Juraj Benić
Jurica Budja
Zagreb

ANNOTATIONS ABOUT CROATIAN AREAL-DIACHRONIC CORPUS BUILDING

UDK: 811.163.42'32(497.583)
Rukopis primljen za tisak 26. 4. 2023.
Pregledni članak
Survey article

The paper is about the areal-diachronic corpus of the Croatian language. Special attention is given to the data from the corpora of old texts (before the standardisation of Croatian), which would give a complete overview of and insight into the development of the language. The paper also describes the areal-diachronic sub-corpus consisting of works from the Makarska coast.

Key words**:** *Croatian corpus linguistics*, *corpora*, *Croatian language*, *Croatian areal-diachronic corpus*.

1. INTRODUCTION

One of the most prominent features of Croatian is, without doubt, its great internal diversity, considering the relatively small geographical area it covers. This diversity is the result of the interplay of several factors:

1. Dialectal composition. Croatian is made up of three macrodialects (which were, arguably, languages of three separate migrational Slavic waves in the early Middle Ages).

2. Political history. From 1102 until the end of World War I what was to become the Croatian nation had been living under foreign rule (although with various levels

of political autonomy), divided into several political entities (the number of which also varied through time). Some regions (such as Istra) were incorporated into the rest of the Croatian ethnic territory only after World War II.

3. Demographic history. In the 15th and 16th centuries, most of the Croatian ethnic territory was affected by the Ottoman military offensives, advance, and eventual conquest, which led to major migrational shifts that had far-reaching consequences for the development of the Croatian language. When the migrational chaos settled down, the area of Čakavian and Kajkavian dialects had become significantly narrowed and the area of Štokavian had significantly expanded.

From the 1830s the Croatian political élite had been enthusiastic about the possibility of the creation of a nation of Southern Slavs – the Yugoslav nation. Among the Southern Slavs (the Bulgarians aside – they had an independent state of their own and did not express the will to unite with the other Southern Slavs), only the Serbs had an independent state (*de jure* from 1835). As a result, they were regarded by the Croats as the prime force to unite the Southern Slaves, which in turn led to an increased inclination towards the Serbian language as a model for creating a new Croatian standard. The influx of elements of the Serbian language (lexicon, morphology, syntax) into Croatian began in 1830s, increased in the 1890s and has lasted to this day. Because of that, the Croatian language changed in the period of 1830 – 2023 probably more than any other European language did.

One way to advance the research of Croatian before the rapid and profound changes it has undergone for almost two centuries now is to build a dialectal and areal corpus that a) covers the period from the earliest evidence up to 1830 – 1890, and b) optimally addresses the problem of the internal diversity of Croatian. The problem of the internal diversity of Croatian could be most conveniently dealt with by dividing the Croatian language territory into areals that reflect the interrelations between its dialectal constitution, political history, and demographic changes. Twentyone areals have been defined.

2. RELATED WORK

There are several on-line Croatian corpora (the best known of these are *The Croatian National Corpus* and *Croatian Linguistic On-line Repository*), however, they are not exclusively diachronically oriented. *Croatian Linguistic On-line Repository* does contain a certain number of works from before 1830, but it is in no way a systematically organized diachronic corpus. The only diachronic corpus of Croatian that is being developed is *Regensburger Diachrones Korpus des Kroatischen (CroDi)*, but it lacks the areal dimension. Moreover, it is syntactically oriented. The largest extant Croatian corpus is hrWaC, which contains texts crawled from the hr top-level domain, and is therefore purely synchronic.

