

A SYNCHRONIC AND DIACHRONIC COMPUTER CORPUS OF MAKARSKA LITTORAL DIALECTS (CROATIA)

JURAJ BENIĆ¹ – LOBEL FILIPIĆ²

¹ Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Zagreb, Croatia

² Institute of Croatian Language and Linguistic, Zagreb, Croatia

BENIĆ, Juraj – FILIPIĆ, Lobel: A synchronic and diachronic computer corpus of Makarska littoral dialects (Croatia). *Journal of Linguistics*, 2021, Vol. 72, No 2, pp. 488 – 501.

Abstract: This paper presents a synchronic and diachronic computer corpus of Makarska littoral dialects. This corpus was created as part of the project to explore the ikavian neoštokavian dialects of the narrow coastal area in Croatian region of Dalmatia around the town of Makarska. The dialectological characteristics of the dialects studied are briefly presented first, followed by presentation of the digital system. The system is logically organized in first part as a corpus of literary texts created from 1729 to 1803 and digitally processed, and in the second part from the materials collected through dialectological questionnaires prepared and methodologically adapted as part of the creation of the Croatian Linguistic Atlas. Methods of collecting linguistic data, method of input into the digital form and methods and possibilities of data processing will be explained. Based on the input and search strategies within the system, the examples will prove the origin of the dialects of the Makarska littoral to be that of the ikavian neoštokavian dialect described in the dialectological literature. This computer-based principle of work is a novelty in Croatian dialectology which has not been digitally processed so far and offers a basis for future dialectological research. This platform can be used in order to shorten the time of data processing and to analyse them more systematically and more efficiently. So far, there has been no such digital repository for any Croatian speech. This project represents a thorough synchronic and diachronic study of one rounded language area.

Keywords: spoken corpus, corpus design, computer corpus, dialect corpus, dialectology, štokavian

1 INTRODUCTION

In 2016, a scientific research project was registered in Croatia with project manager Ivana Kurtović Budja, PhD. The project applied to the Croatian Science Foundation for its funding, and its aim was to conduct dialectological research at the designated points and to record material that will be archived physically and digitally. The default points are located in the Makarska littoral in Dalmatia, Croatia's southern region. The dialects of this area belong to the štokavian dialect, one of three Croatian

dialects (štokavian, čakavian, kajkavian). The characteristics of this dialect with confirmations obtained on the ground will be briefly described in the paper, and the entire material will be computer-processed and entered into the synchronic and diachronic computer corpus which will be accessible online and searchable in synchronic and diachronic mode and also searchable by different language criteria with audio recordings from native speakers.

2 RESEARCH OBJECTIVE AND METHODOLOGY

The aim of the project is a study from the historical linguistic aspect which will monitor specific Makarska littoral linguistic lines. The monitoring process is based on the corpus of old texts from the Makarska littoral, therefore, it will enable the analysis of continuity of the language from the oldest written monuments until nowadays. In addition, a sociolinguistic survey on the attitudes the young speakers have towards the local dialect and a dialect analysis according to different age groups and education levels will be conducted. The dialects of Brela, Podgora, Makarska, Igrane, Zaostrog, Živogošće, Baška Voda, Gradac, Promajna, Tučepi, Drvenik, Drašnice were studied. Apart from them, as control points, the dialects of Žrnovnica, Zadvarje, Raščane, Maslinica, Sumartin, Sućuraj, Račišće, and the three Croatian idioms in Molise (Italy) on phonological, morphological, syntactic, and lexical level will be conducted. Informants, i.e., speakers of organic idioms, i.e., examinees, were audio recorded and questionnaires containing demographic data on informants were manually filled in. Each speech was then dialectologically described, materials were prepared for the synchronic part of the corpus, as were representative samples of the recorded natural idiom of every location to be used as a material basis for the spoken language corpus. Selected texts from the Makarska littoral, created from 1729 to 1803, were prepared for the text corpus.

