The Croatian Web Dictionary Project – Mrežnik

Lana Hudeček and Milica Mihaljević

Institute of Croatian Language and Linguistics, Republike Austrije 16, Zagreb, Croatia
E-mail: lhudecek@ihjj.hr, mmihalj@ihjj.hr

Abstract

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. This paper presents the current, as well as the planned, activities of the Croatian Web Dictionary—MREŽNIK project. The aim of the MREŽNIK project is to compile a free, monolingual, corpus-based, hypertext, easily searchable, online dictionary of Croatian standard language with three modules (for adult native speakers: 10,000 entries, for school children: 3000 entries, and for foreigners: 1000 entries). The dictionary entries will contain links to repositories which will be created as a part of this project (Linguistic Advice Repository: 300 entries, Conjunction Repository: all conjunctions, The Idiom Etymology Repository: 50 idioms, The Repository of Ethnics and Ktetics (place names, feminine and masculine names of the inhabitants and corresponding adjectives): 300 entries) as well as repositories which have already been compiled within other projects at the Institute of Croatian Language and Linguistics: The Verb Valence Repository, The Collocation Repository, The Croatian Terminology Repository (Struna), The Croatian Metaphor Repository, and the website Better in Croatian. The dictionary will be based on these two corpora: the Croatian Web Repository and the Croatian Web Corpus. The dictionary will be compiled using TLex. SketchEngine, a corpus manager and analysis program, and Tickbox Lexicography will be used to search the corpora and extract data from it. As a part of the project, a reverse dictionary will be compiled.

Keywords: e-lexicography; web dictionary; corpus-based dictionary; Croatian language; dictionary grammar

1. Introduction

The fact that Croatia is still one of the countries which do not have a publically-available online corpus-based dictionary of their national language compiled according to the rules of contemporary e-lexicography, or systematic research on e-lexicography, was the reason for starting a new project: Croatian Web Dictionary—MREŽNIK. The project started on the 1ˢᵗ March 2017 and the duration of the project is four years. The result of the MREŽNIK project will be a free, monolingual, hypertext, easily searchable, online dictionary of the Croatian standard language. This dictionary has three different modules: a dictionary for adult native speakers of Croatian, a dictionary for elementary school children, and a dictionary for foreigners. As we are still in the first half of the first year of the project, in this paper we will present the dictionary grammar and style manual for three different modules, which are being compiled at the moment, as well as connected databases and computer tools. Some of the connected databases have already been compiled while others will be compiled at the same time with the dictionary. Some of test definitions and lists of labels have also already been compiled, as well as the pilot reverse dictionary based on the pilot word-list. We will also compare MREŽNIK with *Wrječnik* and *Hrvatski jezični portal* and explain why MREŽNIK will

not be connected with these dictionaries.

2. Foreign E-dictionaries

In modeling the Croatian Web Dictionary many similar foreign dictionaries have been consulted, e.g. *elexiko* (http://www.owid.de/wb/elexiko/start.html) of the Institute of German Language, *Wielki słownik języka polskiego* (http://www.wsjp.pl/) of the Institute of Polish Language, *Swedish online dictionary* (http://spraakbanken.gu.se/karp), *Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart* (https://www.dwds.de/), *Algemeen Nederlands Woordenboek* (http://anw.inl.nl/), *Online-Bildwörterbuch* (http://www.bildwoerterbuch.com/), etc. In particular, *elexiko* was used as an inspiration for modeling some of the dictionary fields presented below.

3. Croatian E-dictionaries

While planning the dictionary grammar and fields, all existing Croatian printed dictionaries and e-dictionaries were consulted.[1] However, MREŽNIK will not be in any way connected with *Wječnik* or with *Hrvatski jezični portal* (Croatian Language Portal; HJP). The reasons for this are numerous: 1. While these two dictionaries are descriptive dictionaries of the Croatian language including dialects, MREŽNIK is a descriptive and prescriptive dictionary of Standard Croatian. 2. While MREŽNIK is corpus-based (see below) *Wječnik* and *Hrvatski jezični portal* are not. 4. HJP presents an online dictionary, which is the result of the collaboration between Novi Liber and Srce (http://hjp.znanje.hr/), today owned by the publishing house Znanje. It is an online version of *Hrvatski enciklopedijski rječnik* (*Croatian Encyclopedic Dictionary*), which was not compiled as an online dictionary, but is a printed dictionary published by the publishing house Novi Liber, and sold in the printed version for the last 15 years. This fact is the reason for many drawbacks of this dictionary. This online dictionary has relatively inefficiently interconnected entries, i.e. only links to other headwords in the etymological part of the entry, and is out of date as it has not been revised for a long time. 5. *Wječnik* is a *Wiktionary* project, a collaborative project, based on crowdsourcing to produce a free-content multilingual dictionary. It is a lexical project based on Wikipedia software (Wikimedia).