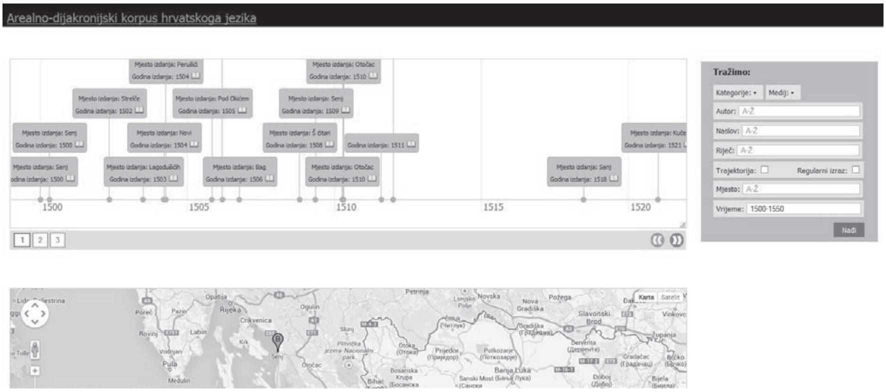
3. CROATIAN AREAL-DIACHRONIC CORPUS

With an areal-diachronic corpus of the Croatian language, all linguistic data would be accurately allocated according to temporal and areal determinants. The body of preserved Croatian language documents is not infinitely large and, in further perspective, a substantial part of it could be entered into the corpus. Each text would have the following determinants (if known): the author's name, the author's gender, the author's social status, name of the text, date of the original, date of the manuscript, page or verse number (or both). The texts are to be classified a) typologically (e. g. Biblical text, *De imitatione Christi* by T. de Kempis, priest handbook on religion, teachings on religion for common folk, discussions on religion, sermons; each type can be further specified, e. g. sermons on Sunday, holiday, Lent, Advent sermons, etc.), b) based on the rhythmic arrangement of the text (prose or verse), c) based on the script (Latin, Glagolitic, Croatian Cyrillic/Bosnian Cyrillic/bosanica). If a Croatian text is a translation from a foreign language, the original text would be available to the user as the related information. Parts of the Croatian text would be connected to parts of the original text at the sentence level.

The access to the corpus would be two-fold (as usual): the regular user would have the option to search and retrieve the information, and the administrator would have the authority to upload and change the information (with various permissions). General users would be able to post their comments next to the retrieved content, as to help the corpus builders in improving the corpus.

The areal-diachronic corpus is created[1] in the web programme which is composed of three basic parts: A) time axis, B) geographic map, C) search engine (Picture 1):

[1] The corpus was originally programmed by Božidar Štimac and Krešimir Milas under the supervision of professor Mario Essert, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb.


Arealno-dijakronijski korpus hrvatskoga jezika
Mjesto izdanja: Peručić Godina izdanja: 1504
Mjesto izdanja: Otočac Godina izdanja: 1510
Mjesto izdanja: Strelče Godina izdanja: 1502
Mjesto izdanja: Pod Okićem Godina izdanja: 1505
Mjesto izdanja: Senj Godina izdanja: 1509
Mjesto izdanja: Senj Godina izdanja: 1500
Mjesto izdanja: Novi Godina izdanja: 1504
Mjesto izdanja: Š ċitari Godina izdanja: 1508
Godina izdanja: 1511
Mjesto izdanja: Kuće Godina izdanja: 1521
Mjesto izdanja: Senj Godina izdanja: 1500
Mjesto izdanja: Lagodušičih Godina izdanja: 1503
Mjesto izdanja: Bag Godina izdanja: 1506
Mjesto izdanja: Otočac Godina izdanja: 1510
Mjesto izdanja: Senj Godina izdanja: 1518
1500
1505
1510
1515
1520
1
2
3
Karta
Satelit
Rijeka
Senj
Otočac
Slavonski Brod
Banja Luka

Tražimo:

Kategorije: ▾ Medij: ▾

Autor: A-Ž

Naslov: A-Ž

Riječ: A-Ž

Trajektorija: □ Regularni izraz: □

Mjesto: A-Ž

Vrijeme: 1500-1550

Nađi

Picture 1. Time axis (top row left); geographic map (bottom row left); search engine (top row right)

A) Time axis. On the time axis either the retrieved documents appear (of any category and/or medium) or the history of a word can be traced. The time axis can be zoomed in or out (with a mouse, along the abscissa) ranging from the largest to the smallest time units, and if a time period has a larger number of documents, it can also be moved along the vertical line (the ordinate). In that case, the display «bubble« with the document finds the appropriate place for viewing. If there are too many documents found to be displayed on the screen simultaneously, pages are created which can be relatively (<<, >>) or absolutely moved (by entering the page number) in the given range. The time axis would show in which period large historical changes happened, as well as how, with the movement of the people, the words moved from one areal to another. Due to the re-settlements in the 15th, 16th and 17th centuries, some areals would show huge losses of territory (e. g. North Čakavian areal), and others, precisely due to the changes in other areals, would emerge (e. g. Lika Štokavian areal, areal of the Šibenik Hinterland).

