Lana Hudeček
Milica Mihaljević
Institute of Croatian Language and Linguistics
Ulica Republike Austrije 16, HR-10000 Zagreb
lhudecek@ihjj.hr
mmihalj@ihjj.hr

THE *CROATIAN WEB DICTIONARY – MREŽNIK* PROJECT – GOALS AND ACHIEVEMENTS[1]

The *Croatian Web Dictionary – Mrežnik* project is at the moment of writing of this paper at the end of the third year. The project goals are to compile an e-dictionary of the Croatian language that is online, free, corpus-based, monolingual, hypertext, searchable, normative, and based on the contemporary results of e-lexicography and computational linguistics. *Mrežnik* consists of three modules: for adult native speakers of Croatian, schoolchildren, and non-native speakers of Croatian. It will be the central meeting point of the existing language resources of the Institute of Croatian Language and Linguistics but also of the language resources created within the project. *Croatian Web Dictionary – Mrežnik* is conceived as a dynamic dictionary that can be further compiled and edited even after the end of the four-year project and the result of the project will be the model for the growing Croatian e-dictionary that will be a long-term project of the Institute of Croatian Language and Linguistics.

1. Introduction

Croatian Web Dictionary – Mrežnik is a four-year project financed by the Croatian Science Foundation and conducted at the Institute of Croatian Language and Linguistics in Zagreb.

The project began on 1ˢᵗ March 2017 and will end on 28ᵗʰ February 2021. After the end of the project, the dictionary will be a long-term project of the Institute of Croatian Language and Linguistics.

The reason for launching the *Mrežnik* project was primarily the lexicographic experience that most of the members of the *Mrežnik* team gained while compiling *Školski rječnik hrvatskoga jezika* (*School Dictionary of the Croatian Language*, Hudeček and Mihaljević 2012) with 30 000 dictionary entries (now also available online at the address rjecnik.hr) and *Prvi školski rječnik hrvatskoga jezika* (*First School Dictionary of the Croatian Language*, Čilaš Šimpraga, Jojić, and Lewis 2008) with 2000 dictionary entries for schoolchildren aged 7 to 9.[2] At the end of these projects, an experienced lexicographic team was formed with a clear idea that compiling a modern e-dictionary of the Croatian language demands a different approach, including new technologies and new knowledge. An additional motive to start work on this project was also the fact that Croatia is still one of the countries which do not have a publically available online dictionary of their national language compiled according to the rules of contemporary e-lexicography, i.e. a born-digital dictionary.[3]

The need for extensive scientific research in the area of e-lexicography, i.e. getting to know the theory and practice of creating e-dictionaries and the possibilities that new dictionary platforms offer, was also recognized. However, as *Mrežnik* is a scientific project; it is compiled taking into account semantic relations and the systematic nature of language. The systematic nature of the dictionary can be seen in these areas: accentuation of entry-words, the selection and accentuation of forms provided in the grammatical block, the definition of words that belong to closed grammatical and semantic groups, etc. For this

---

[2] Encouragement for starting the *Mrežnik* project also came from our colleague Kristina Štrkalj Despot, the member of COST action *European Network of e-Lexicography*, *ENeL*.

[3] More on Croatian online dictionaries see in Hudeček and Mihaljević 2017c, Štrkalj Despot and Möhrs 2015, and Štrkalj Despot et al. 2019.

reason, lists of grammatical and semantic groups with lists of entry-words were composed as shown in Table 1.

Table 1: List of entry-words denoting number

	cardinal number	ordinal number	a group of x men	a group of x people	around x	x year old[4] xth anniversary one xth
10	deset	deseti	desetorica	desetero	desetak	desetogodišnji desetogodišnjak desetogodišnjica desetina desetica
11	jedanaest	jedanaesti				
12	dvanaest	dvanaesti				
13	trinaest	trinaesti				
14	četrnaest	četrnaesti				
15	petnaest	petnaesti				
16	šesnaest	šesnaesti				
17	sedamnaest	sedamnaesti				
18	osamnaest	osamnaesti				
19	devetnaest	devetnaesti				
20	dvadeset	dvadeseti	dvadesetorica	dvadesetoro	dvadesetak	dvadesetogodišnjak dvadesetogodišnji dvadesetogodišnjica
30	trideset	trideseti	tridesetorica	tridesetoro	tridesetak	tridesetogodišnjak tridesetogodišnji tridesetogodišnjica
40	četrdeset	četrdeseti	četrdesetorica	četrdesetoro	četrdesetak	četrdesetogodišnjak četrdesetogodišnji četrdesetogodišnjica
50	pedeset	pedeseti	pedesetorica	pedesetoro	pedesetak	pedesetogodišnjak pedesetogodišnji pedesetogodišnjica
60	šezdeset	šezdeseti	šezdesetorica	šezdesetoro	šezdesetak	šezdesetogodišnjak šezdesetogodišnji šezdesetogodišnjica
70	sedamdeset	sedamdeseti	sedamdesetorica	sedamde-setoro	sedamdesetak	sedamdesetogodišnjak sedamdesetogodišnji sedamdesetogodišnjica
80	osamdeset	osmdeseti	osamdesetorica	osamdesetoro	osamdesetak	osamdesetogodišnjak osamdesetogodišnji osamdesetogodišnjica
90	devedeset	devedeseti	devedesetorica	devedesetoro	devedesetak	devedesetogodišnjak devedesetogodišnji devedesetogodišnjica
100	sto	stoti			stotinjak	stogodišnjak stogodišnji stogodišnjica

---

4 In the table only the male person in given, but in the dictionary both male and female forms are entry-words. More on male/female relations in *Mrežnik* see in Hudeček and Mihaljević 2019d.

Entry-words divided into semantic groups are listed in Table 2.

