Ivana Lovrić Jović and Martina Kramarić

THE DUBROVNIK IDIOM THROUGH TIME
Crafting a Diachronic Dictionary

Abstract This article discusses the project, Dictionary of the Dubrovnik Idiom, conducted at the Institute for the Croatian Language. The project aims to develop a born-digital diachronic dictionary of the Dubrovnik idiom, covering the period from the 16th century to the end of the 20th century. The dictionary will be based on a historical corpus compiled within the project's scope, featuring texts from the same period. Upon completion, this dictionary will be publicly available in digital form, providing valuable insights into the linguistic evolution of the Dubrovnik region. The creation of such a dictionary will meet the needs of the scientific and cultural public as well as the citizens of Dubrovnik. Specifically, the Dubrovnik idiom is relevant because it played an important role in the standardization of the Croatian language. Additionally, well-known Renaissance and Baroque literary works are composed in it, and today it is in decline, which concerns its speakers.

Keywords born-digital dictionary; diachronic dictionary; historical corpus; Dubrovnik idiom; historical lexicography

1. Introduction

The creation of a diachronic dictionary that presents a language before its standardization poses special challenges to lexicographers, who need to rely solely on written sources or previous scientific descriptions of that language. If the dictionary presents a vernacular language,[1] the challenges are even greater, requiring a methodology that combines historical linguistics and dialectology approaches.
In the newly started project, "Dictionary of the Dubrovnik Idiom", we are facing various challenges, which we address in this article

2. Background and the Reasons for the Creation of a Dictionary of Dubrovnik Idiom

2.1 The Value of Dubrovnik Idiom in the Croatian Philology

In the standardization of the Croatian language, the Dubrovnik literacy of the 16th and 17th centuries, based on the Dubrovnik idiom, contributed significantly to establishing the Štokavian dialect as the foundation for the Croatian Standard language. This fact is well known in the history of the Croatian language, and the language of Dubrovnik literacy was analyzed and described quite early (Rešetar 1933, 1941). The lexical

---

[1] In this article, we are using vernacular (colloquial) language as the opposite term to standard language; i. e., the language that is spoken among locals and is not standardized. Furthermore, this dichotomy is used to distinguish between the literary high-register language, referred to as the Dubrovnik language, and the folk colloquial language, referred to as the Dubrovnik vernacular. This dichotomy was introduced by Lovrić Jović (2006; 2014, pp. 6–14).

component of Dubrovnik literacy is well represented in the Academic Dictionary of the Croatian or Serbian Language (ARJ 1880–1976), an extensive historical dictionary of the (not only) Croatian language from the 12th to the 20th century, written in 23 volumes. Due to its extensiveness, it is challenging for the general public to access and find specific lexemes that belong exclusively to Dubrovnik literacy.

Some researchers have also included descriptions of the local vernacular of the 19th century as part of their dialectology studies of the Croatian language (Budmani, 1883). Further analysis of the local vernacular of the 18th century was conducted by Lovrić Jović (2014), who examined its traces in the literary language of *Francesarie* – Molière's comedies adapted into the Dubrovnik idiom. This analysis reveals significant features of the local vernacular of that time. Similar conclusions can be drawn for the administrative language of testimonies, which were transcribed and analyzed by the same author (Lovrić Jović, 2006, pp. 173–192; 2015).[2]

2.2 The History of the Dubrovnik Idiom

The local vernacular of Dubrovnik originally belonged to the Štokavian dialect group, which is part of the South Slavic branch of the Slavic language family. Starting in the 16th century, it began to evolve and incorporate features of the newer Štokavian dialect (Vulić, 2011, pp. 180–182). This process was completed in the 19th century, so today's Dubrovnik speech is undoubtedly New Štokavian with Ijekavian pronunciation.[3] According to Brozović (2005, p. 20), the process of "new Štokavianization" in the 17th and 18th centuries was performed inconsistently, resulting in special, authentic features of the idiom, particularly in terms of prosody.

Throughout its history, the Dubrovnik idiom was heavily influenced by the Italian language. This influence affected the local vernacular, resulting in the incorporation of many Italianisms. Consequently, a significant number of these Italianisms are expected to be part of the word list in the Dictionary of the Dubrovnik Idiom.[4]

As for the Italian dialect that influenced the Dubrovnik idiom, according to Muljačić (1962, p. 341), the Venetian language initially had the greatest influence, but was later supplanted by Tuscan, which the Dubrovnik residents learned from books or in Italy.[5] The local citizens were less familiar with Tuscan. The use of Venetian declined among merchants and sailors, while Italian, supplanted Venetian as the cultural language learned at Italian universities and by merchants.

---

[2] In the project The Dictionary of Dubrovnik Idiom and in the process of its creation of the corpus and dictionary, we will emphasize the difference between high and low register literacy. The special attention will be given to the low register literacy (to be explained) because it represents the local Dubrovnik vernacular in its best manner, and its lexis, albeit stylized, offers the possibility of presentation the local vernacular in the diachronic perspective from the 16th till 20th century.