B) Geographic map. By defining the place of origin for each document, a mark on the Google map would be created. With a click on the mark, the stored document would be opened. A cluster of geographic places can be connected to one areal, which can be determined by the main criterion of geopolitical-dialectal relations in a certain period, as mentioned before. By clicking on the areal, all of the areal's documents appear on the time axis. E.g. by clicking on the *Slavonian-Srijem areal*, all the documents from 1651 onwards appear, as before 1651 in this part of the Croatian linguistic space no preserved testimony of literacy in Croatian has been found.

Geographic distribution of the data, from texts to words or grammatical forms, would reveal mutual strong affinities of certain areals, the cause of which could be traced either upwards (dialectal basis) or downwards (political basis). Political basis implies belonging to the same secular organization (a state) or a religious one (a province – e. g. Bosna Srebrena). Let us take, for illustration, the research possibility of one lexical category, the so-called *verba dicendi*. At a certain time in its history, Croatian had seven core *verba dicendi* (*reći*, *kazati/povidjeti*, **veljeti* (verb without the infinitive form), *diti*, *praviti*, *govoriti*). Today only two remain (*reći* and *kazati* in suppletion and *govoriti*). With the help of an areal-diachronic corpus, it would be possible to reconstruct this in detail:

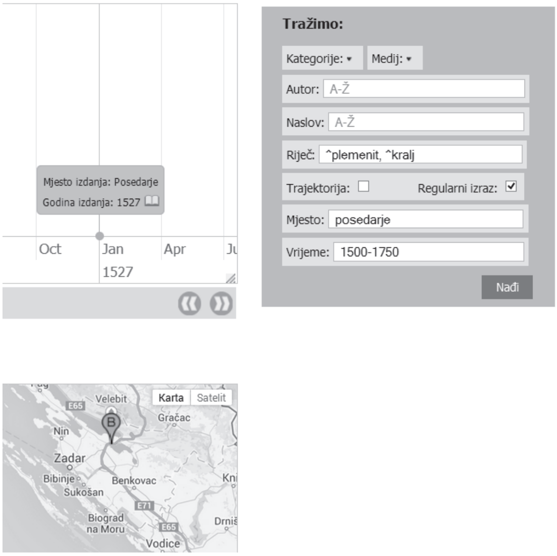
– Areal distribution. E. g. a complementary geographic distribution of the verb *kazati* and *povidjeti* could be described, together with the zone of their overlapping; the relation to the verb *veljeti,* which covered most of the Croatian linguistic space, could be clarified, and also the verb *diti*, characteristic of Čakavian areas; the placement of the verb *praviti* could be determined, which is characteristic of south-western Croatian linguistic space, and so on.

– Historical changes. It would be discernible, for example, how the verbs *praviti* and especially *diti* have been losing their vividness; that in the northern Croatian linguistic space **veljeti* has, to a certain extent, replaced the verb *reći;* that in the north-eastern Croatian linguistic space *kazati* has partly replaced *reći;* that, under the influence of Serbian, *reći* and *kazati* have merged into one suppletive verb, while **veljeti* has been disappearing from use, and so on.

– Semantic nuances. Each of the verbs has its own semantic nuance that would manifest only when it is possible to operate with a large number of occurrences for all seven verbs with a very broad segment of their context.

Eventually, the conditions would be created to answer the question of how it is that Croatian, with seven core *verba dicendi,* which is a lot in the European context, has been reduced to only two of them. Moreover, the changes in the category of *verba dicendi* could be compared with the changes in other semantic categories of Croatian verbs.

C) Search engine. A search engine has been developed to search the documents by author, title, place, or time, as well as by the word(s) from the document, depending on the chosen category and media in the repository. This type of search engine is designed for all possible combinations of the above research areas.



