

Alexander D. Hoyt  
Odsjek za anglistiku  
Filozofski fakultet u Zagrebu  
alex.hoyt@ffzg.hr

## BUILDING A HISTORICAL CORPUS OF CROATIAN CORRESPONDENCE

Presented at the annual conference of the  
Croatian Applied Linguistics Association (HDPL)  
in Osijek, Croatia, in May of 2011

## Building a Historical Corpus of Croatian Correspondence

### Abstract:

My project is an attempt to learn about linguistic change and variation in a linguistic community (Croatian speakers at the turn of the twentieth century) whose written language was undergoing the process of standardization. I am compiling a corpus of personal correspondence, consisting of four types of written messages (traditional letters, postcards, calling cards, and telegrams) written at 10-year intervals between 1875 and 1925, from the personal collection of Istrian politician Vjekoslav Spinčić (1848-1933), kept in the Croatian State Archives. The collection contains some letters written by Spinčić himself, although the vast majority consists of letters sent to him by other people from all levels of society. Most historical corpus projects of this type have focused on English and other Northern European languages with longer traditions of literacy. This is the first such study on Croatian, as well as one of the first historical studies of the Croatian language based upon natural language use rather than, say, the language of literature. Not only will this research be a contribution to the study of language change and variation, but the content of the letters themselves will be of interest to researchers in other fields of study, such as history or ethnology.

### 1. Introduction

Although many different works describe the history of the Croatian language, most of them cover the topic in from the top down, using examples from Croatian literature and official documents published in Croatian in past centuries, or describing the legal and social status of the Croatian language in relation to the languages of power that were used in the lands where Croatian was spoken (Italian, German, Hungarian, and Turkish). Few works focus on the Croatian language as it was used in everyday communication in the past. With this in mind, I have commenced on a project that should serve as an interesting tool for those interested in the study of how Croatian was used in the period when it was becoming standardized. I am currently compiling a corpus of personal letters that were written in the late nineteenth and early twentieth century.

## 2. CEEC – The Model

My project uses as its model the Corpus of Early English Correspondence (CEEC), a project that was initiated in the early 1990s by two Finnish scholars of Early Modern English, Terttu Nevalainen and Helena Raumolin Brunberg. The CEEC was conceived as a tool for the study of historical sociolinguistics. The project has continued since then and at present includes about 12,000 letters written by about 1,200 people between 1403 and 1800 (*Corpora of Early English Correspondence* website). The primary goal of the project is to examine how variationist sociolinguistic models can be applied to historical linguistics. For this purpose, the project originally compiled two complementary corpora: one containing the text from all of the transcribed letters, and the other containing relevant sociolinguistic information about each of the letter writers. The vast majority of the letters in the corpus are from published editions of personal letters in which the original spelling was retained. The CEEC team transcribes these letters by scanning the books in which they were printed and editing them. When problems occur, they consult the original letters themselves, if they still exist. Sociolinguistic data on the letter writers is saved in a data file and shows each correspondent's dates of birth and death, sex, social rank, social mobility in their lifetime, place of birth, main place of residence, native and other languages, among other data. (Nevalainen & Raumolin-Brunberg 50) In order to insure that there is enough linguistic data for each informant, the project set ten medium-length letters as the minimum for a letter writer to be included in the corpus (although this requirement is sometimes waived in cases of people in less well-represented groups, such as the lower social ranks and women). Since the first version of the CEEC (1410-1681) was finished as a plain-text file in 1998, it has been parsed and subsequently made available to the public as the Parsed Corpus of Early English Correspondence (PCEEC). In addition, an extension to the corpus is also being compiled (1681-1800), so that research can be done on the mechanisms of language change over a time span of four centuries (Raumolin-Brunberg & Nevalainen).