The research was organised in several phases and at several levels; the dialectological research is carried out by finding reliable informants who are examined according to the standardized questionnaire for the Croatian Linguistic Atlas. All of the recording is approved by the informant with the statement of approval for participation in the project. Sociolinguistic research is conducted using also the method of field research, filling out questionnaires and processing materials. Historical linguistic and textological research is conducted on the basis of scanned and transcribed materials. Together, everything is prepared for entry into the synchronic and diachronic corpus of Makarska littoral dialects. The obtained results will be computer-processed and searchable according to phonological, morphological, and syntactic characteristics from a synchronic and diachronic aspect, which will demonstrate the continuity or discontinuity of each of the idioms.

3 AREA OF RESEARCH: GEOGRAPHICAL AND DIALECTOLOGICAL FEATURES

3.1 The Makarska littoral

In history, the Makarska littoral was a closed geographical and political unit because of its geographical location. It extends along a narrow coastal area of sixty kilometers in the Croatian region of Dalmatia.

3.2 Dialectological features

The Makarska littoral dialects belong to ikavian neoštokavian dialect [1] (Fig. 1). This dialect borders the eastern Herzegovinian and the neoštokavian ijekavian dialect on the south, and on the sea, just opposite to the Makarska littoral, there is the south-čakavian or ikavian čakavian dialect. The western Bosnian-Herzegovinian dialect, i.e., the ikavian štokavian dialect, is one of the seven štokavian dialects [2]. The largest ikavian štokavian unit consists of western Herzegovina, the Dalmatian mainland, a part of Lika and parts of western and central Bosnia. Today's borders of this dialect were set after the 15th century, in the third period of Croatian language development [3], when Turkish breakthroughs triggered large migrations in the population. For these nonlinguistic reasons, the dialects of today's ikavian štokavian dialect span larger and smaller unconnected regions. It was part of the western-štokavian dialect of the Croatian language, which fell apart after the Ottoman conquest.

The neoštokavian dialects are marked by the so-called neoštokavian language innovations in many dialects. It is about the neoštokavian accentuation (four-accent system), and mainly a systematic transfer of falling accents toward the beginning of the word and to proclitics (Nsg. *òko*, Lsg. *ù oku*). The letter *l* in final position gave way to *o* or *a* (inf. *biti*, r. pr. m. sg. *bìja*), the letter *h* is lost (in standard Croatian *hràst* ('oak'), in dialect Nsg. *ràst*). *Jat* is ikavian (in standard Croatian *mlijéko* ('milk'), in dialect Nsg. *mliko*). The researched dialects are šćakavian (in standard Croatian *iskati*, *tražiti* ('to look for'), in dialect 3. pl. praes. *išćū*). The letter *-m* in final position in suffixes and in inflectional words changed to *-n*. Numerous romanisms as well as turcisms are present in the vocabulary [4]. The group *ra* has two versions; *ra* (in standard Croatian *ràsti* ('to grow'), in dialect inf. *ràst* (Baška Voda)) and *re* (in standard Croatian *ràsti* ('to grow'), in dialect inf. *rést* (Brela)). The phoneme *h* is lost or rare mainly everywhere, and usually switches to *j* (in standard Croatian *grijéh* ('sin'), in dialect Nsg. *grīj*), *v* (in standard Croatian *bùha* ('flea'), in dialect Nsg. *búva*) or disappears in the initial position (in standard Croatian *hlâd* ('shade'), in dialect Nsg. *Lâd*) [5]. All these more distinctive features, as well as many other specific features for this dialect, can be attested for in the corpus, which achieves one of its objectives to make it digitally easily available for dialectological research.



Fig. 1. Map of štokavian dialect with the Makarska littoral marked as follows ([6], [7]) (dialects in the map legend from top to bottom: neoštokavian ikavian, neoštokavian ijekavian, neoštokavian ekavian, slavonian, eastbosnian, zetski)

4 DESIGNING THE SYNCHRONIC AND DIACHRONIC COMPUTER CORPUS OF MAKARSKA LITTORAL DIALECTS

4.1 Collecting materials

The material constituting the corpus consists of data collected from the questionnaire for the Croatian Language Atlas, old texts from the area of the Makarska littoral, and the control texts of other dialects. The texts currently in the digital corpus are Hvarkinja by Martin Benetović (around 1550–1607), Bogoslovje diloredno by Antun Kadčić (1686–1745), Deset pokorni razgovora by Ivan Jožip