[3] This distinguishes it from other Štokavian dialects like the Ekavian and Ikavian varieties.

[4] This complex sociolinguistic situation was described throughout the centuries by Lovrić Jović, 2006, pp. 173–175. Italianisms in the Croatian language were analyzed by Muljačić, 1973; Sočanac, 2004; and Lovrić Jović, 2006.

[5] That language exerted no direct influence, it was acquired from books or introduced by the considerable number of Dubrovnik students who pursued studies in Italy.

2.3 Dubrovnik Idiom in Dictionaries

Despite the significant role that the Dubrovnik idiom played in the standardization of the Croatian language, there is still no comprehensive dictionary of the Dubrovnik idiom based on a corpus and adhering to rigorous lexicographical principles. Instead, recent decades have seen the creation of glossaries by speakers of the Dubrovnik vernacular. These glossaries lack in quality and expertise. However, there does exist a scientific foundation-based, non-corpus-based Dubrovnik idiom dictionary for the 20th century (Bojanić & Trivunac, 2002). For the older linguistic layer, there is the dictionary by Jakov Mikalja (1649/51, reprint Mikalja, 2011), which is based on the Dubrovnik idiom and is considered a precursor to the modern type of dictionary. To this, we can add the annex dictionaries that belong to the monographs of the Dubrovnik idiom (Lovrić Jović, 2014, 2015; Lovrić Jović & Jozić, 2016).

3. The State of the Croatian e-Lexicography

The 'poor' state of Croatian e-lexicography was thoroughly described by Despot & Möhrs (2015, pp. 334–335; pp. 345–346). The situation underwent significant enhancement with the implementation of the project "Croatian Web Dictionary *Mrežnik*" (from 2017) and the initiation of the development of the first corpus-based, born-digital contemporary monolingual e-dictionary of contemporary Standard Croatian, *Croatian Web Dictionary – Mrežnik*.[6] The dictionary is currently in the process of creation, with only a demo version available, which covers entries to the letter F (https://rjecnik.hr/mreznik/). Comprising three modules, the dictionary heavily emphasizes the usage perspective of language, particularly focusing on collocations. Hence, the absence of a historical and dialectal e-dictionary in Croatian philology persists. Our project endeavors to address and fill this void.

4. The Project of a Dictionary of Dubrovnik Idiom

Recognizing the scientific and general need for an integrated dictionary, a project of compiling a dictionary of the Dubrovnik idiom has commenced at the Institute for the Croatian Language (from 2024).[7] This project aims to create a corpus-based born-digital diachronic dictionary of the Dubrovnik idiom covering the period from the 16th till the end of the 20th century.[8] As this dictionary is intended to fill a significant gap, it should serve several user groups. First of all, it would be aimed at language historians and dialectologists, with the possibility of extending it to theatre scholars,

---

[6] The project is being conducted at the Institute for the Croatian Language and is led by Dr. Lana Hudeček, who is a as well member of the Dictionary of Dubrovnik Idiom project. Dr. Hudeček's extensive lexicographical knowledge and expertise will be inherited by the team involved in this project.

[7] Project is led by Dr. Ivana Lovrić Jović.

[8] A similar concept to the dictionary project can be found in the printed dictionary by Bencistà, Alessandro (2021). *Vocabolario del vernacolo fiorentino toscano (con gli esempi delle voci da Dante ai moderni)*. This dictionary presents the Italian Florentine dialect from the 14th century, which was selected as the foundation for Standard Italian language owing to Dante's literary works. In this dictionary, examples from both the vernacular and literary language are integrated, and a diachronic approach is employed. The dictionary contains a total of 400 entries.

dramaturges, and actors. Another important target group would be the citizens of Dubrovnik, especially those who are concerned about the possible disappearance of the Dubrovnik vernacular.[9]

4.1 Terminological Clarifications Regarding Dubrovnik Idiom

We are using "idiom" as the superordinate term, which incorporates both the local Dubrovnik vernacular, both in the form of the local speech and in the form of the stylized literary language presented mainly in low-register literature (comedies, congratulatory folk songs, satirical periodicals, etc.), and the literary language written primarily in high-register literature. Since the vernacular, as an authentic living expression, is impossible to explore for past times, information about it can only be obtained indirectly from those parts of written texts where the Dubrovnik vernacular is best reflected (dramatic works, congratulatory poems, transcribed handwritten testimonies)[10] and in genres that effectively convey spoken language, for example, in epistolary prose. For more recent periods, the lexicon will be excerpted from audio recordings of speakers of the Dubrovnik vernacular, among the most valuable of whom are emigrants to other continents who best preserve the speech of the second half of the 20th century.[11]

5. Material for the Corpus

The Dictionary of the Dubrovnik Idiom will be based on a historical corpus of texts from the same period as the dictionary. It will be compiled within the scope of the project and built from scratch. In selecting material for inclusion in the corpus, emphasis is placed on the diversity of genres, functional styles, and temporal coverage.