## 3. A Historical Corpus of Croatian Correspondence

Inspired by the CEEC and how its compilers were broadening the realm of sociolinguistics so that its methods could be used to study patterns of change in real time over several centuries in the past, it occurred to me that the same method could be used to study language change Croatia. In 2006, as part of an elective sociolinguistics course I was teaching at the University of Zagreb, I decided to take my small class of students to the Croatian

National Archives in order to see what kind of personal letter collections they had in their files. Thanks to the enthusiasm and generosity of several employees at the archives,<sup>1</sup> we were allowed to examine various letters written in the late-nineteenth century. During those visits, I discovered one collection that was different from the rest. This was the massive personal collection of Vjekoslav Spinčić, which spans more than sixty years from the late 1860s until the early 1930s.

### 3. Vjekoslav Spinčić – a multi-faceted man

Vjekoslav Spinčić was born in the Istrian village of Spinčići, near Kastav, a town in the hills overlooking the ports of Rijeka and Opatija, in 1848. After attending the local primary school, he went on to the Croatian Gymnasium in Rijeka, passing his Matura exams in 1868. He then studied theology in Gorizia and Trieste, where he was ordained as a Catholic priest in 1872. Intending to follow a career in education, he proceeded to study history and geography in Prague and Vienna. In 1876 Spinčić found employment at the male teachers' college in Koper, where future teachers of Croatian (as well as Slovene and Italian) were trained to work in primary schools throughout Istria and the Kvarner Islands. In 1882, he was appointed the inspector of schools in the districts of Koper and Voloska, a post which he held until 1888. That year he was transferred to the teachers' college for women in Gorizia. In 1892, Spinčić was relieved of his services as an educator, because of his political activities during the elections of 1891, but mainly for giving a speech at the Zagreb agricultural fair in Autumn 1891 in which he said that Istria, then a province in the Austrian half of the Austro-Hungarian monarchy, should be annexed to Croatia proper, then a kingdom within the Hungarian half of the monarchy (Beuc 9, Strčić & Strčić 99–100).

In 1882 he was first elected as one of three Croats to the Istrian Assembly. He served in that body continuously until its dissolution in 1916. In 1891 he was elected to the House of Representatives of the Imperial Council in Vienna, continuing to serve in that capacity until the end of the Austro-Hungarian state in 1918. He also served in the Istrian provincial government in 1890 and 1891. (Beuc 9)

Spinčić decided upon a career in the clergy because he considered it the best means for him to help the Croats of Istria (Strčić & Strčić 99). In the mid-nineteenth century, Istrian

---

<sup>1</sup> Here I would like to thank my first contact at the Croatian State Archives, Deana Kovačec, who heads the archives' Department of Communication, Promotion and Publishing. Her help in introducing my students to the archives was invaluable. I would also like to thank Melina Lučić, head of the Department of Newer Family and Personal Collections, for her generous assistance, as well as Mario Stipančević, whose selection of late nineteenth-century Croatian letters for my students to examine eventually led to my discovery of the rich collection of correspondence of Vjekoslav Spinčić.

society was unevenly divided between an Italian-speaking, mostly urban minority holding the vast majority of the wealth and power, and a rural, Slavic-speaking majority concentrated in rural areas, most of whom had little access to education or economic improvement. His entire life's work – as priest, educator, historian, politician – was dedicated to defending the rights of the Croatian and Slovene people of Istria. This could be done via two routes of action – by taking part in politics and by improving literacy and education in general among the rural Croatian population. Because Spinčić was a leading actor in both of these areas, he had an enormous number of contacts with whom he corresponded regularly.

As a historian, Spinčić decided to save all of the correspondence that he received. He rightfully believed that this collection of letters would some day serve as a historical record. Late in his life, he donated all of his correspondence to the Croatian national archives in Zagreb, where they are kept today.