Pavlović Lučić (1755–1818) and *Sarod Rakitichah* by Petar Rakitić. The project contains further old texts which have been processed in optical character recognition program (OCR), then manually examined for errors. Those texts are *Razgovor ugodni naroda slovinskoga* (1756) and *Korabljica* (1760) by Andrija Kačić Miošić, *Tridentinskoga sabora naredbe* (1790), *Commentarii morales* (1793) and *Dvi bogoljubne pofale* (1803) by Ivan Jozip Pavlović Lučić. The basis for having those texts in corpus is the birthplaces of the authors which are in Makarsko primorje region. Exceptions are *Hvarkinja* by Martin Benetović and *Sarod Rakitichah* by Petar Rakitić. Those two authors are not from the region, neither are the texts written in the dialect of Makarsko primorje region. They have been chosen as control texts. *Sarod Rakitichah* is štokavian text of other area, not the same as in Makarsko primorje. *Hvarkinja* is a renaissance commedia ridicolosa but with many features common to commedia erudita. It has been chosen as a control text of another dialect but also because one of the characters in the drama is from Makarsko primorje region and speaks in its own vernacular. The idea for organizing the old text corpus, selection of texts and all the major work has been done by Jurica Budja, PhD, an associate on the project. Budja has also reedited *Deset pokorni razgovora* and wrote the introductory study. The plan is to do the same with all the old texts, i.e., books. All of the mentioned texts are part of the all-Croatian corpus, a much bigger set of texts from different authors, regions, dialects, and periods.

4.1.1 Language materials of the Croatian Language Atlas

Croatian local dialects were explored for multilingual atlases: the General Slavic Linguistic Atlas (28 points), the European Linguistic Atlas – Atlas linguarum Europae and the Central South Slavic Dialectological Atlas (called the Croatian-Serbian Dialectological Atlas) (236 points). In 1996, the Croatian Language Atlas (HJA) project was initiated at the Institute of Croatian Language and Linguistics, for which a network of 399 points of Croatian dialects (101 čakavian, 110 kajkavian, 188 štokavian) is envisaged [8]. The Croatian Language Atlas is a set of descriptions of Croatian dialects. The descriptive material was collected by systematic field work, that is, audio and written recording of language status at points in Croatia and locations outside Croatia in which Croatian language is spoken. Researchers, most often dialectologists, in conversation with the informants record linguistic characteristics which are agreed in advance and according to which the field conditions on the phonological (F or f), prosodic (P or p), morphological (M or m), word formation (T or t), lexical (L or l) and syntax (S or S) level can be closely monitored (Fig. 2).

Ksk: u stanju sam to uraditi, ja to...				
fpmsm	18.	1. sg. praes.	<i>mogu</i>	<u>möven</u>
PM	19.	3 sg. praes.	#	<u>möre</u>
FPM	20.	3. pl. praes.	#	<u>mövemo</u>
M	21.	3. sg. aor.	#	<u>-</u>
FPM	22.	r. pr. m. sg.	#	<u>mögä</u>
Ksk: on to nije u stanju uraditi, on to...				
FP	23.	3. sg. praes.	<i>ne može</i>	<u>ne möre</u>
Pok.				
FPLsm	24.	Nsg.	<i>lice</i>	<u>lice</u>
Pok.				
FPSm	25.	Nsg.	<i>čelo</i>	<u>čèlo</u>
Pok.				
tl	26.	Nsg.	<i>sljepoočnica, sljepo oko</i>	<u>sljipóčnica</u>

Fig. 2. The sample page of dialectological paper questionnaire (question asked the informant, linguistical level that is monitored, description of a word word, word told by informant)

The research, based on the Croatian Language Atlas method, is a basis also for the study of the Makarska littoral dialects. The situation is recorded with a questionnaire containing 2122 lexical data divided into 15 chapters covering different areas related to man, his environment and occupation:

- man and body parts
- family
- folk costumes (male and female)
- house and objects in the house
- diet and preparing meals
- grain farming and processing
- domestic animals
- birds and domestic poultry
- wildlife
- trees and fruits
- vegetables and flowers
- diligence
- land, water, natural phenomena
- social life
- supplement.