MREŽNIK, on the other hand, is a scientific project; the collaborators of MREŽNIK are experienced scientists and lexicographers. The dictionary is compiled taking into account semantic relations and the systematic nature of language. We will illustrate this using one simple example, comparing the first definition of seasons in all three dictionaries:

---

[1] For more on Croatian e-lexicography see Jermen et al. (2015); Štrkalj Despot & Möhrs (2015).

	Wječnik	**Croatian Language Portal**	**MREŽNIK**
zima winter	jedno od četiriju <u>godišnjih</u> <u>doba</u>, <u>kalendarski</u> (na <u>sjevernoj</u> <u>Zemljinoj</u> <u>polutci</u>) <u>traje</u> <u>od</u> 21. <u>odnosno</u> 22. prosinca <u>do</u> 21. <u>ožujka</u>, <u>dolazi</u> <u>između</u> <u>jeseni</u> i <u>proljeća</u>.	**a.** kalendarsko doba od 22. prosinca do 21. ožujka **b.** jedno od četiriju godišnjih doba, između jeseni i proljeća	godišnje doba koje najčešće na sjevernoj hemisferi počinje 22. prosinca i traje do 21. ožujka
proljeće spring	Jedno od četiriju godišnjih doba. Kalendarski traje od 22. ožujka do 22. lipnja	**a.** kalendarsko doba od 21. ožujka do 21. lipnja **b.** jedno od četiri godišnja doba (između zime i ljeta)	godišnje doba koje najčešće na sjevernoj hemisferi počinje 22. ožujka i traje do 21. lipnja
ljeto summer	Jedno od četiriju <u>godišnjih</u> <u>doba</u>. Kalendarski traje od 21. lipnja do 22. rujna.	kalendarsko doba od 21. lipnja do 22. rujna, jedno od četiri godišnja doba	godišnje doba koje najčešće na sjevernoj hemisferi počinje 22. lipnja i traje do 22. rujna
jesen autumn	godišnje <u>doba</u> koje traje od 23. rujna do 21. prosinca	**a.** kalendarsko doba između 23. rujna i 21. prosinca **b.** jedno od četiri godišnja doba	godišnje doba koje najčešće na sjevernoj hemisferi počinje 23. rujna i traje do 21. prosinca

Table 1: Comparison of the definitions of four seasons in three dictionaries

From Table 1, it is obvious that only MREŽNIK has all definitions structured in the same way. All definitions start with the same hypernym (godišnje doba – season) written in lowercase letters; definitions have the same syntactic structure and consist only of one sentence; and they give the same data, the date of the beginning and the end of the season. The analysis of semantic fields having more members would show even greater differences between dictionary definitions.

4. Corpus

The Croatian web dictionary MREŽNIK is based on two Croatian corpora: *Croatian Web Repository*, the corpus of the Institute of the Croatian Language and Linguistics

(http://riznica.ihjj.hr/index.hr.html) and the *Croatian Web Corpus* (http://nlp.ffzg.hr/resources/corpora/hrwac). These corpora are managed by the corpus tool Sketch Engine. The corpus is always checked by the lexicographer, as the dictionary is corpus-based and not corpus-driven. One of the reasons for this is: "For most native-speaker dictionaries, corpora are still inadequate in size and are also slightly out of date by the time they are available to lexicographers. So 'reading-and-marking' of the latest newspapers and magazines, and attentive listening to radio and television broadcasts, are still a necessity." (Brown 2006: 250–254). Although this may be less and less true for big languages like English, it is still true for Croatian.