Table 2: Entry-words divided into semantic groups

Croatian	English	
Semantička skupina	**Semantic group**	**Entry-words**
naprava	**device**	baterija, brava, brisač, brklja, crpka, čegrtaljka, čitač, dalekozor, odskočna daska, dubinomjer, giljotina, glazbalo, greda, gromobran, hodalica, hvataljka, karika, kišobran, klapa, klopka, kočnica, kompas, kormilo, kotač, kutomjer, ljestve, ljuljačka, megafon, mijeh, mina, mišolovka, metronom, mjenjač, nosač, nosila, njihaljka, otvarač, padobran, parangal, penjalica, pomagalo, praćka, preča, prekidač, preslica, prigušivač, pročišćivač, projektil, raspršivač, rog, roštilj, ruče, sanjke, sat, sidro, sirena, sisaljka, sito, sjedalo, slavina, stalak, (tkalački) stan, stezaljka, stupica, svjetiljka, šestar, šiljilo, škare, štit, štitnik, štrcaljka, teleskop, termofor, termometar, tijesak, tipkalo, tobogan, toplomjer, truba, tuš, udica, upaljač, upravljač, vadičep, vaga, ventil, vez, vitao, vješala, volan, vratilo, vremeplov, zamka, zviždaljka, zvonce, zvono
naslov osobe	**title of a person**	aga, arkanđeo, ban, barun, beg, biskup, car, dahija, dalaj-lama, derviš, despot, dužd, faraon, grof, herceg, hodža, imam, kardinal, knez, kralj, maestro, maharadža, markiz, metropolit, nadbiskup, nadvojvoda, nuncij, otac, papa, primarijus, satrap, sultan, šah, šejk, šerif, velečasni, vojvoda, župan
nebeska tijela	**celestial bodies**	Merkur, Venera, Zemlja, Mars, Jupiter, Saturn, Uran, Neptun, Pluton, Mjesec, Veliki medvjed, Mali medvjed…
kemijski elementi	**chemical elements**	kovine (metal): aluminij, bakar, željezo, živa… nekovine (non-metal): kisik, ugljik, vodik…
oborine	**precipitation**	kiša, snijeg, susnježica, tuča
obroci	**meals**	zajutrak, doručak, užina, ručak, večera

Dictionary compilation is based on the fact that a dictionary is a highly structured document and the modeling of dictionary micro and macrostructure (dictionary

grammar) is at the center of the project. In modeling the structure of *Mrežnik* many similar foreign dictionaries were consulted: *elexiko* (German, www.owid. de/docs/elex/start.jsp), *Wielki słownik języka polskiego (Polish,* www.wsjp.pl), *Karp* (Swedish, spraakbanken.gu.se/karp), *Algemeen Nederlands Woordenboek* (ANW Dutch, http://anw.inl.nl/), *Slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZ* (Slovenian, fran.si).[5] *elexiko* was the main inspiration for creating some of the dictionary fields. At the very beginning of the project, the editors of *Mrežnik* visited the Institute for the German Language in Mannheim and got acquainted with the structure of *elexiko.* Visits of other project members to the Institute for the German Language[6], the University of Gothenburg[7], the Institute of the Norwegian Language, University of Copenhagen, Danish Society for Language and Literature, Danish Center for Language Technologies, Árni Magnússon Institute for Icelandic Studies, and Fran Ramovš Institute of the Slovenian Language followed.

2. Project goals

The main project goals are:

1. the creation of a basic dictionary of 10,000 entries (with accentuated entry-words and forms in the grammatical block systematically established for each word class, definitions (for adult native speakers, for schoolchildren aged 7 to 9, and for non-native speakers), examples, synonyms, antonyms, sub-entries, idioms), and detailed collocation descriptions;

2. connecting the dictionary with the databases created as a part of the *Mrežnik* project in parallel with dictionary processing: the linguistic advice database (300 pieces of advice), the conjunction database with the description of conjunction groups and modifications (for all conjunctions in the dictionary), the database of idioms (50 idioms), the database of ethnics and ktetics (300 inhabitant names – ethnics and adjectives derived from place-names – ktetics)[8]; Figure 1 shows the

---

[5]   Reports on all visits can be seen on the project webpage http://ihjj.hr/mreznik/page/usavrsavanje-i-diseminacija/4/.
[6]   Report on this visit can be seen in Mihaljević, Josip 2019c.
[7]   Report on this visit can be seen in Lazić 2019.
[8]   Available at the moment on the webpage *Hrvatski u školi* (hrvatski.hr).

database of idioms, Figure 2 shows the database of language advice, and Figure 3 shows the database of ethnics and ktetics;

Ahilova peta
Izraz je utemeljen na grčkome mitu o neustrašivome borcu Ahileju, kojega je njegova majka Tetida, u želji da ga uči neranjivim, nakon rođenja umočila u rijeku Stiks. Pritom ga je držala za petu, koja je ostala suha. Tako je peta postala jedino ranjivo mjesto na njegovu tijelu. Kad kažemo da je što čija Ahilova peta, želimo reći da je to njegovo ranjivo mjesto ili slaba strana.

Arijadnina nit

Augijine štale

baciti bisere pred svinje

crna ovca

Damoklov mač

jabuka razdora

kocka je bačena

Figure 1: Database of idioms

crkva i Crkva

cura > djevojčica, djevojka
Riječ *cura* upotrebljava se u razgovoru. Ona ne pripada hrvatskomu standardnom jeziku. U hrvatskome standardnom jeziku treba umjesto riječi *cura* upotrijebiti riječi *djevojčica* ili *djevojka*. Zato su u njemu točne rečenice *Djevojčica ide u kino.* i *Djevojka ide u kino.*, a nije točna rečenica *Cura ide u kino.*

čistač i čistačica

dečko > dječak, mladić

dešnjak i dešnjakinja

doktor > liječnik

doktorica > liječnica

dućan > prodavaonica, trgovina

Figure 2: Database of language advice for schoolchildren

Beč, Bečanin, Bečanka, bečki
Bêč, G Béča, D Béču, A Bêč, L u Béču, I Béčom
Béčanin, GA Béčanina, DL Béčaninu, V Béčanine, I Béčaninom; *mn.* NV Béčani, G Béčānā, DLI Béčanima, A Béčane
Béčānka, G Béčānkē, DL Béčānki, A Béčānku, V Béčānko, I Béčānkōm; *mn.* NAV Béčānke, G Béčānkā/Béčānkī, DLI Béčānkama
bēčkī

Napomena: U nekim starijim tekstovima i pravopisima s početka XX. stoljeća zapisan je i muški etnik Bečlija. Na kajkavskome području Beč se nekoć nazivao i Dunaj.