The supplement examines onomastic material, some adjective expressions that do not correspond to any of the previous chapters (e.g., sweet, thirsty, solid...), names for colours, pronouns, cardinal and ordinal numbers, adverbs (e.g., up, down, over, from, where, today...), verbs to be and to have in certain forms.

The words are also organized in headwords, that is, several different categories are required for a single flexional word (for flexional words forms in different cases, numbers, and for verbs different person, time, number and mood). A grammatical level is recorded for each entry, that is, showing why a particular word was even chosen to enter the questionnaire. Each headword is accompanied by a question that helps the examiner to obtain the required word (e.g., if the word head is requested, the examiner asks a question: what is the name for (with the abbreviation “Ksz”), and then points (with the abbreviation “Pok”) to their head).

It is also possible that the examiner begins the sentence and does not finish it by asking the examinee to complete it logically. The missing word in the examiner’s sentence is precisely the one to be entered in the questionnaire (e.g., It’s dark in the room; nothing can be... and the examinee is expected to say the word seen in the form present in the idiom). Naturally, such way presents obstacles because it is not always easy to obtain the desired response, the long-term study creates fatigue in the examinee and examiner, and if the examiner is not the native speaker of the examinee’s idiom, or if he speaks at the standard, the examinee will spontaneously start to approach them in their idiom. This can provide compromised data. That is why the strategy of an unstructured conversation is often used, i.e., a spontaneous speech of the examinee, who is asked to talk about their youth, activities, customs, to tell a story, anecdote... is recorded. Such way provides much more material than previously planned. Also, a material which serves as an intangible heritage is obtained because it contains valuable information about life in certain parts of the country.

4.1.2 Sociolinguistic and demographic data

The questionnaire contains detailed metadata that help classify the speech and describe the examinee and the situation in which the study was carried out, for example to which dialect the speech belongs, where the survey was conducted. In particular, the features of the site are recorded in detail:

- approximate number of inhabitants
- name of neighbouring places
- which city serves as a direct centre of the surroundings (economic, administrative, etc.)
- where children go to school, in which companies the adults work
- religion of the population
- what is the occupation of most inhabitants and whether it has always been the case

- inhabitant name for the place and neighbouring places
- spread of surnames over the place (including family nicknames)
- the most common male and female names
- names of streams, hills, rivers, springs...
- is there a proverb, sentence or gesture with which the inhabitants of neighbouring places mock each other
- what are the places around that have the same dialect
- what are the places around that have a different dialect
- origin of the population in the town
- to which animals do people give names
- enumerate plants
- phonological description of the place.

The following part records important details about the examinee, the personal ones and those by which it can be seen how their speech retained the features of the old expression or was changed under the influence of education, media, and migration:

- name and surname
- family nickname
- year of birth
- place of birth
- father and mother's place of birth
- number of ancestors born in the informant's place of birth
- nationality and religion
- the place which the spouse is from and how long they've been married
- where the informant lived in the past
- are they literate and to what extent
- data on education
- what is the degree of their intelligence
- what are their interests (past and present)
- how do they behave in relation to the conducted study (whether they show initiative or are they restrained, fearful, etc.)
- how much are they influenced by the standard Croatian language
- do they make special efforts to speak with the organic idiom (without the influence of the standard language) or are they trying to speak "lordly"
- what are their speech organs and articulation like.

4.2 Organisation of material in the computer corpus

The structure of the corpus consists of two separate parts, data collected from the questionnaire for the Croatian Language Atlas (HJA) and old texts (Fig. 3). The two are organised in two separate databases and are subjected to separate operations.

The HJA database is organized in such a way that the user (investigator) is connected to the examinee, and through the examinee the connection continues to the place where the speech is investigated. On the other hand, the user is related to the answers, i.e., any data entered in the questionnaire. Separate data is a part of the grammatical structure (division) concerning one word (all grammatical forms of flecional word). The division is related to one question, that is, a sentence by which the examiner receives information about the division, that is, about the initial grammatical category from which the entire grammatical structure of a headword is developed.