The corpus-based principle was used in *elexiko* as Klosa points out: *elexiko* is basically corpus-based, i.e. there are no lexicographic entries in *elexiko* which do not come from the *elexiko*-Korpus and there is no information that is simply taken over from other dictionaries. (Klosa 2011: 16.) However, she also points out some negative aspects of this principle while some topics appear more often in the newspaper corpus. (Klosa 2011: 58.) and *elexiko*-Korpus is basically a newspaper corpus. Another reason for using the corpus-based and not corpus-driven approach is the normative aspect of MREŽNIK. The corpus will provide bases for creating the list of headwords, differentiating meanings, selecting derivatives, compounds and collocations, composing definitions, selecting or creating examples (depending on the module). Three different modules have three different approaches to the corpus as will be shown below.

5. Three Modules

MREŽNIK is a scientific dictionary which is also user-friendly and fulfils different needs of different user groups. Thus, the dictionary gives as the information the user needs and is connected to many other databases. A similar idea when speaking about *elexico* has been stated in Haß (2005: 3) as she points out that *elexico* can fulfill different user needs and interests and this approach would not be possible in a printed dictionary.

MODULES
module for adult native Croatian speakers – 10 000 entries
module for school children – 3000 entries
module for foreigners – 1000 entries

Table 2: Three modules of MREŽNIK with the number of entries

MREŽNIK consists of three separate modules which are connected by the fact that all given data is coordinated and synchronized. However, each module functions as a separate dictionary compiled for a different target group of users. The first module is a dictionary for adult native speakers of Croatian consisting of 10,000 entries. The second module is a dictionary for elementary school children consisting of 3000 entries,

and the third is a dictionary for foreigners consisting of 1000 entries. Each dictionary module has different dictionary grammar which is based on the specific needs of the dictionary user.

Different modules have a different approach to examples from the corpus. In the dictionary for adult users, the lexicographer will select the examples from the corpus. Each meaning and definition will have examples from the corpus. The approach to the corpus of three different modules is shown in Table 3:

Three modules	Module for adult native Croatian speakers	Module for elementary school children	Module for foreigners
Explanation	The headword will be a direct link to the corpus. For each meaning, examples will be taken from the corpus. These examples will be selected by the lexicographer. In addition, each headword will have a link to the corpus.	For each meaning very simple examples will be devised by the lexicographer.	For each meaning the example will be taken from the corpus and simplified by the lexicographer.
Example	Djelatnici Inspektorata Ministarstva zaštite okoliša i kriminalistička policija na zgarištu su u Parku prirode Kopački rit proveli više od deset sati kako bi se utvrdio uzrok požara koji je u nedjelju navečer poharao ovaj baranjski biser i to njegov najvredniji dio - poseban ZOO rezervat u kojemu se gnijezde rijetke i zaštićene vrste ptica.http://riznica.ihjj.hr/philocgi-bin/search3t?dbname=Cijelihr&word=ptica&OUTPUT	Ptica leži na jajima u gnijezdu.	U parku se gnijezde rijetke vrste ptica.

Table 3: Approach to the corpus in three different modules

6. Word List and the Corpus

Three different modules will have three different lists of headwords. The starting point for the word list for adult native speakers is the corpus from which the 10,000 most-frequent lemmas will be extracted. The words extracted from the corpus will be manually checked by dictionary editors, compared with the word list which has been compiled manually by the authors and editors of the dictionary and supplemented using the criteria of word formation and semantic fields. As the compilation of the corpus

extracted word list is as yet in progress, this will be illustrated from the word list that has been compiled manually.

autobus	autoportret	bacač
autobusni	autoput	bacil
autocesta	autor	baciti
autogram	autostop	bačva
automat	avantura	badem
automatski	avion	badminton
automehaničar	b	Badnjak
automobil	baba	badnji
automobilizam	babaroga	

Table 4: Extract from the pilot wordlist of the MREŽNIK project

This word list will be supplemented for example by the words *automehaničarka, automobilistički, bacačica.*