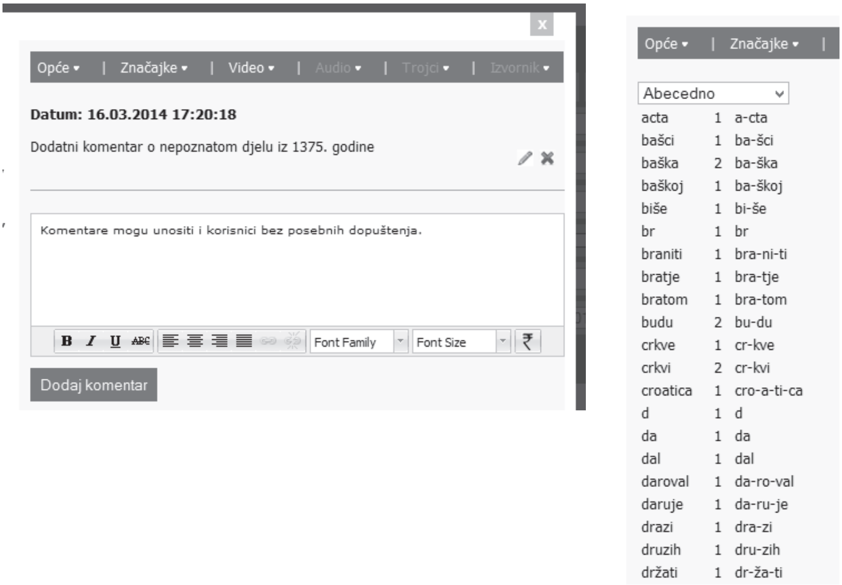
Mjesto izdanja: Posedarje
Godina izdanja: 1527

Tražimo:
Kategorije: ▾ Medij: ▾
Autor: A-Ž
Naslov: A-Ž
Riječ: ^plemenit, ^kralj
Trajektorija: □ Regularni izraz: ☑
Mjesto: posedarje
Vrijeme: 1500-1750
Nađi

Picture 2. Example of having defined a) place (Posedarje), b) time span (1500 – 1750), and c) key words (*plemenit* ‘noble’ and *kralj* ‘king’)...

4. BASIC AND ADDITIONAL INFORMATION

The retrieved information, a document in the corpus, opens with a click on the marker on the map (pin) or the marker in the bubble (book) on the time axis. The window with the traced information always has two parts: the main part (on the left) and the additional part (on the right). In the main part, the basic information is displayed (e. g. a text), and on the rightside additional information can be added if needed. The additional information can be user comments, certain components of the main information (e. g. a dictionary with all the words occurring in the text in all their grammatical forms, with the possibility of sorting them by their alphabetical order or by the frequency of their occurrence), audio or video recordings attached to the document, document’s triplestore LLOD – linguistic linked open data and links to the document, saved for example in PDF form or already published elsewhere. Currently, the components of the main text include only the list of words, but in the future the text is going to be visually marked by different criteria, e. g. lemma tagging, syntactic function labelling, sentence structure tagging (bracketing layer), syntactic category annotation, semantic role labelling, sense tagging and similar. A special visual tool, TEIMark, has been designed for these types of operations.



Datum: 16.03.2014 17:20:18

Dodatni komentar o nepoznatom djelu iz 1375. godine

Komentare mogu unositi i korisnici bez posebnih dopuštenja.

Abecedno

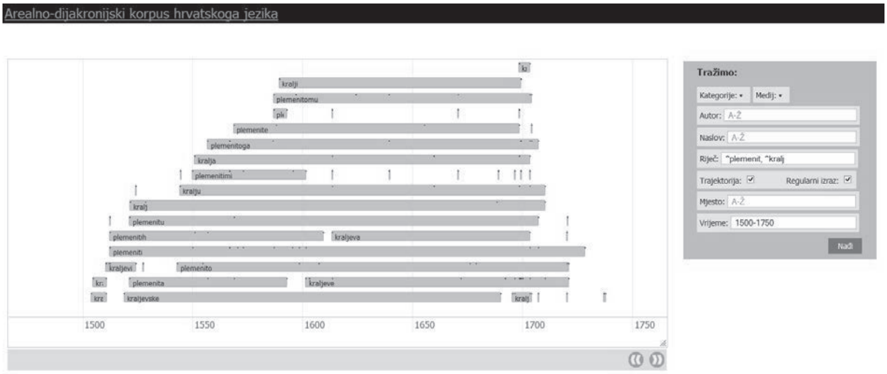
acta	1	a-cta
bašci	1	ba-šci
baška	2	ba-ška
baškoj	1	ba-škoj
biše	1	bi-še
br	1	br
braniti	1	bra-ni-ti
bratje	1	bra-tje
bratom	1	bra-tom
budu	2	bu-du
crkve	1	cr-kve
crkvi	2	cr-kvi
croatica	1	cro-a-ti-ca
d	1	d
da	1	da
dal	1	dal
daroval	1	da-ro-val
daruje	1	da-ru-je
drazi	1	dra-zi
druzih	1	dru-zih
držati	1	dr-ža-ti