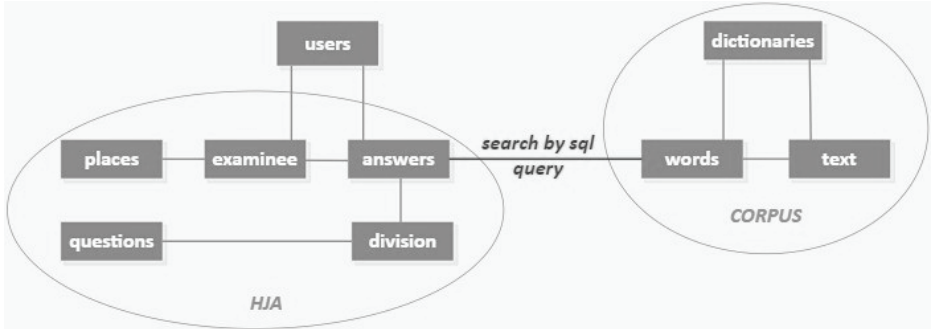


Fig. 3. The organisational scheme of computer corpus

The second database contains a corpus of old texts. This database contains complete texts and also all the words separated from them. Tokens were obtained by tokenization to the level of words, with punctuation marks neglected and not entered into the database as tokens.

Separate operations can be conducted on each database; content input and content analysis. However, to achieve a diachronic aspect of the corpus and enable diachronic analysis, it was necessary to enable communication between the two databases in order to compare historical and contemporary language data. Since these are separate databases, they are interconnected by sending SQL queries from one database to another and vice versa. The process of sending SQL queries for diachronic analysis is explained at the end of chapter 4.2.1 when comparing questionnaire data and old text corpus data.

4.2.1 The material from the questionnaire for the Croatian Language Atlas

The material collected through the questionnaire is transferred to the computer corpus on a website that has a data entry form with the same structure as the paper version of the questionnaire. Before entering the main part of the questionnaire – dialectological material or words – data on the place where the speech is examined and data on the informant are entered.

In the questionnaire, the phonetic characteristics of the phonemes are not recorded, but only the accent characteristics of vowels and syllabic r: quantity (long, short) and tone (falling, rising). Therefore, accentuation in ikavian neoštokavian idioms implies four accents with distribution as in the standard Croatian language, but with some exceptions [7]. This means that there are four different accents: short-falling –̀, short-rising –́, long-falling –^ and long-rising –´. Some vowels that are not stressed but are long are marked with length –̄. When entering each word, it is possible to obtain a drop-down box in which all special characters are offered to enter the correct accent or length (Fig. 4).

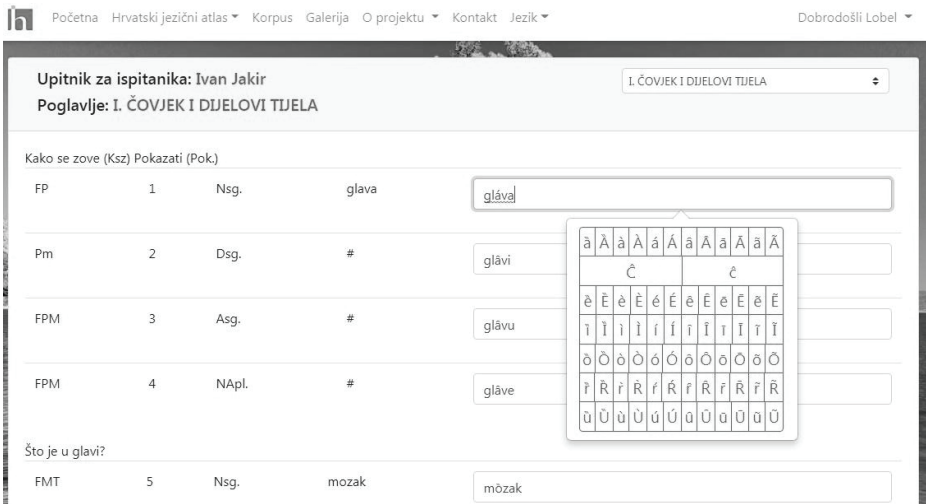


Fig. 4. Inserting the data with accent entry possibilities

The database was created in SQL and therefore data entry and review have a slight flaw. Some phonemes are specially written; ĉ, ž, ž, í, n̄. However, SQL does not support encoding special characters corresponding to their phonological transcription. Therefore, these signs are entered as ĉ → ĉ, ž → dž, ž → đ, í → lj, n̄ → nj. They reflect the dialectological notation tradition from the transition of the 19th to the 20th century, although the versions on the right side of the arrow are increasingly used today [9].