The corpus is compiled from various sources, including different historical texts that represent various literary and functional styles of the language. It will contain existing philological descriptions of the Dubrovnik idiom: grammar, glossaries, dictionaries. Further it will include administrative-legal texts such as testimonies from the 17th and 18th centuries (Lovrić Jović, 2015), archival documents like *Lamenta Ragusina* and *Lamenta del Criminale* from the 17th and 18th centuries, personal letters and epistolary prose from the 19th century, folk congratulatory songs, literal and dramatic texts such as Držić's comedies and *Francesarie* – adaptations of Molière's comedies into the Dubrovnik idiom written in the 18th century (22 comedies, Lovrić Jović, 2014), memoir texts of prominent citizens, published chronicles, etc. Further, we will incorporate sources from a special section of Dubrovnik press from the 20th century (including satirical newspapers published from 1901 until the end of the 20th century), and a cookbook representing Dubrovnik dishes throughout history and written in the Dubrovnik vernacular (Ničetić, 2012).

---

[9] The phonological, morphological, syntactical, and prosodic features that rendered it authentic are nearly extinct.

[10] Deeper theoretical explanation of this terminology see in Lovrić Jović, 2006, pp. 173–175; 2014, pp. 6-9.

[11] Audio recordings were recorded ten years ago among local inhabitants of the city of Dubrovnik by Ivana Lovrić Jović.

In addition, we will use as a reference source the previously mentioned Dictionary of Mikalja and the ARJ, in which the canonical high literature was processed and lexical material of prominent Dubrovnik writers through the centuries was mentioned (Đore Držić, Šiško Menčetić, Mavro Vetranović, Marin Držić, Ivan Gundulić, etc.), along with early testimonies of literary works such as prayer books, lectionaries and miscellanies. As an additional source for the dictionary, we will include audio recordings of speakers from Dubrovnik.

Once the corpus is completed, it will be representative in terms of covering different time periods and functional styles of language. Both stylized vernacular, colloquial low-register vocabulary from comedies, congratulatory songs, satirical periodicals, etc., as well as literal high-register vocabulary, will be presented. The representation of various sources from different time periods will enable the analysis of lexicon in a diachronic overview and facilitate the recognition of changes that have occurred over time in the lexicon (disappearance, expansion, or narrowing of meanings, changes in forms, etc.).

Currently, the project is in the phase of creating the corpus using the Sketch Engine, which involves the digitization and critical interpretation of written sources from various centuries. The corpus is derived from printed documents, manuscripts, and TIFF image files. All the documents will be digitized for the first time within the scope of the project. In addition to digitized sources, handwritten manuscripts such as personal letters from the 19th century will not be transcribed fully, but they will be consulted as sources for examples.[12] The same approach will be applied to audio recordings, which will not be fully transcribed. Instead, only examples will be excerpted. All excerpted material will be adapted according to contemporary transcription rules, and all data in the dictionary will be provided in the transcribed form.[13] For consistency, some words from printed material will be adapted according to the transcription principles adopted for this dictionary. This is considering the fact that printed texts may contain forms written according to different transcription principles. Once the corpus is finished, it will be searchable, upgradable, and partly publicly available, with exceptions for works that are still restricted by legal rights.[14]

---

[12] For sources written in manuscripts during a time when there was no Standard Croatian orthography yet, and Italian orthography was used on the coast, especially for consonants with diacritics (such as *č, ć, ž, š, dž, đ*, but also *lj* and *nj*), transcribing all the personal letters in full would be very time-consuming.

[13] The same rules were used, for instance, in the transcription of testimonies by Lovrić Jović, 2015. In addition to the transcribed texts, the corpus may also contain non-transcribed texts written in the old, non-standardized orthography. These may be digitized printed material from the 19th century or older, and later in the text we will explain how we will extract data for the dictionary from these texts. If the user of the corpus wants to search for an element in the corpus, they need to know how this element might have been written down over the centuries.