Picture 3. General comments (left) and components (right)

All the materials (main and additional) are placed, replaced or deleted in a simple way; the only thing needed is a WEB-viewer (browser). The uploaded document can be worked on on-line with a tool similar to Word. For the purposes of using Croatian characters with accents, a module is developed for the insertion of any of the UTF-8 symbols.

One of the especially valuable possibilities of the web tool for the tracking of the corpus in space and time are the trajectories of the words – they could also be tracked in these frames, selecting them either in the usual or structured way (via the regular expressions). In that way, the life of the Croatian words through the centuries could be studied: their origin, development, and disappearance.

The black dots on the trajectory refer to the documents (their time coordinate) in which the word appears, and by turning off the option of «trajectory«, the user can open the chosen document. It is a powerful tool for monitoring and tracing lexical changes, sometimes very turbulent, which Croatian vocabulary underwent, particularly in the 19th century.



Arealno-dijakronijski korpus hrvatskoga jezika

Tražimo:

Kategorije: ▾ Medij: ▾

Autor: A-Ž

Naslov: A-Ž

Riječ: ^plemenit, ^kralj

Trajektorija: ☑ Regularni izraz: ☑

Mjesto: A-Ž

Vrijeme: 1500-1750

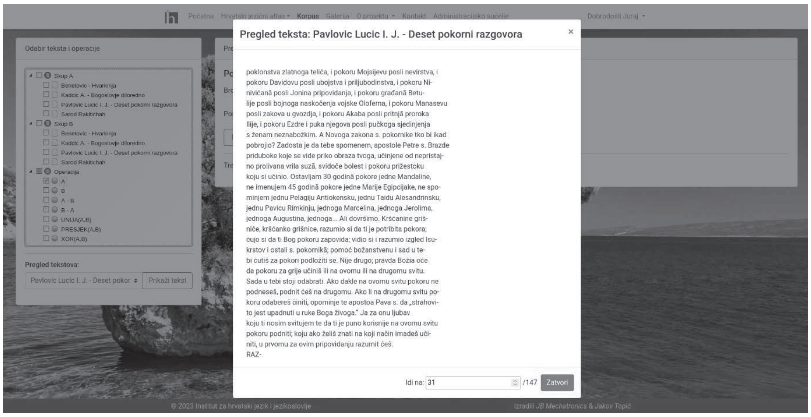
Nađi

Picture 4. Tracking the duration of words

5. AREAL-DIACHRONIC SUB-CORPUS OF TEXTS FROM THE MAKARSKA COAST

As part of the project Dialects of Makarska coast – diachrony and synchrony, an areal-diachronic corpus of texts written in the Makarska coast area was compiled. These diachronic linguistic data are compared with synchronic dialectological data. The corpus consists of three chronicles from the 17th and 18th centuries; 18th-century works of A. Kačić Miošić and I. J. Pavlović Lučić; and 19th-century works of don M. Pavlinović, who collected folk literature of the Makarska coast along with the father and son Alačević. So far there are several available online: «Hvarkinja« by M. Benetović, «Bogoslovje diloredno« by A. Kačić, «Deset pokorni razgovora« by I. J. Pavlović Lučić (Picture 5.). Books prepared for computer processing have previously been transcribed. The program makes it possible to research the concordance of each individual book, to determine which words are found in all books or which words appear in only some or only one of the books. This makes it easier to research the language of each individual writer or individual period.

In parallel with the digital corpus, the transcribed works contained in the corpus are also published, with accompanying introductory studies providing a detailed linguistic analysis. (Budja 2022). The project »…will enable the analysis of continuity of the language from the oldest written monument until today« (Benić, Filipić, 2021: 489).