#### 4. The Spinčić Collection

The Spinčić collection consists of 132 archival boxes containing Spinčić's personal notes and journals, documentation related to his activities in Istrian politics and education, manuscripts of his literary compositions and historical treatises, personal identification documents, his correspondence, and miscellaneous newspapers and newspaper clippings. The entire collection has been recorded on microfilm, which can be viewed at the Croatian State Archives in both Zagreb and Pazin in Istria. Correspondence alone accounts for 59 of the boxes (nos. 63–121), which contain 16,790 letters (Beuc 5) written to Spinčić by relatives, friends, colleagues, citizens from all walks of life, societies, and companies. A small fraction of the letters were written by Spinčić to others (many of which are in the form of letter drafts written in pencil). Some of the letters in the collection were written in other languages, such as Slovene, German, Czech, Italian, and even Latin, but the vast majority were written in some form of Croatian, representing either what was then the standard language or a local variety.

#### 5. Four Types of Correspondence

The Spinčić collection includes not only letters, but other types of correspondence as well. The most common correspondence types are the following: traditional letters, postcards, telegrams, and calling cards. These types differ according to several variables: purpose, means of transportation, physical material, and typical length.

A. Traditional Letters (*Pisma*) include any form of correspondence written on one or more sheets of paper and sent either via post or human carrier. Letters in the collection can be very brief, written on one side of a sheet of paper, but can also be as long as eight pages or more. Longer letters were typically written on medium-sized sheets (somewhat smaller than A4 or letter-sized paper) that were folded in half, so that each folded sheet of paper would have four pages of written text. Paper quality ranged greatly, probably depending on the writer's financial situation. A few correspondents had their own stationery personalized with their name or a stylized monogram. Writers in mourning often used writing paper with a black border.

A relatively short letter can be seen in Figure 1. Dr. Stjepan Bratanić wrote this letter to Spinčić in 1900 in order to notify him that he had found an apartment for him in the house of a good Christian family. He writes that the family wanted to rent the apartment for 30 forints per day, but that he told them he couldn't pay more than 25. He thinks Spinčić will be able to pay 25, but asks him to contact him if he isn't satisfied, so that he can notify the family. This was when Spinčić was a representative in the Imperial Council in Vienna.

A transcription of this letter can be seen below. The first six lines give the user important information about the letter and contain codes that are partially based on those used in the CEEC, but some codes are specific to this corpus. The first line identifies the collection (Spinčić), the unique identification number of the letter within the Spinčić corpus (year 1900, letter no. 42), and the code of the letter's author. The second line identifies the letter's date. The third and fourth lines indicate the author (A) and recipient (R) of the letter. The fifth line indicates the type of correspondence (L). In this case, the code PI stands for *pismo* 'letter'. Finally, the sixth line indicates the page number. Since all of the correspondence so far transcribed for this corpus are original documents, an attempt has been made to represent them as closely as possible to how they look in the original. In other words, if one line is centered on the page or indented, that line should be spatially positioned similarly in the transcription. Whenever there is a new line in the original text, a new line is started in the corpus transcription. If a word is broken in two at the end of a line, however, that word is kept intact in the corpus and the new line starts with the following word. If it is not clear how the author spelled a particular word or phrase (due to illegible handwriting, damage to the letter etc.), that word or phrase will be placed between two number-signs (see “#prosto#” in the transcription below).