[14] Currently, the corpus is only accessible to members of the project team.
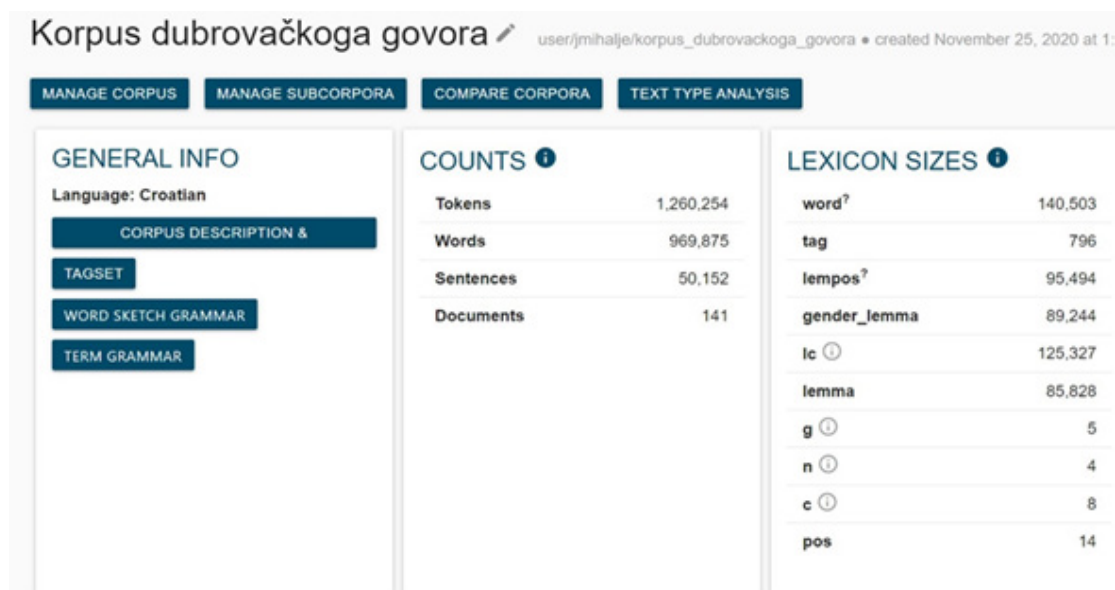
Fig. 1: Current status of the corpus

6. Methodology

In crafting the Dictionary of Dubrovnik Idiom, we will integrate methodologies from both historical linguistics and contemporary dialectology dictionaries. Once the dictionary is finished, it will provide crucial information for understanding the diachronic development of the described language. Additionally, it will offer valuable insights for dialectologists in terms of accentuation and sound recording of the headwords.

As mentioned before, the methodology of the dictionary follows the *Mrežnik* dictionary with the exception that only one module will be created. This means that a corpus-based methodology will be employed.[15] The microstructure of the dictionary entry will be similar to the mentioned dictionary but with exceptions due to historical component of the dictionary and the limitations of the corpus.[16] Therefore, only very typical collocations and phrases will be provided, in contrast to *Mrežnik*, where collocations play an important role (Hudeček – Mihaljević, 2020b, p. 78). T-Lex was chosen as the DWS for both projects. The microstructure of the dictionary entry is shown in the diagram below.[17]

---

[15] In the *Mrežnik* project, lexicographers use a corpus, but they have the freedom to decide what should be included in the dictionary entry. All data extracted from the corpus serve only as guidelines (www.mreznik.ihjj; Hudeček – Mihaljević, 2020a, p. 657).

[16] The second major difference concerns the fact that *Mrežnik* is the normative and the Dictionary of the Dubrovnik Idiom the descriptive dictionary. About the normative aspect of *Mrežnik* see in Hudeček & Mihaljević 2020a, pp. 658–659.

[17] The microstructure of the dictionary entry was created by Ivana Lovrić Jović and Lana Hudeček.
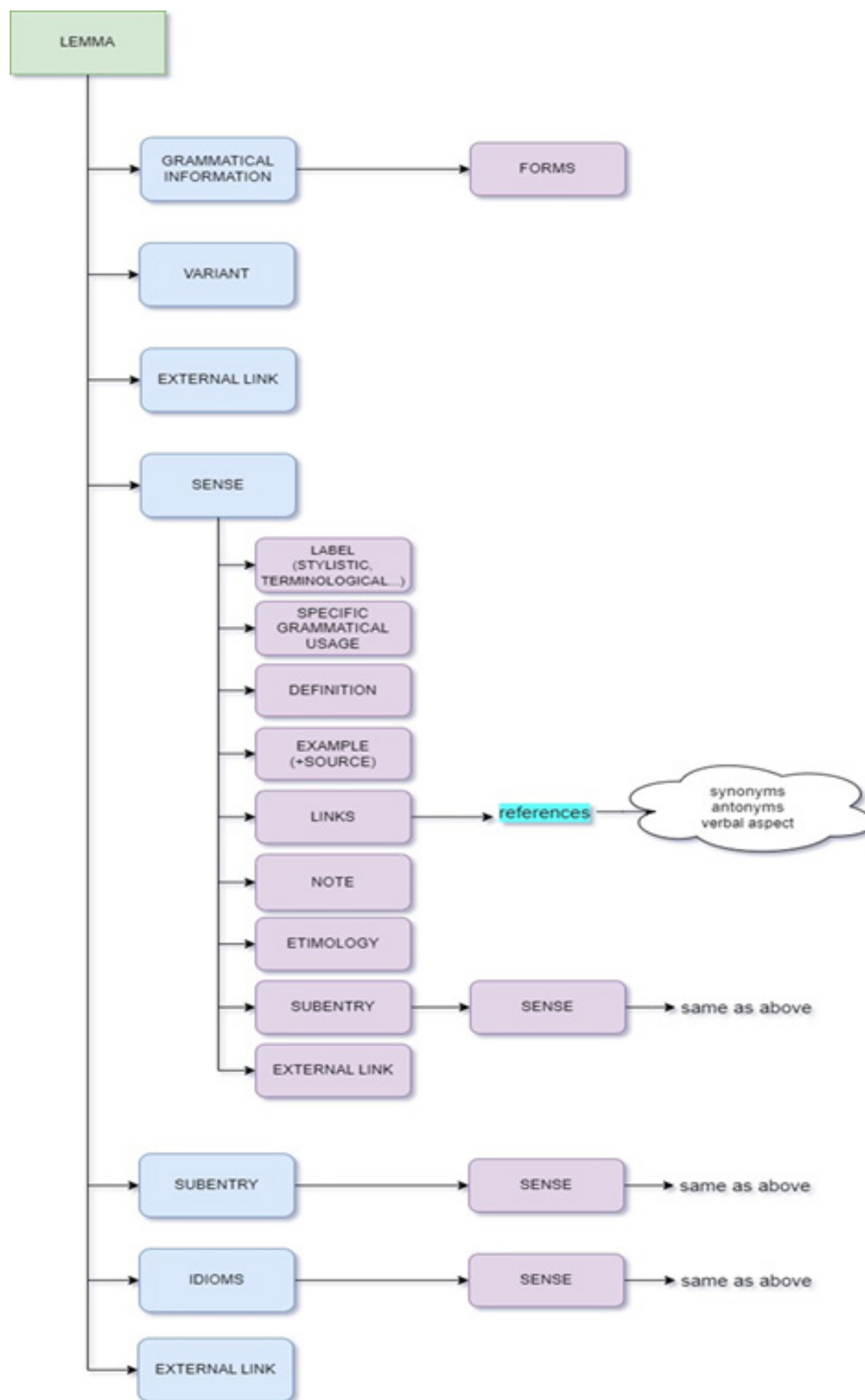
Fig. 2: The microstructure of the dictionary entry