As there are no specialized corpora for elementary school children and foreigners, the list of headwords for these users has to be derived manually. Fortunately, some of the members of the MREŽNIK project have experience in writing lexicographic works for school children as some of them are authors of *Prvi školski rječnik hrvatskoga jezika* (The First Dictionary of the Croatian Language—Čilaš Šimpraga, Jojić & Lewis, 2008), *Školski rječnik hrvatskoga jezika* (School dictionary of the Croatian language—Birtić at al, 2012), and *Prvi školski pravopis* (First Orthographic Manual Hudeček, Jozić, Hudeček, Lewis & Mihaljević, 2016) and are the editors of *School portal*, which is one of the elements which will be connected with MREŽNIK (see below). A member of the MREŽNIK team works with foreigners learning Croatian in Croaticum at the Faculty of Humanities and Social Sciences and, on the basis of her experience and textbooks for foreigners learning Croatian, 1000 words for foreigners will be selected.

7. Dictionary Grammar

A dictionary is a highly structured document. The 'dictionary grammar' is at the center of the project. It names the different fields of information and says how they are to be nested and ordered, and which are obligatory and which are optional. (*Encyclopedia of Language & Linguistics* 2006: 783–793). Simultaneously with the extraction of the headwords, the editors of the dictionary are working on the 'dictionary grammar' of MREŽNIK. The idea is to have a three-module and three-dimensional dictionary. The

representation of the dictionary in each module will consist of fields giving basic data for each entry and links giving additional information which the user can see by clicking on the links. Dictionary grammar will be analyzed for each module separately.

7.1 Module for Adult Speakers

The dictionary grammar consists of these elements: accentuated headword (direct link to the type in the corpus), homonym mark, grammatical information, accentuated inflectional forms, link to inflectional forms, masculine/feminine pairs, perfect/imperfect pairs, cross-references to other entries, accentuated sub-entry, grammatical label, stylistic label, usage label, field label, differentiation of meaning, grammatical restriction, definition, examples from the corpus, link to collocations, link to pragmatic comments, link to semantic relations, phrase, idiom, word formation analysis of the headword, link to derivatives and compounds from the corpus. Semantic relations are divided into synonyms, antonyms, hyponyms and, co-hyponyms. Some elements re-occur as many times as needed, e.g. differentiation of meaning, definition, semantic relations, link to pragmatic explanation are given for the headword, phrase, and idiom. All the fields are optional, except the headword and the grammatical information. Dictionary grammar is shown in Table 10 in the Appendix.

The fields in the table will be illustrated by examples from different entries as it is impossible to find an entry containing all these elements: accentuated headword: **kȕća** (house), grammatical information: *im. m.*, accentuated forms <G kȕćē; *mn.* N kȕće, G k ćā>.

In the four-year duration of the project from each headword a link to all forms that will be automatically derived and manually checked will be attached, e.g.

	Singular	Plural
N	kuća	kuće
G	kuće	kuća
D	kući	kućama
A	kuću	kuće
V	kućo	kuće
L	kući	kućama
I	kućom	kućama

Table 5: Example of word forms for the headword *kuća* (house).

In the future the plan is to have all these forms accentuated as well, e.g.:

	Singular	Plural
N	kùća	kùće
G	kùće	k ćā
D	kùći	kùćama
A	kùću	kùće
V	kùćo	kùće
L	kùći	kùćama
I	kùćom	kùćama

Table 6: Example of accentuated word forms for the headword *kuća* (house).

This is not a problem with nouns, but there are very many verbal forms, and this will probably not be possible within the four-year duration of the project.

An important part of the dictionary are the masculine-feminine pairs, e.g. the masculine noun *učitelj* and the feminine noun *učiteljica* (teacher) will be interconnected:

> **ùčitelj** *im. m.*
>
> **učitèljica** *im. ž.*

The criterion for such interconnection is not only word-formation, but also semantics. So e.g. nouns *jelen* (deer) and *košuta* (doe) will also be connected:

> **jȅlēn** *im. m.*
>
> **kòšuta** *im. ž*

Such interconnection is also possible for phrases, e.g.:

> **medicinska sestra** (nurse feminine)
>
> **medicinski tehničar** (nurse masculine)

The connection to the pair will be a link if the pair has its own entry in the dictionary.