Belgija, Belgijac, Belgijka, belgijski

Beli Manastir, Belomanastirac, Belomanastirka, belomanastirski

Figure 3: Database of ethnics and ktetics

3. connecting the basic dictionary with other web sources currently being compiled at the Institute of Croatian Language and Linguistics, e.g. the Croatian Verb Valency Database *e-Glava*, the Croatian Collocation Database, the Croatian

Special Field Terminology database *Struna*[9];

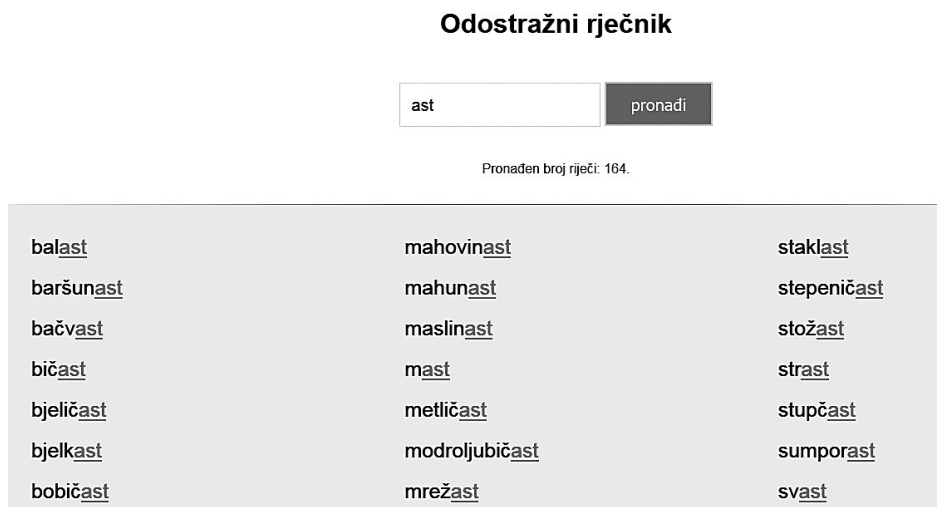4. the creation of a reversed dictionary based on the *Mrežnik* wordlist[10] as shown in Figure 4.

Odostražni rječnik

ast	pronađi

Pronađen broj riječi: 164.

bal<u>ast</u>	mahovin<u>ast</u>	stakl<u>ast</u>
baršun<u>ast</u>	mahun<u>ast</u>	stepenič<u>ast</u>
bačv<u>ast</u>	maslin<u>ast</u>	stož<u>ast</u>
bič<u>ast</u>	m<u>ast</u>	str<u>ast</u>
bjelič<u>ast</u>	metlič<u>ast</u>	stupč<u>ast</u>
bjelk<u>ast</u>	modroljubič<u>ast</u>	sumpor<u>ast</u>
bobič<u>ast</u>	mrež<u>ast</u>	sv<u>ast</u>

Figure 4: Demo version of the reversed dictionary

The additional goals of the project were listed in the project application as follows:

a) initiate stronger research of monolingual e-lexicography in Croatia[11];

b) organize a scientific conference on monolingual e-lexicography[12];

---

9   More about the links to other external databases see in Hudeček and Mihaljević 2019b.

10   A pilot version of the reverse dictionary has been compiled and is publically available on the site of the project *Muško i žensko u hrvatskome jeziku* (Male and Female in the Croatian Language, address http://ihjj.hr/projekt/musko-i-zensko-u-hrvatskome-jeziku/72/). For more details about this reverse dictionary see Lewis and Mihaljević 2018. The final version of the reverse dictionary will be compiled after the compilation of the final wordlist of *Mrežnik* and will be available on the *Mrežnik* website.

11   At the moment of writing of this paper, 53 papers (scientific, case studies, preliminary publications, review papers, professional papers, and papers for the popularization of science) have been published; in 2020 – 5 papers, in 2019 – 20 papers, in 2018 – 15 papers, and in 2017– 13 papers have been published.

12   The international conference *E-dictionaries and E-lexicography* was held on 10-11 May 2019 in Zagreb. The conference aimed at bringing together Croatian and foreign experts in e-lexicography, to present Croatian e-lexicography to the wider public, as well as to get acquainted with contemporary research and ideas in the field of e-lexicography. The conference focused on the following topics: e-dictionaries, e-encyclopedias, corpora in lexicography, databases on language and terminology, e-terminology, retrodigitalization, computer tools for lexicography. More detail about the conference can be found on http://ihjj.hr/mreznik/page/e-rjecnici-i-e-leksikografija/8.

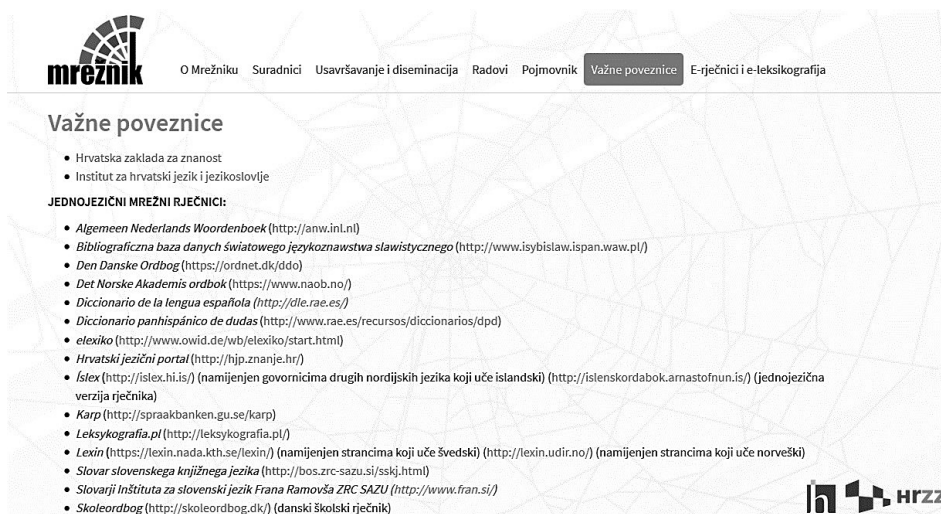c) compile a specialized website on monolingual e-lexicography in Croatia[13] as shown in Figure 5;

Figure 5: *Mrežnik* webpage with data on e-dictionaries

d) publish an e-monograph on *Mrežnik* and the proceedings of the conference.[14]

One of the additional results not mentioned in the project application is the compilation of the *Glossary of e-lexicography* which is published on the project webpage (http://ihjj.hr/mreznik/page/pojmovnik/6/). This glossary is being compiled in collaboration with the project *Croatian linguistic terminology – Jena*[15]. The glossary (Figure 6) lists the most important Croatian terms of e-lexicography and names relevant for e-lexicography with English equivalents, a brief definition, useful links, and illustrations.