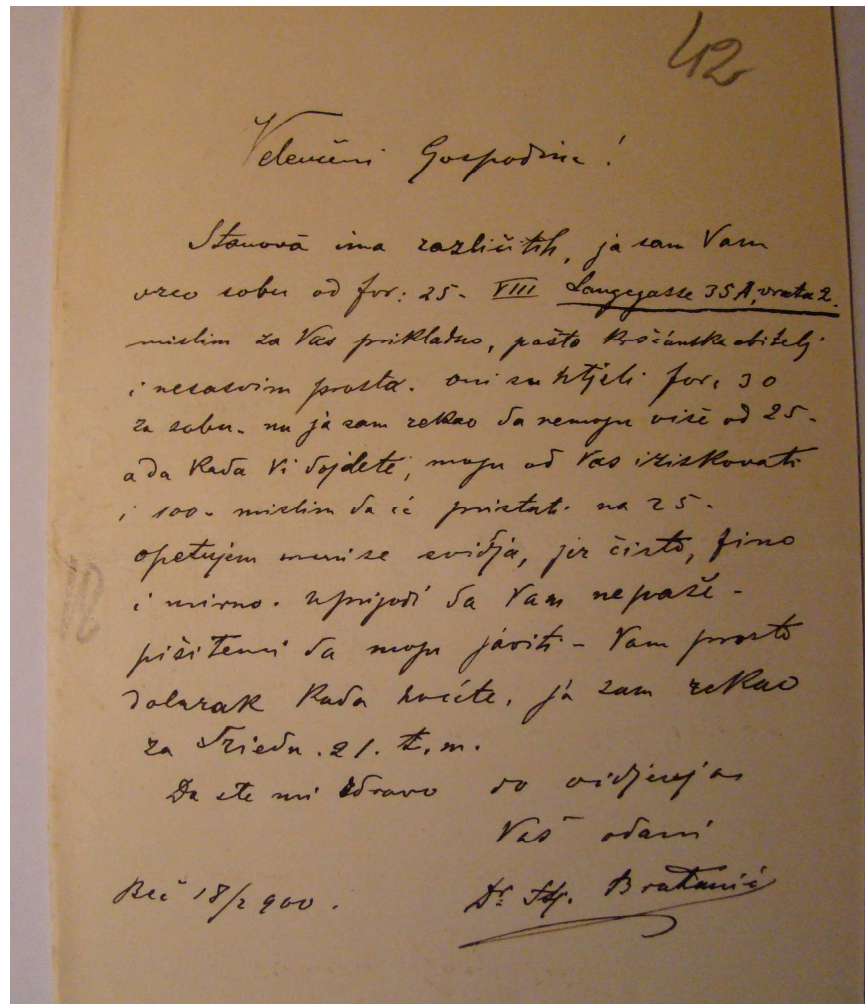


Figure 1

Letter dated 18 February 1900 from Stjepan Bratanić to Vjekoslav Spinčić  
HR-HDA-819. Spinčić Vjekoslav, no. 42, box 86.

Transcription of Letter 1900-042:

<Q SPI 1900-042 SBRATANIC >

<D 1900-02-18>

<A STJEPAN BRATANIĆ>

<R VJEKOSLAV SPINČIĆ>

<L PI>

<P 1>

Veleučeni Gospodin!

Stanova ima različitih, ja sam Vam  
uzeo sobu od for: 25- VIII Langegasse 35A, vrata 2.  
mislim za Vas prikladno, pošto kršćanska obitelj  
i nerasvim prosta. Oni su htjeli for: 30  
za sobu. nu ja sam rekao da nemogu više od 25.  
a da kada Vi dojdete, mogu od Vas iziskovati  
i 100- mislim da će pristati na 25.  
opetujem meni se svidja, jer čisto, fino  
i mirno. uprigodi da Vam ne paše -  
pišitemi da mogu javiti - Vam #prosto#

dolazak kada hoćete. ja sam rekao  
za sriedu. 21. t. m.  
Da ste mi zdravo do vidjenja  
Vaš odani  
Beč 18/2 900. D=r= Stp. Bratanić

This letter was presumably written by a well-educated person, judging by use of the title *Dr.*, yet he almost regularly starts each sentence without a capital letter and rarely ends sentences with periods. One wonders whether he did this on purpose, in the fashion of many emails of today, or whether he had poor writing skills in Croatian.

B. Postcards (*Dopisnice*) – Most of the postcards in this collection are post-office issues without pictures. On one side was a form with spaces for the recipient's name, address, and postal-route information, and the other side was reserved for the sender's message. This type of correspondence was probably cheaper to send than a regular letter. They differed from letters in that the message length was limited, and anyone who handled the letter could read it because it was not sent in an envelope. Postcards in Austria-Hungary differed according to the official languages of the province that issued each one. Those sent from Prague, for example, were bilingual, with the instructional words (e.g. "address,") written in both German and Czech. In most cases the language(s) of the form's instructions would be written at the bottom in fine print. For example, a postcard issued in Prague bear the inscription "(Böhm.)", indicating that (besides German) the instructions on the form were also in Czech. In 1870, postcards issued in Istria were printed in Italian or bilingually in Italian and German. By 1900, they were issued in three languages, German, Croatian, and Italian.