In the same way, perfect and imperfect verbs will be connected, e.g.:

ocijéniti *gl. svrš. prijel.* (evaluate)

ocjenjívati *gl. nesvrš. prijel.*

Headwords will be cross-referenced with the label *v.* (vidi – see) which directs a headword not belonging to the standard language to its standard equivalent. These words will usually also have a usage *label* and/or will be connected with linguistic advice.

Sub-entry is used for reflexive verbs which are analyzed under the main verbal entry, i.e.:

sèliti gl. *prijel./neprijel.* (move)

- **sèliti se** *povr.*

Grammatical restriction is used in the cases when a specific grammatical description applies only to a specific meaning, e.g.:

mȉš *im. m.* <G mȉša, A mȉša/mȉš, L mȉšu/mìšu; mn. N mȉševi, G mȉšēvā> 1. <A mȉša, L mȉšu> *zool.* 2. <A mȉš, L mìšu> *inform., tehn.* (mouse)

Each meaning has a definition. Definitions usually start with hypernyms which are links to the dictionary entry, e.g.:

mànastīr *im. m.* <G mànastīra, I mànastīrom; mn. N mànastīri, G mànastīrā> crkva … (monastery)

cȑkva im. ž. <G cȑkvē; mn. N cȑkve, G cȑkāvā/c kvā/cȑkvī> 1. Građevina … (church)

gràđevina *im. ž.* <G gràđevinē; mn. N gràđevine, G gràđevīnā> grad. 1. objekt … (building)

This is the reason why hyperonyms are not stated among semantic relations in the dictionary grammar.

Examples follow each meaning and are selected from the corpus as described above.

Special attention in MREŽNIK will be paid to collocations. Collocations are differentiated from phrases and idioms. They will be analyzed according to the model from *elexiko* and derived from the corpus via Word Sketches. In *elexiko* e.g. headwords *girl* and *boy* are analyzed according to these questions: *What are the characteristics of a X?, What does X do?, What happens to X?, Which themes are used with X?* (http://www.owid.de/wb/elexiko/gruppen/maedchen-junge.html).

Some entries will have a link to the pragmatic comment. Pragmatic comment will be given, e.g. with pronouns *ti* and *Vi* (you), greetings *dobar dan*, *dobro večer*, *dobro jutro*, *zdravo*, *bok,* etc.

Special attention will also be paid to semantic relations which will be attached to each meaning: synonyms, antonyms, and hyponyms, e.g.:

> **döbar** 1. koji ima pozitivne osobine ili poželjna svojstva ***sin.*** valjan; ***ant.*** loš 2. koji je onakav kakav treba biti, koji ispunjava očekivanja ***ant.*** loš 3. koji čini i želi dobro ***ant.*** zao, zločest 4. koji je ispravno utemeljen i logičan ***sin.*** pravi *razg.*, valjan; ***ant.*** loš 5. <neodr.; u im. funkciji> srednja školska ocjena označena s 3; ***sin.*** trojka razg. 6. <sup.> koji ima najpozitivnije osobine i najpoželjnija svojstva]; ***sin.*** (optimalan)

Three different classes of labels will be used as shown in Table 7.

Grammatical labels	Field labels	Usage labels
m. – muški rod male	anat. – anatomija anatomy	razg. razgovorno – colloquial
s. – srednji rod neuter	astr. – astronomija astronomy	reg. regionalizam regional
ž. – ženski rod female	astrol. – astrologija astrology	žarg. – žargonizam – jargon
pl. tantum – *pluralia tantum*	biol. – biologija biology	
sg. tantum – *singularia tantum*	bot. – botanika bothanics	
neprijel. – neprijelazni glagol intransitive verb	el. – elektrotehnika electrical engineering	
povr. – povratni glagol reflexive verb	farm. – farmacija pharmacy	
prijel. – prijelazni glagol – transitive verb...	fil. – filozofija philosophy	
	fiz. – fizika physics	
	fiziol. – fiziologija physiology	
	geol. – geologija geology...	

Table 7: Three different classes of labels in MREŽNIK

The dictionary entries contain links to repositories which have already been compiled within other projects conducted at the Institute of Croatian Language and Linguistics: Valence database e-Glava, Repository of Metaphors, Terminology database STRUNA , Better in Croatian and to databases which are created as a part of this project and compiled simultaneously with the dictionary. These repositories are: Linguistic Advice Repository, Conjunction Repository, Repository of Idioms, Repository of Ethnics and Ktetics, Male/female repository, and Pragmalinguistic repository.