---

[13]   The address of the website is http://ihjj.hr/mreznik.

[14]   On project goals see also on the project website http://ihjj.hr/mreznik/page/o-mrezniku/1.

[15]   More on the collaboration of these two projects see in Hudeček and Mihaljević 2019a, and more on *Jena*, another project financed by the Croatian Science Foundation (Struna-2017-09-05), see on the project webpage http://ihjj.hr/projekt/hrvatsko-jezikoslovno-nazivlje-ndash-jena/85/.

Figure 6: Glossary of e-lexicography

3. Technical aspects of *Mrežnik*

The two basic computer tools for the compilation of this three-module diction-
ary are *Sketch Engine*, a corpus query system (loaded with the corpora) to sup-
port the analysis of language, and *TLex*, a dictionary writing system.

Data extraction from the corpora is performed with the *Sketch Engine* web tool, which allows the display of the lexeme context through *Word Sketches*, the most common collocations sorted into syntactic categories and the discovery of good examples of word usage or collocations. *Word Sketches* are specially adapted to the needs of the project, which are based on developed *Sketch Grammar. Sketch Grammar* for Croatian was written by Nikola Ljubešić for the *Mrežnik* project. *Word Sketches* were checked by the *Mrežnik* team and some categories were added to *Word Sketches*. However, for some categories, regular expressions still had to be used[16], e.g. with interjections and conjunctions where *Sketch Grammar* didn't give adequate results. Table 3 gives an overview of the additions to *Word Sketches* by categories requested by the *Mrežnik* team.[17]

Table 3: Addition to *Word Sketches* by categories

Categories added to *Word Sketches*					Regular expressions used to search the corpus	
Verb (V)	Noun (N)	Adjective (Adj)	Adverb (Adv)	Particle (P)	Interjection (I)	Conjunction (C)
V + *se*, *se* + V, V + pronoun + P	N + Pr, Pr + N	Adj + Pr, Pr + Adj	Pr + Adv, Adv + P	P + Adv, P + Adj, Pr + P, P + Pr	Adv + I, V + I, I + Pr, I + N, I + i + I, I + I, I + Pr, P + I	Adv + C, P + C, C + Adv

Mrežnik is compiled in the *TLex* program, which has been adapted to the needs of the project. Figure 7 shows the fields of the entry *pisač* ('printer') in *TLex*.

---

[16] On the use of regular expressions on the Croatian corpus see more in Blagus Bartolec and Matas Ivanković 2017.

[17] More on these additions to *Word Sketches* by the *Mrežnik* team see in Hudeček and Mihaljević 2018b.
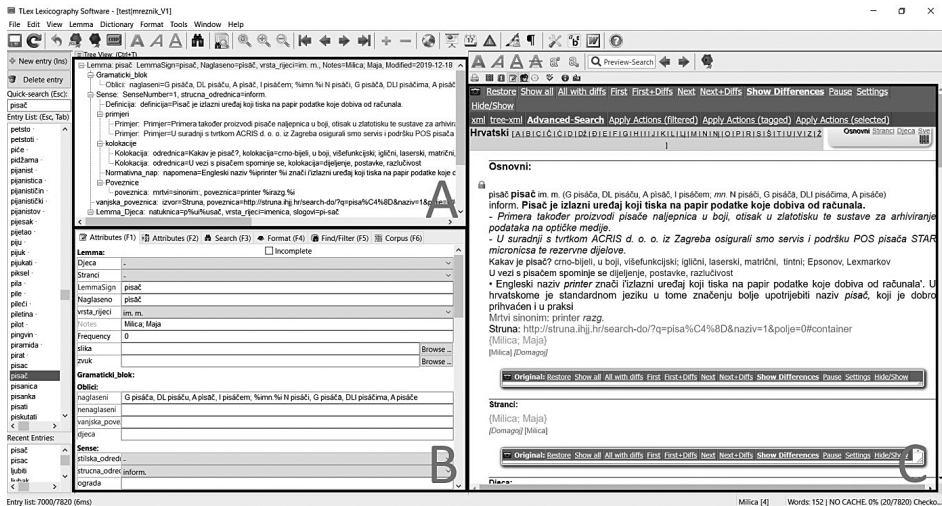
Figure 7: Entry *pisač* ('printer') in *TLex*

Table 4 explains the three fields of each entry in *TLex*.

Table 4: Parts of *TLex* interface

A	**The three view** – for structuring and selecting entries
B	**Fields** – for entering data for each entry element
C	**Preview** – for viewing the currently edited entry

The entry *pisač* ('printer') was selected because of its shortness and simplicity. Figure 8 shows the parts of the entry in the preview. The entry is linked to the terminological database *Struna*.

Figure 8: Parts of the entry *pisač* ('printer')

After lexicographic processing is completed, the data will be exported from *TLex* to both the web application and the CLARIN European science infrastructure

repository (clarin.si repository and the github.com public data management system). This will make *Mrežnik* available for use both via a web application and for machine implementation by downloading data from the CLARIN repository.

4. Corpora and wordlist

Mrežnik is based on two Croatian corpora: the *Croatian Language Repository* (http://riznica.ihjj.hr/index.hr.html) and the *Croatian Web Corpus hrWaC* (http://nlp.ffzg.hr/resources/corpora/hrwac). These corpora differ in size and style: *hrWaC* is much bigger and consists of texts mostly from newspapers, forums, blogs, etc. written in the journalistic and/or colloquial style. *CLR* is smaller and consists mostly of older texts (written before 2009) often belonging to the literary style. These corpora are managed by *Sketch Engine*. *Mrežnik* is corpus-based and not corpus-driven. The reason for using the corpus-based and not corpus-driven approach is the normative aspect of *Mrežnik* but also the fact that the two available Croatian corpora are neither the corpora of the standard Croatian language nor are they representative corpora of the Croatian language. In addition to these corpora, all other available print and web sources are taken into account in writing definitions and providing examples and meanings.