The corpus of the Dubrovnik idiom is lemmatized, and we will use the Sketch Engine web tool, but the extraction from the corpus will be performed mainly manually.[18] As the additional tool, we will use *Word Sketches* for Croatian which were written for the *Mrežnik* project (Hudeček & Mihaljević, 2020a, p. 654).[19] Since the Dubrovnik idiom is based on the same Štokavian dialect as standard Croatian, we will be able in some extent to display lexeme context through *Word Sketches*, even if we are working with historical lexemes that sometimes have a different form than contemporary lexis.

6.1 Word List

Following the principle of effectiveness (Morkovkin, 1996), the dictionary will be distinctive in a broader dialectal sense, adhering to the concept of differentiating vocabulary where items "that are equal to the standard" are not included (in the very broad sense of the word). This means that, since this isn't a standard language dictionary, only those lexemes that are unfamiliar to other speakers of the Croatian language and that constitute the general lexicon of Dubrovnik residents (not specialized) will be excerpted. Words that differ only in their phonological and morphological form from their standard equivalents will not be processed in this version. Professional vocabulary will not be considered, or rather, it will be considered only to the extent that it is usually represented in standard language dictionaries. Lexemes belonging to the standard language (of which there are many in every idiom, e.g., *water, head, house, child, sky, live, know, wait*, etc.) will not be included.[20]

Initially containing around 730 entries, the dictionary will be open to further expansion. Since the project is currently in the corpus creation phase and the lexicographic processing of headwords has not yet begun, we will present each component of the envisioned dictionary entry and outline the challenges expected in the processing considering the limited corpus and lack of all confirmations, as well as the variability of forms, which is an important characteristic of historical and non-standard language.

---

[18] In view of the fact that there is not yet a tool that can cope with the non-standardised spelling in the corpus, we have to carry out the search manually. Since the members of the project team are philologists and historical linguists, they all know the possible spellings of a lemma in the old orthography. For example, *mobilja* 'furniture' is written as *mobiglia*, *crkva* 'church' as *zarqua*, *poći* 'to go' as *pochi*. The old ortography differs from the standard orthography only in certain letters and is well known to the project members. Once the data written in non-standard orthography is extracted from the corpus, the headwords and examples of usage will be transcribed for the dictionary entry.

[19] All corrections in the forms of headwords will be performed manually. The display of lexeme context through *Word Sketches* enables an overview of the most common collocations sorted into syntactic categories, facilitating the discovery of good examples of word usage or collocations (Hudeček & Mihaljević, 2020a, p. 654).

[20] Furthermore, the decision on word lists for the dictionary cannot be based solely on the corpus, frequency of occurrence of an element, or collocations, but rather on experience from which it is possible to conclude whether a lexeme is sufficiently confirmed in the Dubrovnik idiom to deserve a place in the dictionary. Namely, there are lexemes appearing in a part of the corpus that will not enter the dictionary because they are confirmed only once, unknown to speakers, and based on the genre of the text, it can be concluded that they are failed neologisms. The majority of them come from various folk texts, from congratulatory poems to satirical periodicals, but also in Dubrovnik adaptations of Molière's comedies.

7. Dictionary Entry

7.1 Headword: Form, Accentuation, Variants

The headword will be accentuated and given in the canonical form (e.g., for nouns: nominative singular, for verbs: infinitive form), and in the form transcribed according to contemporary transcription rules.