7.2 Module for Elementary School Children

The second module of the project is the module for school children. The aim is to contribute to Croatian language learning in schools as it is evident that e-dictionaries have a much greater chance of being accepted in schools than classical dictionaries.[2] The dictionary grammar is shown in Table 11 in the Appendix.

This dictionary will consist of the headword with marked accentuation place (djev**o**jčica), some grammatical information (part of speech, gender), syllable marking (e.g. dje-voj-či-ca), simple definitions, examples written by the lexicographer, very few synonyms, collocations, idioms. In this module some of the entries will contain illustrations. The principles of using appropriate illustrations and their role in the understanding of semantic relations in language manuals for children have been given in Hudeček & Mihaljević (2015).[3]

Some dictionary entries contain links to simple language advice and explanation of idioms for school children in the repository Croatian in School (http://hrvatski.hr/).

The block of fields starting with differentiation of meanings re-occurs as many times as needed. All the fields are optional except the headword and the grammatical information.

7.3 Module for Foreigners

The third module of the project is the module for foreigners. This module will contain audio recording of the pronunciation of each headword. It will also provide information (pragmatic, cultural, collocations) useful to foreigners learning Croatian.[4] The dictionary grammar is shown in Table 12 in the Appendix.

The block of fields starting with differentiation of meanings, re-occurs as many times as needed. All the fields are optional except the headword and the grammatical information. Some entries will be linked to simplified language advice and explanation of idioms.

8. Reverse Dictionary

In the fourth year of the project, a reverse dictionary in which entries are alphabetized from the end is planned. This dictionary is very important for the analysis of word-formation. This dictionary will be based on the completed list of 10,000 words. At this phase of the project a pilot reverse dictionary based on the manually compiled word-list has already been compiled and is available to team members.

---

[2] The mariginal role of dictionaries in teaching German as a mother tongue in German schools is also mentioned by Töpel (2014: 291).

[3] Illustrations have also an important function in *elexiko*. Müller-Spitzer (2005: 212).

[4] Similar ideas appear in Möhrs (2014: 322).

baba	oporaba	skladba
visibaba	zloporaba	obradba
tužibaba	uporaba	preradba
koraba	zlouporaba	doradba
poraba	žaba	razradba

Table 8: Extract for the pilot reversed wordlist of the MREŽNIK project

9. Computer Tools

The two basic computer tools for the compilation of this three-module dictionary are SketchEngine, a corpus query system (loaded with corpora) to support the analysis of the language, and TLex, a dictionary writing system to support the preparation of the dictionary text. The dictionary will be compiled using TLex, a professional software application for compiling dictionaries. SketchEngine, a corpus manager and analysis program, will be used to search the corpora and extract data. SketchEngine can be used to retrieve the context in which a word is usually found using word sketches, grouping the strongest collocations into syntactic categories and finding adequate examples of lexemes and collocations. Data are selected from the corpus by simply ticking boxes inside a lexicographic interface (Tickbox Lexicography). Data selected in this way are automatically saved in TLex and other lexicographers can continue describing the lexemes on all other levels. For the morphological description, the morphological lexicon hrLex (http://nlp.ffzg.hr/resources/lexicons/hrlex/) will be used, the content of which will be adapted and connected with MREŽNIK. The data acquired from hrLex will be checked by the lexicographer and accentuated by the accentologist.

After completion of the dictionary entries, the data will be exported from TLex, in order to be used in the Web application, which will be developed for the dictionary and the CLARIN repository; a European research network working in the field of archiving and processing of language-related resources (https://www.clarin.eu/). MREŽNIK will in this way become available for use through web application. It will also be available for various purposes at the CLARIN repository.

10.   Conclusion

In the paper, the dictionary grammar of a three-module three-dimensional corpus-based dictionary is presented. As the project only began in March 2017, many elements are still being developed and programmed in TLex so the fields are presented in tables and not as screenshots of the program. Table 9 sums up the connection of three dictionary modules with special databases.