The frequency lists of *hrWaC* (first 12,000 words) and the *Croatian Language Repository* (first 10,000 words) were overlapped, all words present only in *Croatian Language Repository* and not present in *hrWaC* were extracted, their frequency was multiplied by four, and they were added to the shared list. (The alphabetical list also has a word type filter [POS], allowing e.g. only verbs to be filtered). This wordlist (first 8,000 entries) was juxtaposed with two separate wordlists: the wordlist for the module for children (which was extracted from textbooks for the first four grades of elementary school with some additions by the collaborators of *Mrežnik*) and the wordlist for the module for non-native speakers (which includes 1,000 words selected from lists in textbooks for non-native speakers) to ensure that words found in these lists (which partly overlap) appear in the list for adult native speakers. This wordlist was supplemented with male/female[18] and aspectual pairs, possessive and descriptive adjectives, adverbs

---

[18] More on male and female pairs in *Mrežnik* see in Mihaljević 2018 and Hudeček and Mihaljević 2019d.

derived from adjectives from the list, nouns ending in -*ost* derived from adjectives from the list, numerous grammatical and semantic groups, etc. Nevertheless, this was not done at all costs, e.g. if these additions are not supported by corpus and Internet data and/or the language experience of native speakers. This means that e.g. *brodski kuhar* ('ship's cook') or *trkaći inženjer* ('race engineer') will not have their female pairs *brodska kuharica* or *trkaća inženjerka* as there are no attestations of these collocation.

This resulted in a wordlist of 10,000 words and phrases (entry-words and sub-entries) with two separate wordlists of 3,000 words (for children) and 1,000 words (for non-native speakers).

The wordlist of the module for children was considered as the basic wordlist (the majority of the 1,000 entries for non-native speakers were also included in the module for children)[19] and the compilation started with entry-words from this list to make processing for the module for children as compatible as possible with that for adult native speakers.[20] The decision to create a wordlist this way (and not from balanced and overlapped wordlists of both corpora) was motivated by the idea that the compilation of *Mrežnik* will not stop when the project is over and therefore the choice of headwords included in the dictionary during the project period doesn't have to be based on frequency. Rather it is based on the language system, word formation, and semantic relations.

As *Mrežnik* is a corpus-based, not a corpus-driven dictionary the corpus and all data extracted from it serve only as guidelines. The glossary on the *Mrežnik* website *ihjj.hr/mreznik* defines a corpus-based dictionary as follows: *a dictionary for which the lexicographer uses a corpus, but can freely decide what should be included in the dictionary, allowing the dictionary to be supplemented with words from other sources if necessary, as well as collocations and meanings not attested in the corpus.*

It follows that, in compiling a dictionary entry, meanings, and collocations can be added to a particular entry or collocation field regardless of word sketches and corpus attestations.

---

19   More on the compilation of wordlists for children and non-native speakers see in Mihaljević 2018a.
20   See information on the webpage http://ihjj.hr/mreznik/page/o-mrezniku/1/ and in Hudeček and Mihaljević 2018.

In addition to *Word Sketches*, for each entry dozens of pages in *hrWaC* and *Croatian Language Repository* have also been checked, as this sometimes resulted in more typical collocations than those found in *Word Sketches*, e.g. *Word Sketches* for *kuharica* ('female chef') do not give the collocation *glavna kuharica* ('head female chef'), however, this collocation is found if one directly searches the corpus.

Figure 9 shows the wordlist of *Mrežnik* in *TLex*.

Figure 9: Lemma *list* in the *TLex* program

5. Normative aspects of *Mrežnik*

The Croatian Web Dictionary – Mrežnik is a normative dictionary. Its compilation is followed by research of normative criteria to be used in a lexicographic work.[21] Its normative nature is apparent in the following:

---

[21]   On normative criteria see Hudeček and Mihaljević 2019c.

1. the selection of entry-words and referencing (the wordlist is not established automatically based on a list created from the overlap of the frequency lists of the *Croatian Language Repository* and *hrWaC*; instead, in composing the word-list, the fact that *Mrežnik* is a corpus-based dictionary and not a corpus-driven dictionary was taken into account; due to frequency, entry-words were included that are not a part of the Croatian standard language or are not recommended in the standard language as they belong to its colloquial register; the dictionary entry of these entry-words (definitions, meanings, examples, collocations, etc.) was compiled, however, a normative note has been included explaining why the word (phrase) does not belong to the standard language; this normative note also appears in the entry of synonymous words that belong to the standard language (e.g. for both *stomatologica* and *stomatologinja* 'female dentist');

2. the accentuation of entry-words based on neo-Štokavian accentuation rules, i.e. forms don't have accents on the last syllable or descending accents in the middle of the word;[22]

3. the selection of forms in the grammatical block; only forms acceptable by the norm are given;

4. the selection of examples (the collaborators try to select examples with no language errors while examples with language errors are edited);

5. giving linguistic advice in all three modules:

- the advice is taken from *Jezični savjetnik* (Language advice, jezicni-savjetnik.hr) and *Bolje je hrvatski* (Better in Croatian, bolje.hr)[23] and if needed it is modified according to the needs of the entry or a new piece of advice is written especially for *Mrežnik*, e.g. in the entry of the word *razdoblje* ('period') the advice is given that the common phrase *vremensko razdoblje* ('time period') is a pleonasm and should be replaced by just *razdoblje* ('period'), as shown in Figure 10.

---

[22]   See Vidović 2020a, Vidović 2020b.

[23]   The normative advice in the entry *pisač* in Figure 8 is from the website *Bolje je hrvatski*.

Figure 10: Entry *razdoblje* ('period') with a normative note

6. Micro and macrostructure of *Mrežnik*

Mrežnik consists of these three modules:

– the module for adult native speakers of Croatian;[24]

– the module for elementary school children;[25]

– the module for non-native speakers of Croatian.[26]

Data in all these modules are coordinated and synchronized. However, each module functions as a separate dictionary compiled for a different target group of users.

The module for non-native speakers will contain audio recordings and games that will make learning Croatian easier and more fun. This is made in collaboration with *Croaticum*, Centre for Croatian as Foreign and Second Language, and Dinka Pasini, who works in *Croaticum*, is a member of the *Mrežnik* team. The structure of the module for non-native speakers was discussed with the students of *Croaticum* on three workshops in 2017, 2018 (see Hudeček, Mihaljević, and Pasini 2018), and 2019.