Headwords that are not neutral will be labeled with stylistic and terminological labels.

In non-standard dictionaries, it is anticipated to encounter phonological variants of the headwords (e.g., *grakàtež/grahatež/grakatež* for 'grater'), which in this dictionary will all be listed. These variations offer us a crucial piece of information about the development of the language, which is an important part of the diachronic perspective of this dictionary. For example, in the word *pacijencija/pacijenca* ('patience'), phonological variations exist, and the first version is older. Similarly, for the word *gȍspoja/gospòđa*, phonological variations exist, and the first form is older and belongs to literary language. In the verb *afiktat(i)/afitat(i)* ('to rent'), both variants are recorded in the same period, in the same testimony from 1767, but the first version (with the consonant cluster *-kt-* indicating an older (Latin) layer of loanword entry into the Dubrovnik vernacular) has completely disappeared to this day, and the second one is still in use. This information will be provided through stylistic labels followed by explanatory notes as part of the dictionary entry.

If the canonical form of the word is not attested in the corpus, it will be reconstructed on the basis of the phonological and morphological features of the Dubrovnik idiom. If there are two or more possible forms for this construction, all of them will be listed. The headword will be accentuated, and a sound recording will be provided.[21] As the corpus generally does not offer information about accents, deciding whether to record them presents a significant challenge.[22] It will have to be sought in the scientific lexicographic literature.[23] Indirectly, information will be sourced from scientific phonological descriptions of the older Dubrovnik language in literature, from dramatic texts in which the accents of Dubrovnik actors born at the beginning of the 20th century, etc. In addition, the aforementioned audio recordings of Dubrovnik speakers will be used for accent reconstruction. In cases where the accent is disputed, it will be reconstructed based on living speakers. The sound recording of the headword will be captured in a music studio and pronounced by the project leader, with collaboration from a Dubrovnik actor possessing extensive experience in portraying Dubrovnik characters in dramas by Dubrovnik authors spanning from the 16th to the 20th century.

---

[21] The accentuation of the headword in the dictionary will be performed by a member of the dialectology team, with suggestions from the project leader.

[22] For example, we will provide an overview of the accent reconstruction for the headword *kondicion* ('condition'), an Italian loanword. In the Dubrovnik idiom and in the Standard Croatian language, the stress behaves differently. In the standard language, there is a shift in stress (from the original *condizióne*) by one syllable closer to the beginning of the word. In the Dubrovnik idiom, the stress from the Italian origin, which falls on a diphthong with the first element being a semivowel /j/, shifts directly to that semivowel, which becomes the bearer of stress: Italian *kondit'tsjo* – Dubrovnik *kondiciōn*. In the standard language, the stress moves one syllable further from the end of the word: *kondícija*, or according to the Neoštokavian accentual system, even further: *kòndicija*.

[23] In ARJ headword is accentuated.

7.2 Grammar Block

A grammar block will contain information about the morphological features of the word, the part of speech, gender, and usually for nouns, the form of the genitive singular case, and for verbs, the aspect information will be given.[24] For adjectives, we will provide information about the definitive and indefinite forms. Exceptions to that rule will be mentioned as specific grammatical usage if an unexpected form of some word occurs.[25]

If the grammatical information is not attested in the corpus, some forms will need to be reconstructed according to available insights from phonological and morphological descriptions – e.g., *pantarula* Gsg, Nsg *pantaruo* (not *pantarul*) 'fork'.[26]

7.3 Sense Block

A sense block will contain definitions covering all meanings and examples of usage based on the corpus. The definitions will be conveyed either through a Croatian standard synonym (e.g., *vizita – posjet* for 'visit', *kondicion – uvjet* for 'condition') or explained via paraphrase (e.g., *pižuo –* 'specific kind of stone'). Examples of usage will be drawn from the corpus based on their degree of clarity, with the explanatory note indicating from which text it was excerpted and the time period and genre/functional style. This will enable users of the dictionary to determine whether a word belongs more to colloquial, low-register language or literal, high-register language.

In the absence of examples of usage from the corpus or if they are not sufficiently explanatory, they may be found in philological works, ARJ, or from existing dictionaries (with a critical approach), or from data provided by living speakers. Typical semantic relations, like synonymy and antonymy, between words will be recorded. For synonymy relations, we will mention only synonyms belonging to the Dubrovnik idiom because we already used a synonym from the standard language for the definition. So, for the lexeme *parac* ('patron saint'), the Dubrovnik synonym will be *protektur* and the standard definition *svetac zaštitnik*. For the lexeme *pedepsa* ('punishment'), the synonym will be *kastig* and the standard definition *kazna.*

---

[24] Once more we will mention the word *kondicion*, which exhibits many specificities of the Dubrovnik idiom on several levels: phonological, morphological, and prosodic. Regarding its grammatical gender specificity compared to the standard language, its ending *-(i)on* can be used for feminine nouns in Dubrovnik, while in the standard language, they are masculine nouns (*bastion*, *peron*, *žeton*). The noun *kondicion* originates from the Italian form *condizione*, more likely the Venetian form *condizion*, and it is one of many borrowed into the Dubrovnik idiom with this ending or rarely its phonological variant *-ijun* (*okupacion*, *okazion*, *ezekucion*, *ordinacion*, *komisijun*). In the standard language, loanwords of this type adapt on a derivational level and take the ending *-ija* (*kaucija*, from German *Kaution* ← Latin *caution*; *institucija* from Latin *institutio*; *špedicija* from Italian *spedizione*). In the Dubrovnik idiom, such loanwords either drop the initial vowel *-e* (from Latin or Italian origin) or they enter the language without it (from Venetian origin).