The postcard shown in Figures 2a and 2b was written by Vjekoslav Spinčić's long-time friend and fellow politician Matko Mandić in 1875, when they were both university students. It was sent from Prague to Kastav, as can be seen on the two postmarks. Mandić writes that at the library he was able to renew Spinčić's textbook (Prescott) without any problem. He says that tomorrow he will attempt to do the same with Höfler, wearing a fake moustache, and the next day with Lafluente, wearing a fake beard, so that the librarian will not recognize him. Although the main text is written in the standard language of that time, Mandić adds a message at the top of the card obviously in their local Istrian dialect: *Va preše sam naopako napisal.* 'In a hurry I wrote [this card] upside down.'





Figure 2a: Postcard dated 5 May 1875 from Matko Mandić to Vjekoslav Spinčić  
HR-HDA-819. Spinčić Vjekoslav, no. 36, box 64.

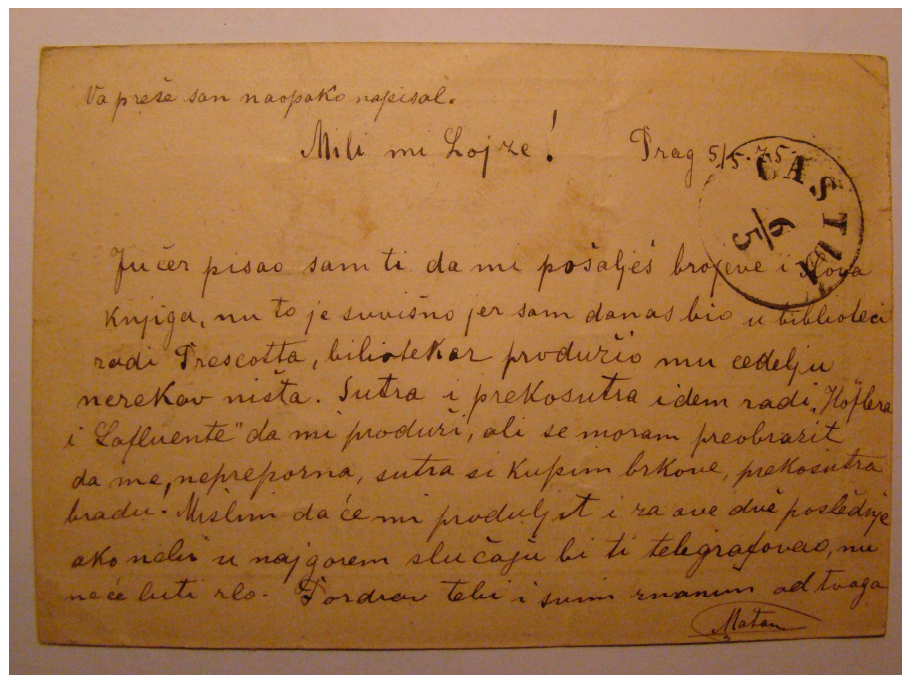


Figure 2b: Message side of postcard in Figure 2a  
HR-HDA-819. Spinčić Vjekoslav, no. 36, box 64

Below is a transcription of the same postcard as it appears in the corpus. In order to give as complete a representation of the document as possible, all of the printed text is

included, but marked with an opening and closing code – i.e. ( ˇCorrespondenz-Karte. ˇ) – so that the user can discern between what was originally printed on the card and what the author wrote.