Entering data from the questionnaire provides a table in which each entity (the entire table line) is described by four attributes or features (table columns). The attributes of the table are the grammatical level for which the requested word was examined, the grammatical determination of the requested word, the word in its standard Croatian language form and number of how many dialectological results were entered for each standard language word (Fig. 5).

Početna Hrvatski jezični atlas Korpus Galerija O projektu Kontakt Jezik Dobrodošli Lobel

Pretražite Hrvatski jezični atlas

Prikaži 25 rezultata po stranici Pretraži:

Razine	Gramatika	Standard	Broj rezultata
Kako se zove (Ksz) Pokazati (Pok.)			
FP	Nsg.	glava	2
Pm	Dsg.	#	2
FPM	Asg.	#	2
FPM	NApl.	#	2
Što je u glavi?			
FMT	Nsg.	mozak	2
fm	Gsg.	#	2
m	Lsg.	#	2
Kad nam tko nešto jasno tumači, mi ga dobro...			
fpm	1. pl. praes.	razumijemo	2

Fig. 5. Digitized dialectological questionnaire (columns: linguistic levels, grammar, standard form, number of results)

Clicking on an entity opens an overview showing separately all occurrences in all surveyed local dialects. The review is organised by providing metadata, i.e., a description of the entity, in the upper part. The chapter in which the word is included shall be indicated, the question assisting the examiner to obtain an answer, the grammatical level, the grammatical description and the standard version of the word. Below that there is a table that shows in which places the word appears, in which form and what its frequency is (Fig. 6). In addition to these data, one can find additional options in the last column of the table. By clicking on the note icon one can get an audio clip and hear how the native speaker pronounces the word, and by clicking on the book icon there is an insight into which old texts contain that exact word and what its frequency is. An overview of the content from the questionnaire is a synchronic overview of the local dialects, and a comparison with the old texts (book icon) is a diachronic overview (Fig. 7).

4.2.2 Old texts in corpus

The second part of the computer corpus consists of texts created from 1729 to 1803. They were selected as representative examples of the language of the surveyed area. The first published editions were put into the optical character recognition system (OCR). They were then tokenized – the text is divided into occurrences (Fig. 8).

Početna Hrvatski jezični atlas Korpus Galerija O projektu Kontakt Jezik Dobrodošli Lobel

Rezultati pretraživanja

Odlomak: I. ČOVJEK I DIJELOVI TIJELA
Pododlomak:
Pitanje: Kako se zove (Ksz) Pokazati (Pok.)
Razina: FP
Gramatika: Nsg.
Standard: glava

Prikaži 25 rezultata po stranici Pretraži:





Mjesto	Odgovor	Učestalost	Dodatno
Baška Voda	gláva	1	 
Brela	gláva	1	 
Mjesto	Odgovor	Učestalost	Dodatno





Fig. 6. Description of each entity and its variants in local dialects (columns: place, answer, frequency)

Početna Hrvatski jezični atlas Korpus Galerija O projektu Kontakt Jezik Dobrodošli Lobel

Rezultati pretraživanja

Odlomak: I. ČOVJEK I DIJELOVI TIJELA
Pododlomak:
Pitanje: Kad tko priča nešto što
Razina: PSm
Gramatika: 1. pl. praes.
Standard: znamo

Prikaži 25 rezultata po stranici Pretraži:

Mjesto	Odgovor	Učestalost	Dodatno
Baška Voda	znâdemo	1	 
Brela	znâmo	1	 
Mjesto	Odgovor	Učestalost	Dodatno

Rezultati pretraživanja korpusa za riječ: znamo

- Benetović - Hvarčinja 1
- Kadcic A. - Bogoslovje diloredno 2

Zatvori

Fig. 7. Finding the word in old literary texts (for each word there is a number of occurrences in each of them)

That process enabled comparison of the text database with the questionnaire database at the level of words. This gives a diachronic insight into the similarities and differences between the historical and present language forms, that is, the development of words at grammatical levels can be historically monitored. Each text in the corpus can be viewed in its entirety, in the form of text rather than tokenized units. In this case, the text is divided into pages, as it is divided into pages in the original.