[25] For example, regarding the word *kondicion*, in addition to the genitive form *kondicioni*, we will also discuss the instrumental form, which is found several times in testimonies from the 17th and 18th centuries, confirming the old case ending for the instrumental singular: *kondicionim*. Alongside the archaic form, it is necessary to mention the contemporary form of the instrumental singular, as it is confirmed in the same texts: *kondicioni*.

[26] The expected form of the nominative singular *pantarul* according to the genitive singular *pantarula* does not belong to the Dubrovnik idiom because the vocalization *l* to *o* appears in the nominative singular. That change is typical for the Štokavian dialect.

In addition, for the verbs, we will provide both the perfective and imperfective forms. It is common for perfective verbs to be formed from basic verbs by adding prefixes, and these prefixes can vary. Therefore, for the Latinism *kordati* ('to arrange, set in order, organize, agree, come to terms'), two versions of the perfective aspect will be listed: *uskordati* and *akordati*.

When the corpus allows, we will mention collocations or typical phrases.

7.4 Etymology

For loanwords, an etymological note will be provided. In the Dubrovnik idiom, aside from Italianisms, these include Latinisms, Hellenisms, Turkisms, and occasionally rare Germanisms, Hungarianisms, and Hispanicisms. Italianisms will sometimes be specified as Venetianisms if the information is certain (if it's probable but uncertain, that will also be noted).

The note section will be used for various types of notes, including different variants of lexemes for different grammar forms, examples of usage, etymology, etc.

7.5 The Sub-Entry

The sub-entry will have the same structure as the main dictionary entry.

7.6 The External Link

The external link will be used to connect the dictionary entry with external sources. For example, if a satirical newspaper is mentioned as the source, we will link it to its digitized form in the *Digital Repository of the Scientific Library of the Dubrovnik Libraries* (https://zdur.dkd.hr/?pr=l&mr%5B27848%5D=a&ps=100).

8. Conclusion

The primary objective of the Dictionary of the Dubrovnik Idiom project is to develop a diachronic, corpus-based, born-digital dictionary that encompasses the Dubrovnik idiom from the 16th to the 20th century. This paper outlines the methodology used to compile the corpus and the planned structure of the dictionary. It also describes the microstructure of the dictionary entries and discusses the anticipated challenges in creating the dictionary, such as the limitations of the available corpus and the task of reconstructing the vernacular from written sources. Additionally, it examines both the diachronic dimension of the dictionary and the dialectal aspects, particularly focusing on the accentuation and sound recordings of headwords.

For a project like this in Croatian philology, there is a significant need because a dictionary of the Dubrovnik idiom, created based on rigorous lexicographic principles, does not yet exist. Given the diachronic perspective, diversity, and

comprehensiveness of the texts that will comprise the corpus for this dictionary, as well as the pivotal role the Dubrovnik idiom has played in the standardization of the Croatian language and its importance in dialectological research, the corpus and the resulting e-dictionary will hold considerable scientific, expert, heritage, and cultural value. This value extends primarily to philologists and linguists but also to the broader Croatian public and an important cultural sector: the theatrical world, where actors often face challenges in conveying Dubrovnik texts and understanding the nuances of Dubrovnik sentences.

A special value will be brought by the sound recordings of the headwords. This element is a welcomed addition in contemporary lexicography, and in this project, it will present a significant challenge. Determining the correct accentuation will rely on systematic methods, linguistic and dialectological expertise, experience in working with the Dubrovnik idiom, and the authenticity provided by the project leader, who is a native speaker of the Dubrovnik idiom. This inclusion of sound recordings will enhance the dictionary's usability and authenticity, offering invaluable support to users in understanding and pronouncing the headwords accurately.

The geopolitical situation in the Dubrovnik area has led to demographic changes that are not favorable to the Dubrovnik idiom, so it is now heard very rarely in Dubrovnik. Therefore, this is the last opportunity to rescue its vocabulary from oblivion.

References

ARJ – *Rječnik hrvatskoga ili srpskoga jezika* [Academic Dictionary of the Croatian or Serbian language] (1880–1976) I – XXIII. JAZU.

Bencistà, A. (2021). *Vocabolario del vernacolo fiorentino toscano (con gli esempi delle voci da Dante ai moderni).* Sarnus.