Transcription of Postcard no. 1875-036:

```
<Q SPI 1875-036 MMANDIC>
<D 1875-05-05>
<A MATKO MANDIĆ>
<R VJEKOSLAV SPINČIĆ>
<L DO>
<P 0>

                ( ˇCorrespondenz-Karte. ˇ)
                ( ˇKorespondenčni listek. ˇ)
( ˇAdresse} ˇ)
( ˇAdresa} ˇ) Gospodinu Vjekoslavu Spinčiću
                absolviranomu filosofu
                ( ˇin ˇ)
                ( ˇv ˇ)u

        über St. Peter                                Kastvu
                                                (Istrien)

                                                ( ˇBöhm. ˇ)

<P 1>
Va preše san naopako napisal.
                Mili mi Lojze!                Prag 5/5 75
Jučer pisao sam ti da mi pošalješ brojeve i slova
knjiga, nu to je suviše jer sam danas bio u biblioteci
radi Prescottta, bibliotekar produžio cedulju
nerekav ništa. Sutra i prekosutra idem radi Höflera
i Lafluente da mi produži, ali se moram preobraziti
da me neprepozna, sutra si kupim brkove, prekosutra
bradu! Mislim da će mi produljiti i za one dne posljednje
ako nebi u najgorem slučaju bi ti telegrafovao, nu
neće biti zlo. Pozdrav tebi i svim znanim od tvoga
                Mate
```

C. Telegrams (*Brzjavke*) – Telegrams were an important means of communication in Austria-Hungary. If a person needed to send an urgent message, one which needed to be delivered on the same day, this was the way of doing it. Since the length of the message determined the price, they tended to be very short. In this corpus, typical telegrams would remind the recipient that he was expected at dinner that evening, notify him of someone's death, or warn him about what he should or should not say at a session of parliament. The telegrams in the Spinčić collection tell us perhaps less about the language of the sender because they provide us with much less text and because transmitting the message required two "middle men" – the sending and receiving telegraph operators – who could be

responsible for linguistic interference or outright errors, especially if they were not native speakers of the language of communication. On the other hand, telegrams do provide a good example of how Austria-Hungary, as it developed into a modern multi-ethnic state, dealt with multilingualism. In Figure 3, all of the printed information on the telegram form is in three languages: German, Croatian, and Italian. This telegram was sent by Mirko Jelušić, mayor of the municipality of Kastav, to Matko Leginja, also a close friend of Spinčić's and a well-known politician. The telegram starts by notifying Leginja of the death and funeral of Eugenija Grosman, but also tells Leginja to relay a request to Spinčić that he attend the local school board meeting in Kastav.

TELEGRAMM  
BRZOJAVKA  
TELEGRAMMA

Eingangs-Nr. 849  
Dolaska Br. 849  
No. d'arrivo 849

Text. — Doslovni sadržaj. — Testo.

9/3 1905

...necis umrla eugenija grosman pogreb subotu pred podne ako  
profesor spinčić tu javite mu stigla tabla kvalifikacija  
aducitaljice ovdje bi li mogao prisustvovati sjednici mjestnoga  
školskoga vjeća cim prije koji dan zupnika veprinackoga birati će  
obcinari veprinca bez obzira na obzira na obcinu kastav danas  
kapatana ovamo sluzbem javlja. - Jelušić nacelnik =

Figure 3

Telegram sent by Kastav Mayor Jelušić to Matko Leginja on 3 September 1905.

HR-HDA-819. Spinčić Vjekoslav, no. 37, box 91.

D. Visiting Cards (*Posjetnice*) – The visiting card is the fourth type of correspondence that can be found in the Spinčić collection. These were cards of varying sizes that were usually left by the owner when he or she visited the recipient but did not find them at home. The printed side of the card had the owner's name, sometimes alone and sometimes with an accompanying phrase or a description of the owner's profession. Most of them have a



message that was written on the blank side, sometimes in minute handwriting, so that, in spite of the small size of the card, the messages on some visiting cards are sometimes as long as those on postcards. The message would often end on the printed side, with the printed name replacing the person's signature after the closing greeting. As with letters, it was common to have special visiting cards printed with a black border when the owner of the card was in mourning.