---

[24] More on this module see in Hudeček 2018.

[25] More on this module see in Hudeček and Mihaljević 2019, Mihaljević, Milica 2018a.

[26] More on this module see in Hudeček, Mihaljević, and Pasini 2018, Mihaljević, Milica 2018a.

The microstructure of *Mrežnik* is shown in the diagram below. The diagram shows the microstructure of the module for adult native speakers. The microstructure of other modules is just the reduction of this basic module, so it is not shown by a separate diagram.
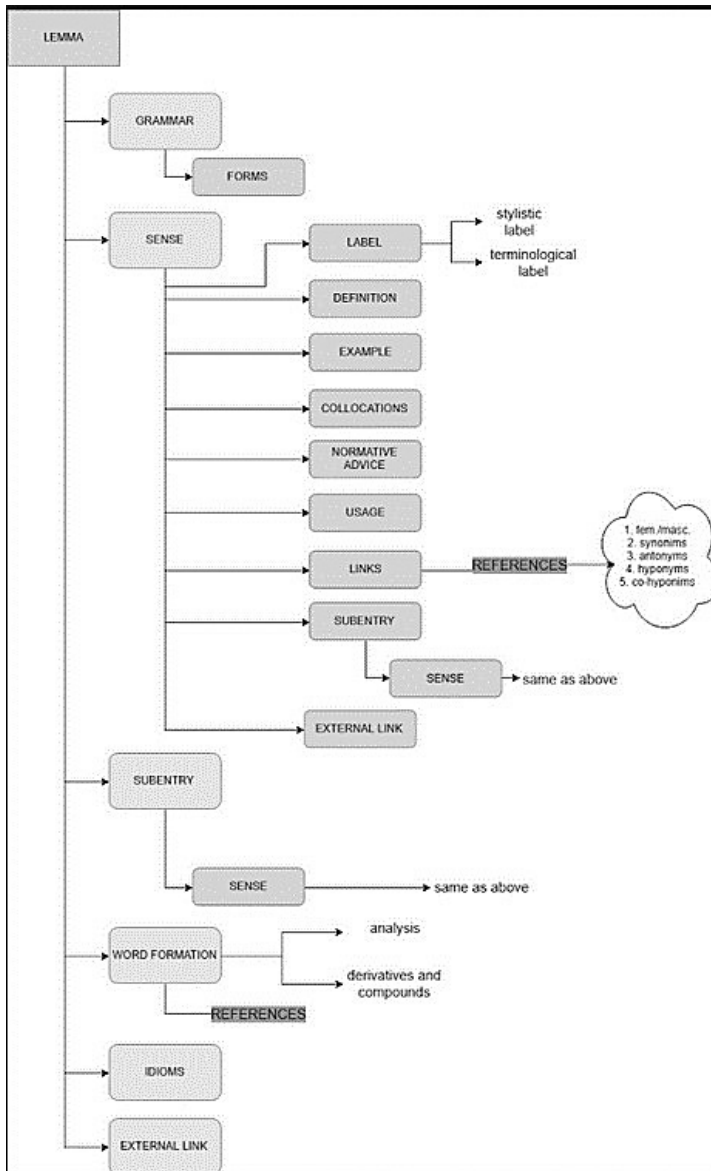

Figure 11: The microstructure of *Mrežnik*

The approach to some elements of the microstructure will be explained below:

- Definitions are full sentences. (*List je… / A leaf is…*). They are formed systematically according to grammatical and semantic groups, e.g. closed grammatical groups: interjections, conjunctions, particles, prepositions, adverbs (not derived from adjectives), and semantic groups: days of the week, months, chemical elements, units of measurement, periods in art and history, abbreviations.

- Examples are taken from the corpus and the compilers try to find examples that require no intervention. However, they are, if needed, edited by correcting obvious mistakes following the *Croatian Orthography Manual* (Jozić et al. 2013), deleting names, shortening sentences, etc.

- Collocations are based on *Word Sketches*. The project collaborators make informed choices from *Word Sketches* to provide collocations that are representative of the Croatian standard language and not of the corpus. In choosing collocations, all features of the corpus are taken into account and their suitability for inclusion in *Mrežnik* is evaluated regardless of their degree of attestation in the corpus (e.g. the collocations *brkata konobarica* ('waitress with a mustache'), *sisata konobarica* ('large-breasted waitress'), *silovana konobarica* ('raped waitress'), *pohotna pekarica* ('lusty baker') would not be suitable collocations for *Mrežnik* although they appear in the corpus).[27]

- Normative and usage advice is given in all cases where the user might not be sure which word to use or when to use a specific word.

- Internal links are links that link one *Mrežnik* entry (i.e. lemma or sense) to another. Links that connect one lemma to another are used for compounds and derivatives while links that link senses are used for synonyms, antonyms, hyponyms, co-hyponyms, meronymes, and masculine/feminine pairs.[28]

External links are links to databases being created by project collaborators or other Institute members within the framework of other projects. These projects

---

[27] See Hudeček and Mihaljević 2018b.
[28] See Hudeček and Mihaljević 2017b, and Hudeček and Mihaljević 2019b.

are the *Valence Database* (http://ihjj.hr/projekt/baza-hrvatskih-glagolskih-valen-cija), *Repository of Metaphors* (http://ihjj.hr/metafore), *Terminology Database Struna* (http://struna.ihjj.hr/), etc. and databases composed within the framework of the *Mrežnik* project are *Language Advice for Schoolchildren* (http://hrvatski.hr/savjeti), *Repository of Conjunctions*, *Repository of Idioms* (http://hrvatski.hr/frazemi), and *Repository of Ethnics and Ktetics* (http://hrvatski.hr/etnici-i-ktetici).

Word formation models are given for all derivatives and compounds.

7. Conclusion

In a time in which the concepts of dictionary and (online) e-dictionary have become nearly equivalent, as have the concepts of lexicography and e-lexicography, Croatia still belongs to the ever-smaller number of countries with no free online national language dictionary founded on modern e-lexicography, nor has systematic scientific research been carried out in this area. Therefore, the basic goal of the project *Croatian Web Dictionary – Mrežnik* was to change this in both of the aforementioned aspects. In the paper, we show the approach to a Croatian e-dictionary and what has been achieved in three years from the beginning of the project. As *Mrežnik* is a scientific project it is compiled taking into account semantic relations and the systematic nature of language. It is based on the fact that a dictionary is a highly structured document and the 'dictionary grammar' with its micro and macrostructure is at the center of the project. Entries, entry-words, and sub-entries are interconnected, as well as linked with other databases.