Početna
Hrvatski jezični atlas
Korpus
Galerija
O projektu
Kontakt
Administracijsko sučelje
Dobrodošli Juraj
Odabir teksta i operacije
Skup A
Benetovic - Hvarkinja
Kadcic A. - Bogoslovje diloredno
Pavlovic Lucic I. J. - Deset pokorni razgovora
Sarod Rakitichah
Skup B
Benetovic - Hvarkinja
Kadcic A. - Bogoslovje diloredno
Pavlovic Lucic I. J. - Deset pokorni razgovora
Sarod Rakitichah
Operacija
A
B
A - B
B - A
UNIJA(A,B)
PRESJEK(A,B)
XOR(A,B)
Pregled tekstova:
Pavlovic Lucic I. J. - Deset pokor
Prikaži tekst
Idi na: 31
/147
Zatvori
© 2023 Institut za hrvatski jezik i jezikoslovlje
Izradili JB Mechatronics & Jakov Topić

Pregled teksta: Pavlovic Lucic I. J. - Deset pokorni razgovora

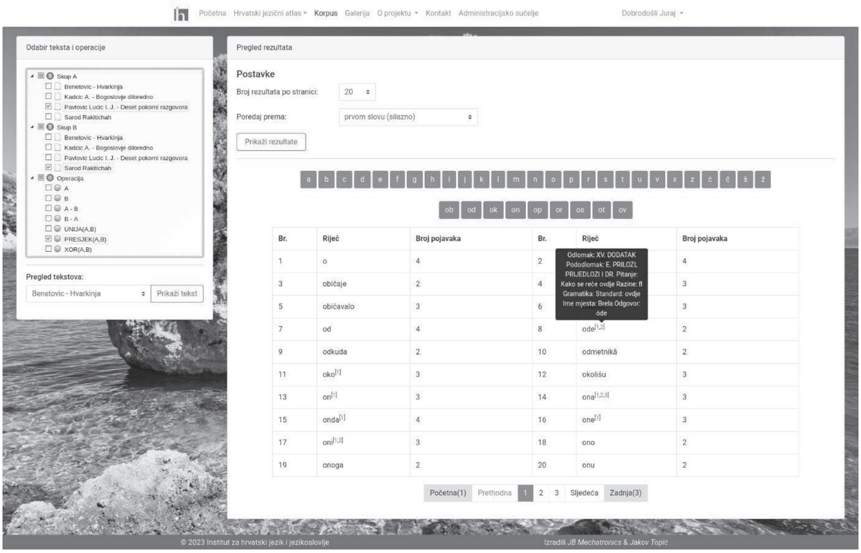
poklonstva zlatnoga telića, i pokoru Mojsijevu posli nevirstva, i
pokoru Davidovu posli ubojstva i priljubodinstva, i pokoru Ni-
nivićanā posli Jonina pripovidanja, i pokoru građanā Betu-
lije posli bojnoga naskočenja vojske Oloferna, i pokoru Manasevu
posli zakova u gvozdja, i pokoru Akaba posli pritnjā proroka
Ilije, i pokoru Ezdre i puka njegova posli pučkoga sjedinjenja
s ženam neznabožkim. A Novoga zakona s. pokornike tko bi ikad
pobrojio? Zadosta je da tebe spomenem, apostole Petre s. Brazde
priduboke koje se vide priko obraza tvoga, učinjene od nepristaj-
no prolivana vrila suzā, svidoče bolest i pokoru prižestoku
koju si učinio. Ostavljam 30 godinā pokore jedne Mandaline,
ne imenujem 45 godinā pokore jedne Marije Egipcijake, ne spo-
minjem jednu Pelagiju Antiokensku, jednu Taidu Alesandrinsku,
jednu Pavicu Rimkinju, jednoga Marcelina, jednoga Jerolima,
jednoga Augustina, jednoga... Ali dovršimo. Kršćanine griš-
niče, kršćanko grišnice, razumio si da ti je potribita pokora;
čujo si da ti Bog pokoru zapovida; vidio si i razumio izgled Isu-
krstov i ostali s. pokornikā; pomoć božanstvenu i sad u te-
bi ćutiš za pokori podložiti se. Nije drugo; pravda Božia oće
da pokoru za grije učiniš ili na ovomu ili na drugomu svitu.
Sada u tebi stoji odabrati. Ako dakle na ovomu svitu pokoru ne
podneseš, podnit ćeš na drugomu. Ako li na drugomu svitu po-
koru odabereš činiti, opominje te apostoa Pava s. da „strahovi-
to jest upadnuti u ruke Boga živoga." Ja za onu ljubav
koju ti nosim svitujem te da ti je puno korisnije na ovomu svitu
pokoru podniti; koju ako želiš znati na koji način imadeš uči-
niti, u prvomu za ovim pripovidanju razumit ćeš.
RAZ-