Bojanić, M., & Trivunac, R. (2002). *Rječnik dubrovačkog govora.* Srpski dijalektološki zbornik XLIX. SANU.

Brozović, D. (2005). Dva moguća fonološka tumačenja prozodijskoga sustava u cavtatskome govoru. Translation from English: H. Heffer. *Croatica et Slavica Iadertina, 1*(1), 19–30.

Budmani, P. (1883). Dubrovački dijalekat, kako se sada govori. *Rad JAZU, 65*, 154–179.

Despot, K. Š., & Möhrs, C. (2015). Pogled u e-leksikografiju. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje, 41*(2), 329–353.

Foji od Pazara (1901 – 1941). *Digital Repository of the Scientific Library of the Dubrovnik Libraries.* Retrieved April 3, 2024, from https://zdur.dkd.hr/?pr=l&mr%5B27848%5D=a&ps=100

Gabrić-Bagarić, D., Horvat, M., Lovrić Jović, I., & Perić Gavrančić, S. (2012). *Jakov Mikalja. Blago jezika slovinskoga (1649./1651.): transkripcija i leksikografska interpretacija.* Institut za hrvatski jezik i jezikoslovlje.

Hudeček, L., & Mihaljević, M. (2020a). The Croatian Web Dictionary – Mrežnik project – goals and achievements. *Rasprave, 46*(2), 645–667. https://doi.org/10.31724/rihjj.46.2.11

Hudeček, L., & Mihaljević, M. (2020b). Collocations in the Croatian Web Dictionary – Mrežnik. *Slovenščina* 2.0., *8*(2), 78–111. https://doi.org/10.4312/slo2.0.2020.2.78-111

Lovrić Jović, I. (2006). Fonološka adaptacija talijanizama u dubrovačkim oporukama iz 17. i 18. stoljeća. *Rasprave Instituta za hrvatski jezik i jezikoslovlje, 32*(1), 173–192.

Lovrić Jović, I. (2014). *O starome dubrovačkom govoru nazbilj. Jezična analiza dubrovačkih frančezarija.* Institut za hrvatski jezik i jezikoslovlje.

Lovrić Jović, I. (2015). *Ja, Krsto Lučin Dubrovčanin, činim sad ovi testamenat. Jezična analiza dubrovačkih oporuka iz 17. i 18. stoljeća s transkripcijom i aneksnim rječnikom.* Institut za hrvatski jezik i jezikoslovlje.

Lovrić Jović, I., & Jozić, Ž. (2016). *Libro od svetoga Vlaha. Jezična studija s rječnikom.* Institut za hrvatski jezik i jezikoslovlje.

Metzeltin, M. (1992). La Dalmazia e l'Istria. In F. Bruni (Ed.), *L'italiano nelle regioni: lingua nazionale e identità regionali* (pp. 316–335). UTET.

Morkovkin, V. V. (1996). The Theoretical Basis of a Systematic Approach to Dictionary Making. In: M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström, C. Röjder Papmehl. *Euralex '96 Proceedings I—II.* Retrieved March 15, 2024, from https://www.euralex.org/elx_proceedings/Euralex1996_2/005_Valery%20V.%20Morkovkin %20-The%20 Theoretical%20Basis%20of%20a%20Systematic%20Approach%20to.pdf

Muljačič, Ž. (1962). Dalmatski elementi u mletački pisanim dubrovačkim dokumentima 14. stoljeća. Prilog raguzejskoj dijakronoj fonologiji i dalmatsko-mletačkoj konvergenciji. *Rad JAZU, 327*, 237–380.

Muljačić, Ž. (1973). Problemi della simbiosi slavo-romanza nell'Adriatico. In M. Cortelazzo (Ed.), *Miscellanea II* (pp. 21–39). Facoltà di lingue e letterature straniere dell'Università di Trieste.

Ničetić, J. (2012). *Dubrovački komini.* Self edition.

Rešetar, M. (1933). Jezik Marina Držića. *Rad JAZU, 248*, 99–240.

Rešetar, M. (1936). Jezik pjesama Ranjinina zbornika. *Rad JAZU, 255*, 77–220.

Rešetar, M. (1941). Glavne osobine Gundulićeva jezika. *Rad JAZU, 272*, 1–44.

Sočanac, L. (2004). *Hrvatsko-talijanski jezični dodiri (s rječnikom talijanizama u standardnome hrvatskom jeziku i dubrovačkoj dramskoj književnosti).* Nakladni zavod Globus.

Vulić, S. (2011). Štokavski hrvatski književni jezik. In A. Bičanić, R. Katičić, & J. Lisac (Eds.), *Povijest hrvatskoga jezika, 2. knjiga: 16. stoljeće.* (pp. 125-187). Društvo za promicanje hrvatske kulture i znanosti CROATICA.

Acknowledgments

Contact Information

Ivana Lovrić Jović
Institute for the Croatian Language
ilovric@ihjj.hr

Martina Kramarić
Institute for the Croatian Language
mkramar@ihjj.hr