The visiting card in Figures 4a and 4b were written by Sofija Šabec née Jenko of Podgrad, a Slovenian town north of Kastav, in February of 1900. She apparently visited Špinčič's household, bearing a letter and a turkey sent to him by a certain Miss Živković. Špinčič had been expected to arrive that day but apparently did not, and Šabec writes that they will have to eat it in his absence. It is interesting to note that although the message is written in Slovene, on the front of the card one can see the slogan of the Croatian Party of Rights, *BOG I HRVATI* 'GOD AND CROATS', to which Špinčič, Mandić, and Laginja belonged.

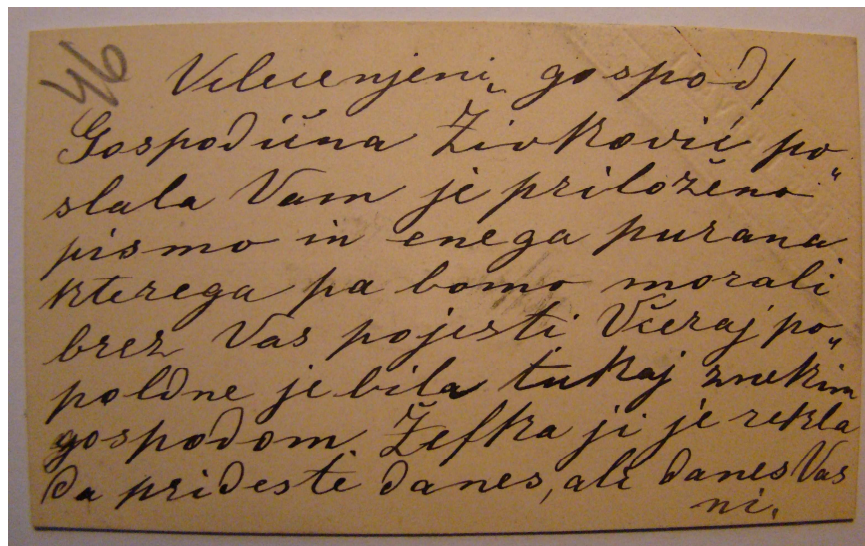


Figure 4a

Back of visiting card of Sofija Šabec

HR-HDA-819. Špinčič Vjekoslav, no. 46, box 86.

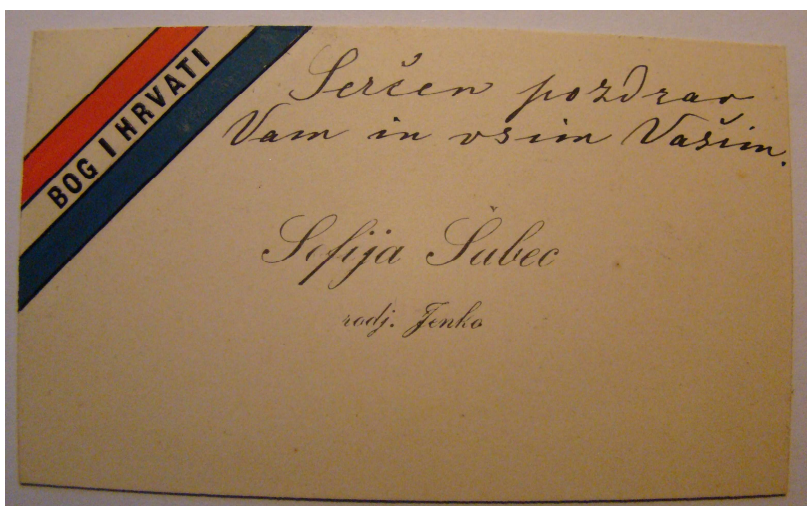


Figure 4b

Front of visiting card of Sofija Šabec

HR-HDA-819. Spinčić Vjekoslav, no. 46, box 86

## 6. Conclusion

The Historical Corpus of Croatian Correspondence is still in its early stages, but hopefully it will steadily grow and eventually be available to users. With this in mind, I feel its contents should be made public via the internet once enough texts have been transcribed and once I have decided upon the best format. This article explains the types of correspondence that are included in the corpus and demonstrates the methods that I use in transcribing them. Obviously, the Spinčić collection cannot be transcribed in full, but if a good selection is made, it is rich enough to provide language produced by a variety of informants with very different social backgrounds. For a scholar interested in linguistic variation and how it relates to language change, this is good data. I also believe that the content of these letters would be interesting to historians, ethnographers, sociologists, and f. who study the Austro-Hungarian period and the aftermath of the First World War, because they provide a wealth of information about how people conducted their everyday lives, what things were important to them, and what problems they had to deal with.