Picture 5. The complete transcribed text is available in the corpus part of the project

In history, the Makarska coast was an enclosed geographical and political unit due to its geographic position. In the North it borders Poljica and in the South the Neretva Valley. This political constellation has existed since ancient times. It correlates to today's dialectological image, in which the dialect of coastal Poljica is Čakavian, the one spoken on Makarska coast is Neo-Štokavian šćakavian, and the Neretva Valley is Neo-Štokavian štakavian.

The idioms of the Makarska area belong to the Neo-Štokavian ikavian. On the south it borders with the Eastern Herzegovinian, i.e. Neo-Štokavian ijekavian, and on the islands just across Makarska's coastal area the local idioms are Southern Čakavian i.e. Čakavian ikavian dialect. On the Pelješac peninsula, the local idioms are separated in a way that the western side of the island is Southern Čakavian while the eastern side is Neo-Štokavian.[2] The language contact between the Štokavian and Čakavian dialects in the Makarska coast is very interesting. Some characteristics of the Neo-Štokavian dialects of the Makarska coast are found in the South Čakavian dialects, but are not found in the nearby Štokavian dialects (lengthening before sonants, i.e. *sîr*, or consonant cluster simplification, i.e. *kȕška*, *mȁška*). Dialectological data are collected according to the HJA (Croatian linguistic atlas) questionnaire. These data are computer-linked with textual data, which allows the development of a particular language line to be followed over the course of several centuries (Picture 6).

[2] The most recent work on the local idioms of the Pelješac peninsula is the thesis of M. Tomelić Ćurlin *Jezične posebnosti peljeških govora. Fonologija* (2019).


Početna
Hrvatski jezični atlas
Korpus
Galerija
O projektu
Kontakt
Administracijsko sučelje
Dobrodošli Juraj
a b c d e f g h i j k l m n o p r s t u v x z ć č š ž
ob od ok on op or os ot ov
Početna(1) Prethodna 1 2 3 Sljedeća Zadnja(3)
© 2023 Institut za hrvatski jezik i jezikoslovlje
Izradili JB Mechatronics & Jakov Topić

Odabir teksta i operacije

- Skup A
  - ☐ Benetovic - Hvarkinja
  - ☐ Kadcic A. - Bogoslovje diloredno
  - ☑ Pavlovic Lucic I. J. - Deset pokomi razgovora
  - ☐ Sarod Rakitichah
- Skup B
  - ☐ Benetovic - Hvarkinja
  - ☐ Kadcic A. - Bogoslovje diloredno
  - ☐ Pavlovic Lucic I. J. - Deset pokomi razgovora
  - ☑ Sarod Rakitichah
- Operacija
  - ☐ A
  - ☐ B
  - ☐ A - B
  - ☐ B - A
  - ☐ UNIJA(A,B)
  - ☑ PRESJEK(A,B)
  - ☐ XOR(A,B)

Pregled tekstova: Benetovic - Hvarkinja | Prikaži tekst

Pregled rezultata

Postavke

Broj rezultata po stranici: 20

Poredaj prema: prvom slovu (silazno)

Prikaži rezultate

Br.	Riječ	Broj pojavaka	Br.	Riječ	Broj pojavaka
1	o	4	2		4
3	običaje	2	4		3
5	običavalo	3	6		3
7	od	4	8	ode[1,2]	2
9	odkuda	2	10	odmetnikā	2
11	oko[1]	3	12	okolišu	3
13	on[1]	3	14	ona[1,2,3]	3
15	onda[1]	4	16	one[1]	3
17	oni[1,2]	3	18	ono	2
19	onoga	2	20	onu	2

Odlomak: XV. DODATAK Pododlomak: E. PRILOZI, PRIJEDLOZI I DR. Pitanje: Kako se reče ovdje Razine: fl Gramatika: Standard: ovdje Ime mjesta: Brela Odgovor: óde

Picture 6. On the left is a list of transcribed works that have been prepared for digital search

There is a check mark next to the book *Deset pokorni razgovora* by I. J. Pavlović Lučić, which means that its language and concordance will be studied. On the right is a box in which data from the marked book are compared with dialectological data.