Mrežnik is compiled bearing in mind that it will be a long-term project after the end of the project period. During the four year project, a model will be compiled which will, hopefully, be supplemented in the many years to come.

References

Blagus Bartolec, Goranka; Matas Ivanković, Ivana. 2017. Kad nam korpus ispunjava želje. *Hrvatski jezik* 4/3. 25–28.

Čilaš Šimpraga, Ankica; Jojić, Ljiljana; Lewis, Kristian. 2008. *Prvi školski rječnik hrvatskoga jezika*. Školska knjiga – Institut za hrvatski jezik i jezikoslovlje. Zagreb.

Hudeček, Lana. 2018. Izazovi leksikografske obrade u jednojezičnome mrežnom rječniku (na primjeru *Hrvatskoga mrežnog rječnika – Mrežnika*). *Вісник Львівського університету. Серія філологічна* 69. Ed. Salyha, Т. Львівський національний університет імені Івана Франка. Львів. 29–38. [Hudeček, Lana. 2018. Izazovi leksikografske obrade u jednojezičnome mrežnom rječniku (na primjeru *Hrvatskoga mrežnog rječnika – Mrežnika*). *Visnik Lʹvivsʹkogo universitetu. Serija filologična* 69. Ed. Salyha, T. Lʹvivsʹkij nacionalʹnij universitet imeni Ivana Franka. Lʹviv. 75–98.]

Hudeček, Lana; Mihaljević, Milica (eds.). 2012. *Školski rječnik hrvatskoga jezika*. Institut za hrvatski jezik i jezikoslovlje – Školska knjiga. Zagreb.

Hudeček, Lana; Mihaljević, Milica. 2017a. A New Project – Croatian Web Dictionary MREŽNIK. *The Future of Information Sciences. INFuture2017, Integrating ICT in Society.* Eds. Atanassova, Iana et al. Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences. Zagreb. 205–213. doi.org/10.17234/infuture.2017.21.

Hudeček, Lana; Mihaljević, Milica. 2017b. Hrvatski mrežni rječnik – Mrežnik. *Hrvatski jezik* 4/4. 1–8.

Hudeček, Lana; Mihaljević, Milica. 2017c. The Croatian Web Dictionary Project – Mrežnik. *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference.* Eds. Kosem, Iztok et al. Lexical Computing CZ s.r.o. Brno – Leiden. 172–192.

Hudeček, Lana; Mihaljević, Milica. 2018a. *Hrvatski mrežni rječnik – Mrežnik (Upute za obrađivače)*. Radna inačica. http://ihjj.hr/mreznik/uploads/upute.pdf (accessed December 2019).

Hudeček, Lana; Mihaljević, Milica. 2018b. Croatian Web Dictionary Mrežnik: One year later – What is different?. *Proceedings of the Conference on Language Technologies & Digital Humanities.* Eds. Fišer, Darja; Pančur, Andrej. Znanstvena založba Filozofske fakultete v Ljubljani. Ljubljana. 106–113.

Hudeček, Lana; Mihaljević, Milica. 2019a. Croatian Web Dictionary – Mrežnik vs. Croatian Linguistic Terminology – Jena. *INFuture 2019: Knowledge in the Digital Age.* Eds. Bago, Petra; Hebrang Grgić, Ivana; Ivanjko, Tomislav; Juričić, Vedran; Miklošević, Željka; Stublić, Helena. Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences. Zagreb. 22–31. doi.org/10.17234/infuture.2019.4.

Hudeček, Lana; Mihaljević, Milica. 2019b. Croatian Web Dictionary – Mrežnik – Linking with Other Language Resources. *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference.* Eds. Kosem, Iztok et al. Lexical Computing CZ, s.r.o. Brno. 72–98.

HUDEČEK, LANA; MIHALJEVIĆ, MILICA. 2019c. Kako normativnu preporuku donosi terminolog, a kako standardolog. *Jezik i um. Zbornik radova s međunarodnoga znanstvenog skupa Hrvatskoga društva za primijenjenu lingvistiku održanoga od 3. do 5. svibnja 2018.* Eds. Matešić, Mihaela; Vlastelić, Anastazija. Srednja Europa. Zagreb. 3–30.

HUDEČEK, LANA; MIHALJEVIĆ, MILICA. 2019d. Profesijski nazivi u hrvatskoj e-terminografiji i e-leksikografiji. *Studia lexicographica* 13/24. 75–95. doi.org/10.33604/sl.13.24.2.

HUDEČEK, LANA; MIHALJEVIĆ, MILICA. 2019e. Uloga e-učenja i e-rječnika u usvajanju hrvatskoga jezika. *Dijete i jezik danas – Razvoj pismenosti u materinskom i inom jeziku.* Eds. Majdenić, Valentina; Trtanj, Ivana; Živković Zebec, Vedrana. Fakultet za odgojne i obrazovne znanosti. Osijek. 85–111.

HUDEČEK, LANA; MIHALJEVIĆ, MILICA; PASINI, DINKA. 2018. Radionica na Croaticumu – provjera strukture modula za strance *Hrvatskoga mrežnog rječnika – Mrežnika.* *Hrvatski jezik* 5/2. 9–12.

JOZIĆ, ŽELJKO et al. 2013. *Hrvatski pravopis.* Institut za hrvatski jezik i jezikoslovlje. Zagreb.

LAZIĆ, DARIA. 2019. Mrežni izvori Sveučilišta u Göteborgu. *Hrvatski jezik* 6/3. 21–26.

LEWIS, KRISTIAN; MIHALJEVIĆ, JOSIP. 2018. Odostražni rječnik – što je, kako ga izraditi i čemu služi. *Hrvatski jezik* 5/2. 21–25.

MATIJEVIĆ, MAJA; MIHALJEVIĆ, JOSIP. 2019. Arabic Speakers as Croatian Language Learners – Electronic Educational Games as a Support for Learning. *INFuture 2019: Knowledge in the Digital Age.* Eds. Bago, Petra et al. Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences. Zagreb. 135–145. doi.org/10.17234/infuture.2019.17.

