskoga jezika. *Filologija* 30–31, 337–347.

Stjepko TEŽAK, 1999: *Hrvatski naš (ne)zaboravljeni*. Zagreb: Tipex.

## OSTANKI SRBOHRVAŠKE LEKSIKE V DANAŠNJI HRVAŠČINI

V hrvaščini je vse od razpada Jugoslavije opazen proces čiščenja, tj. izločanja vseh besed, za katere se predpostavlja, da so prispele iz srbščine oz. so bile značilne za srbohrvaščino, saj vzbujajo strah, da gre za srbizme. Ta raziskava je bila izvedena z namenom ugotoviti, v kolikšni meri je proces ločevanja hrvaškega in srbskega jezika uspel v vsakdanjem življenju. Želeli smo preveriti, ali govorci hrvaškega jezika poznajo in uporabljajo tiste leksikalne elemente, za katere na splošno velja, da so bolj značilni za srbščino. V ta namen smo uporabili dva vira podatkov: korpus hrvaškega jezika in anketni vprašalnik (N = 40). Korpusna raziskava je pokazala, da se je raba srbizmov zmanjšala v besedilih, objavljenih po letu 1990. S pomočjo vprašalnika je bilo preverjeno, ali govorci poznajo pomen posameznih srbizmov, kako pogosto jih uporabljajo v vsakdanjem govoru in kakšen je njihov odnos do srbščine in srbizmov. Rezultati kažejo, da se govorci hrvaškega jezika v veliki meri zavedajo razlik med hrvaščino in srbščino. Večina govorcev sicer pozna pomen leksemov, ki so bili uporabljani v srbohrvaškem obdobju, a jih večinoma ne uporabljajo. Zaključimo lahko, da se proces čiščenja hrvaščine uspešno izvaja (četudi še zdaleč ni dokončan), pri čemer avtor nasprotuje upravičenosti tega postopka.