Br.	Riječ	Broj pojavaka	Br.	Riječ	Broj pojavaka
1	k	3	2	ka ^[1]	3
3	kacaj	1	4	kad	3
5	kada	3	6	kadcich	1
7	kadgod	1	8	kadgodir	1
9	kadijom	2	10	kadimo	1
11	kadčiča	1	12	kaifásu	1
13	kajat	2	14	kajati	1
15	kajaše	1	16	kaje	1
17	kajem	2	18	kaješ	1
19	kajo	1	20	kaju	2

Fig. 8. The occurrences in old text corpora (columns: number, word, number of occurrences)

It is also possible to perform logical operations over different sets of texts. The system offers to form two sets of texts. Following the formation of the two sets, possible operations are provided: view set A, view set B, A–B, B–A, UNION(A,B), INTERSECTION(A,B), XOR(A,B).

5 CONCLUSION

The synchronic and diachronic computer corpus of Makarska littoral dialects is a corpus accessible online, divided into two larger data groups. One group consists of data collected from the dialectological questionnaires for the Croatian Linguistic Atlas – a list of points where the dialectological situation is examined. Based on it, the phonological descriptions of Croatian dialects are made, which constitutes one synchronic level. These descriptions are used to create dialectological maps and to monitor the distribution status of certain language phenomena. The second group of data is a corpus of linguistically representative old texts from the Makarska area, which serve as a basis for comparison with the current language situation. The texts present new synchronic level of their time each, but together they make a diachronic cross-section of literary texts. Finally, old texts and materials from the questionnaires together can be compared by applying principles of data research and comparison to investigate synchronic and diachronic differences and similarities. This computer system is the first example of the computer processing of dialectological questionnaires in Croatia. Although it was drafted within the specific project which

explores the dialects of a smaller area and a single dialect, the principles of computer processing and subsequent research are applicable for all Croatian dialects and the processing of all questionnaires filled in so far and for all questionnaires to be done. Indeed, in the process of collecting materials it is now possible to skip the step of manual paper questionnaire filling out but the materials from the field can be immediately entered into the computer system which can later generate paper-based edited questionnaires. It is also possible to extend the text corpus, which may cover different periods and thus provide insight into the language phenomena of certain times and places. Further upgrading of the system will enable morphosyntactic processing of words.

References

- [1] Štokavski dijalekat. In Enciklopedija Jugoslavije. (1960). Leksikografski zavod FNRJ. Zagreb.
- [2] Lisac, J. (1998). Štokavski i torlački idiomi Hrvata. In Najwse dzieje języków słowiańskich. Editor Lončarić, M. Opole (Poland). Uniwersytet Opolski – Instytut Filologii Polskiej.
- [3] Brozović, D. (1970). O Makarskom primorju kao jednom od središta jezično-historijske i dijalekatske konvergencije. In Makarski zbornik 1, pages 381–405.
- [4] Lisac, J. (2003). Hrvatska dijalektologija 1. Hrvatski dijalekti i govori štokavskog narječja i hrvatski govori torlačkog narječja, pages 50–71, Zagreb, Hrvatska.
- [5] Govori Makarskoga primorja – sinkronija i dijakronija. Accessible at: <http://elte2.fsb.hr:8000/gomapri>, (Cited 20. 3. 2021).
- [6] Edited by map in Hrvatska enciklopedija. (2002). Vol. 4, Fr–Ht, page 676. Zagreb. Leksikografski zavod Miroslav Krleža; graphic design by Tomislav Kaniški.
- [7] Croatica. Povijest hrvatskoga jezika. (2018). Vol. 5, page 540. Zagreb.
- [8] Kurtović Budja, I. (2020). Upitnici za Hrvatski jezični atlas i Zlatna formula hrvatskoga jezika ča-kaj-što. In Conference abstracts of Computational Linguistics and Golden Formula of Croatian Language. Osijek.
- [9] Brozović, D. (1981). O fonetskoj transkripciji. In Fonološki opisi srpskohrvatskih/hrvatskosrpskih, slovenačkih i makedonskih govora obuhvaćenih opšteslavenskim lingvističkim atlasom, pages 17–25.