MIHALJEVIĆ, ANA; MIHALJEVIĆ, JOSIP. 2019. Mrežne igre u poučavanju i učenju hrvatskoga jezika. *Dijete i jezik danas – Razvoj pismenosti u materinskom i inom jeziku.* Eds. Majdenić, Valentina; Trtanj, Ivana; Živković Zebec, Vedrana. Fakultet za odgojne i obrazovne znanosti. Osijek. 113–137.

MIHALJEVIĆ, JOSIP. 2018. Zabavna glagoljica. *Hrvatski jezik* 5/4. 25–27.

MIHALJEVIĆ, JOSIP. 2019a. Games for Learning Old and Special Alphabets – The Case Study of Gamifying *Mrežnik.* *CLiC-it 2019: Italian Conference on Computational Linguistics.* Eds. Bernardi, Raffaella; Navigli, Roberto; Semeraro, Giovanni. AILC. Bari.

MIHALJEVIĆ, JOSIP. 2019b. Gamification in E-Lexicography. *INFuture 2019: Knowledge in the Digital Age.* Eds. Bago, Petra et al. Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences. Zagreb. 155–164. doi.org/10.17234/infuture.2019.19.

MIHALJEVIĆ, JOSIP. 2019c. Jezične tehnologije i mrežni izvori Instituta za njemački jezik. *Hrvatski jezik* 6/2. 22–25.

Mihaljević, Milica. 2018a. Hrvatski mrežni izvori za djecu i strance. *Вісник Львівського університету. Серія філологічна* 69. Ed. Salyha, T. Львівський національний університет імені Івана Франка. Львів. 75–98. [Mihaljević, Milica. 2018a. Hrvatski mrežni izvori za djecu i strance. *Visnik Ľvivsʹkogo universitetu. Serija filologična* 69. Ed. Salyha, T. Ľvivsʹkij nacionalʹnij universitet imeni Ivana Franka. Ľviv. 75–98.]

Mihaljević, Milica. 2018b. Muško i žensko u e-rječniku. *Jezik i njegovi učinci: Zbornik radova s međunarodnoga znanstvenog skupa Hrvatskoga društva za primijenjenu lingvistiku održanoga od 4. do 6. svibnja 2017.* Eds. Stolac, Diana; Vlastelić, Anastazija. Srednja Europa. Zagreb. 209–228.

Štrkalj Despot, Kristina, Möhrs, Christine. 2015. Pogled u e-leksikografiju. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 41/2. 329–353.

Štrkalj Despot, Kristina; Hudeček, Lana; Stojanov, Tomislav; Ljubešić, Nikola. 2019. State-of-the-art on monolingual lexicography for Croatia (Croatian). *Slovenščina 2.0.* 7/1. 65–77. doi.org/10.4312/slo2.0.2019.1.65-77.

Vidović, Domagoj. 2020a. Hrvatski naglasni sustav između deskriptivnoga i preskriptivnoga pristupa. *Jezikoslovlje* 21/1. 7–25. doi.org/10.29162/jez.2020.1.

Vidović, Domagoj. 2020b. O naglasnoj normi na temelju obradbe naglasaka u Školskome rječniku hrvatskoga jezika i Mrežniku. *Od norme do uporabe 3.* Eds. Rišner, Vlasta; Glušac, Maja; Mikota, Jadranka. Filozofski fakultet Sveučilišta Josipa Jurja Strossmayera u Osijeku – Hrvatska sveučilišna naklada. Zagreb.

Data bases and webpages

Algemeen Nederlands Woordenboek ANW. http://anw.inl.nl (accessed December 2019).

Baza hrvatskih glagolskih valencija – GLAVA. http://ihjj.hr/projekt/baza-hrvatskih-glagolskih-valencija (accessed June 2019).

Bolje je hrvatski. http://bolje.hr (accessed June 2019).

elexiko. www.owid.de/docs/elex/start.jsp (accessed December 2019).

Hrvatski u školi. hrvatski.hr (accessed December 2019).

Hrvatsko jezikoslovno nazivlje – Jena. http://ihjj.hr/projekt/hrvatsko-jezikoslovno-nazivlje-ndash-jena/85/ (accessed December 2019).

Hrvatsko strukovno nazivlje – Struna. http://struna.ihjj.hr/ (accessed December 2019).

Jezični savjetnik. http://jezicni-savjetnik.hr/ (accessed June 2019).

Karp. http://spraakbanken.gu.se/karp/ (accessed December 2019).

Muško i žensko u hrvatskome jeziku. http://ihjj.hr/projekt/musko-i-zensko-u-hrvatskome-jeziku/72/ (accessed December 2019).

Slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZ. http://fran.si (accessed December 2019).

Wielki słownik języka polskiego. www.wsjp.pl (accessed December 2019).

Projekt *Hrvatski mrežni rječnik – Mrežnik* – ciljevi i postignuća

Sažetak

Projekt *Hrvatski mrežni rječnik – Mrežnik* u trenutku pisanja ovoga rada na kraju je treće projektne godine. Cilj je projekta objava slobodno dostupnoga, na korpusu utemeljenoga, hipertekstoga, pretraživoga normativnog jednojezičnog mrežnog rječnika hrvatskoga jezika utemeljenoga na postignućima suvremene e-leksikografije i računalnoga jezikoslovlja. Mrežnik uključuje tri modula: modul za odrasle govornike hrvatskoga jezika, modul za učenike nižih razreda osnovne škole te modul za osobe koje uče hrvatski kao ini jezik. Mrežnik je zamišljen kao mjesto povezivanja postojećih jezičnih izvora Instituta za hrvatski jezik i jezikoslovlje, ali u sklopu projekta bit će izrađeni i novi jezični izvori. Nakon završetka četverogodišnjega projektnog razdoblja zamišljeno je da projekt *Hrvatski mrežni rječnik – Mrežnik* postane trajni projekt Instituta za hrvatski jezik i jezikoslovlje te da se rječnički sadržaj stalno proširuje i dopunjuje te s vremena na vrijeme objavljuju nove inačice.

Keywords: *Mrežnik*, e-dictionary, e-lexicography, dictionary microstructure, dictionary macrostructure
Ključne riječi: *Mrežnik*, e-rječnik, e-leksikografija, rječnička mikrostruktura, rječnička makrostruktura