## Literature

- Beuc, Ivan. (1993). *Rukopisna ostavština Vjekoslava Spinčića* (Znanstveno-Obavijesna Arhivska Pomagala 3). Zagreb: Arhiv Hrvatske.
- Corpora of Early English Correspondence*. (2011). Helsinki: Varieng.  
<<http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/index.html>> (30 September 2011).
- Nakaš, Lejla. (2010). *Jezik i grafija krajišničkih pisama* (The language and writing of letters from the border region). Sarajevo: Slavistički komitet.
- Nevalainen, Terttu & Raumolin-Brunberg, Helena (eds.). (1996). *Sociolinguistics and language history: Studies based on the Corpus of Early English Correspondence* (Language and Computers: Studies in Practical Linguistics 15). Amsterdam – Atlanta, GA: Rodopi.
- Novak, Kristian. (2010). *Rekonstrukcija jezičnih biografija pripadnika ilirskoga pokreta - prilog povijesnoj sociolingvistici hrvatsko-njemačkih jezičnih dodira* (The reconstruction of linguistic biographies of members of the Illyrian Movement: A contribution to the historical sociolinguistics of Croatian-German language contact). Ph.D. dissertation. Zagreb: University of Zagreb.
- Raumolin-Brunberg, Helena & Nevalainen, Terttu. (2011). "General Introduction." *Corpora of Early English Correspondence*. Helsinki: Varieng.  
<<http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/generalintro.html>> (30 September 2011).
- Spinčić, Vjekoslav. (1867-1931). Correspondence. HR-HDA-819. Boxes 63-121. Zagreb: Hrvatski državni arhiv/Croatian State Archives.
- Strčić, Mirjana & Strčić, Petar. (1996). "Sazrijevanje Vjekoslava Spinčića". *Hrvatski istarski trolist*. Rijeka: ICR, pp. 97–110.

## Konstrukcija povijesnog korpusa hrvatske korespondencije

U radu se govori o istraživanju jezičnih promjena i varijacija među govornicima hrvatskoga jezika čiji je pisani jezik na prijelazu iz 19. u 20. stojeće bio zahvaćen procesima standardizacije. U tu se svrhu izgrađuje korpus osobnih pisama, sastavljen od četiri tipa pisanih poruka (tradicionalnih pisama, razglednica, posjetnica i telegrama), nastalih u desetogodišnjim intervalima u razdoblju između 1875. i 1925. godine, a koja se čuvaju u Hrvatskom Državnom Arhivu kao dio osobnog fonda istarskog političara Vjekoslava Spinčića (1848-1933). Zbirka se tek manjim dijelom sastoji od pisama koja je napisao sam Spinčić, dok najveći udio u njoj imaju pisma koja su Spinčiću pisali njegovi korespondenti, inače pripadnici svih društvenih slojeva toga vremena. Većina postojećih historijskih korpusa pisane korespondencije temelji se na engleskom i drugim sjevernoeuropskim jezicima, koji imaju dužu tradiciju pismenosti. Ovo je prvi takav projekt za hrvatski jezik, a ujedno i jedno od prvih povijesnih istraživanja hrvatskoga jezika koje se temelji na uporabi svakodnevnog jezika, za razliku od jezika književnosti. U skladu s time, predloženo će istraživanje biti važan doprinos proučavanju jezičnih promjena i varijacija, dok će sadržaj samih pisama, objedinjenih u sustavno izgrađenu i elaboriranu korpusu, zasigurno biti od koristi istraživačima drugih interesnih područja, kao što su povijest i etnologija.