This sub-corpus offers the possibility of comparing diachronic and synchronic data and has been computerized in the way that, by adding sub-corpora from other areas, it will be possible to compare historical and dialectological data for the entire Croatian language area.

6. CONCLUSION AND FUTURE WORK

Without an all-encompassing corpus of the Croatian language, which would contain all the historic and current texts, the historical development of the Croatian language cannot be described. A corpus would help illuminate the dialectal diversity, describing the proto-standard Croatian languages in the 18th century. An areal-diachronic corpus of Croatian is equipped with all the necessary computer tools for text searching at all grammatical and lexical levels. This corpus should in the future be broadened by adding the corpus of spoken language (standard language as well as dialects) to complete and frame the obtained picture of the Croatian language.

LITERATURA

Budja, J. (2022). *»Deset pokorni razgovorā« Ivana Josipa Pavlovića Lučića.* Zagreb: Institut za hrvatski jezik i jezikoslovlje.

Benić, J. i Filipić, L. (2021). A synchronic and diachronic computer corpus of Makarska littoral dialects (Croatia). *Jazykovedný časopis*, 72, 2, 488–501.

Tomelić Ćurlin (2019). *Jezične posebnosti peljeških govora. Fonologija.* Split: Filozofski fakultet Sveučilišta u Splitu.

Hansack, E. Hansen, B., Wald, V., Horvat, M. i Perić Gavrančić, S. (2016). Regensburški dijakronijski korpus hrvatskoga jezika – CroDi. *Rasprave: časopis Instituta za hrvatski jezik i jezikoslovlje*, 42, 1, 1–19.

Mrežni izvori
<http://riznica.ihjj.hr/philologic/Cijeli.whizbang.form.hr.html>
<https://bastina.jezik.hr/gomapri/o_projektu?tip=opci_podatci>

NAPOMENE UZ (IZ)GRADNJU AREALNO-DIJAKRONIJSKOGA KORPUSA HRVATSKOGA JEZIKA

Sažetak

U radu se donosi opis arealno-dijakronijskoga korpusa hrvatskoga jezika. Posebna se pozornost posvećuje podatcima iz korpusa starih tekstova nastalih prije standardizacije hrvatskoga jezika, a koji pomažu za cjelovit pregled i uvid u i na razvoj hrvatskoga jezika. Dalje se u radu opisuje iz korpusa izabrani arealno-dijakronijski potkorpus, izgrađen iz pisanih djela i tekstova nastalih na Makarskome primorju.

Ključne riječi: *hrvatska korpusna lingvistika*, *korpusi*, *hrvatski jezik*, *hrvatski arealno-dijakronijski korpus*.

NOTE ALLA COSTRUZIONE DI UN CORPUS AREALE-DIACRONICO DELLA LINGUA CROATA

Riassunto

Nel lavoro si presenta la descrizione di un corpus areale-diacronico della lingua croata. Si dedica particolare attenzione ai dati dei corpora dei testi antichi risalenti a prima della standardizzazione della lingua croata e che sono d'ausilio per un compendio e una disamina completi dello sviluppo della lingua croata. Nel seguito del lavoro si descrive un sub corpus areale-diacronico, costituito da opere e testi attinenti la Riviera di Makarska.

Parole chiave: *linguistica croata dei corpora*, *corpora*, *lingua croata*, *corpus areale-diacronico croato*.

Podatci o autorima

Dr. sc. Juraj Benić zaposlen je na Fakultetu strojarstva i brodogradnje Sveučilišta u Zagrebu na Katedri za strojarsku automatiku. Njegovo područje rada obuhvaća mrežno programiranje te regulacija mehatroničkih sustava.
E-adresa: juraj.benic@fsb.hr

Dr. sc. Jurica Budja zaposlen je u Institutu za hrvatski jezik na Odjelu za opće jezikoslovlje. Znanstveni mu je interes osobito upravljen na tvorbu riječi i povijest hrvatskoga jezika.
E-adresa: jbudja@ihjj.hr