Dictionary module	Connected databases
dictionary for adult native speakers 10,000 entries	Linguistic Advice Repository http://jezicni-savjetnik.hr/
	Conjunction Repository
	Repository of Idioms
	Repository of Ethnics and Ktetics
	Portal Bolje je hrvatski Better in Croatian http://bolje.hr/
	Male/female portal
	Pragmalinguistic portal
	Repository of Metapors http://ihjj.hr/metafore/
	Valency database e-Glava: Baza hrvatskih glagolskih valencija http://ihjj.hr/projekt/baza-hrvatskih-glagolskih-valencija/27/
	Croatian Special Field Terminology http://struna.ihjj.hr/
dictionary for elementary school children 3000 entries	Croatian in schools – explanation of idioms and language advice for children http://hrvatski.hr/
dictionary for foreigners	explanation of idioms explanation of usage simple language advice

Table 9: Connection of three dictionary modules with special databases

11. Acknowledgements

12. References

Baza hrvatskih glagolskih valencija – GLAVA. Accessed at http://ihjj.hr/projekt/baza-hrvatskih-glagolskih-valencija/27/ (18 May 2017)

Birtić, M. & et al. (2012). *Školski rječnik hrvatskoga jezika.* Zagreb: Školska knjiga – Institut za hrvatski jezik i jezikoslovlje.

Bolje je hrvatski. Accessed at http://bolje.hr/ (18 May 2017)

Brown, K. (ed.) (2006). *Encyclopedia of Language & Linguistics* (second edition) Elsevier Ltd.

Čilaš Šimpraga, A.; Jojić, Lj. & Lewis, K. (2008). *Prvi školski rječnik hrvatskoga jezika.* Zagreb: Školska knjiga – Institut za hrvatski jezik i jezikoslovlje.

Haß, U. (2005). Das Bedeutungsspectrum. In U. Hass (ed.) Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz. Berlin/New York: De Gruyter, pp163-181.

Hrvatski jezični portal. Accessed at http://hjp.znanje.hr/ (30 April 2017)

Hrvatsko strukovno nazivlje – STRUNA. Accessed at http://struna.ihjj.hr/ (18 May 2017)

Hudeček, L. & Mihaljević, M. (2015). Dječji jezik i jezični priručnici za djecu in In L. Cvikić; B. Filipan-Žignić, I. Gruić, B. Majhut & L. Zergollern-Miletić (eds.) *Istraživanje paradigma djetinjstva i obrazovanja.* Zagreb: Učiteljski fakultet, pp. 46-67 Available at http://hrvatski.ihjj.hr/content/hr-nastava-1.pdf

Hudeček, L., Jozić, Ž., Lewis, K. & Mihaljević, M. (2016). *Prvi školski pravopis.* Zagreb: Institut za hrvatski jezik i jezikoslovlje.

Jermen, N., Kraus, C. & Starčević Stančić, I. (2015). Lexicography and Encyclopaedistics in the Digital Environment in K. Anderson, L. Duranti, R.Jaworski, H. Stančić, S. Seljan & V. Mateljan. *Infuture 2015. The Future of Information Sciences. E-Institutions Openness, Accessibility, and Preservation,* Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, pp. 65-75.

Jezični savjetnik Accessed at http://jezicni-savjetnik.hr/ (18 May 2017)

Klosa, A. (ed.). (2011). elexiko. *Erfahrungsberichte aus der lexikographischen Praxis eines Internetwörterbuchs.* Tübingen: Narr. Verlag.

Möhrs, Ch. (2014): Landeskundliche Wortschatzübungen auf der Basis von Kollokationen. Zur Nutzung von *elexiko* für Deutschlehrende. In A. Klosa (ed): *Themenheft "Dateninterpretation und -präsentation in Onlinewörterbüchern am Beispiel von elexiko".* Deutsche Sprache 4/2014, pp. 309-324.

Müller-Spitzer, C. (2005). Illustrationen in elexiko. In U. Haß (ed.). *Grundfragen der elektronischen Lexikographie. elexiko - das Online-Informationssystem zum deutschen Wortschatz.* Berlin/New York: De Gruyter, pp. 204-226.

OWID. Accessed at http://www.owid.de/wb/elexiko/gruppen/maedchen-junge.html (15 May 2017)

Repozitorij metafora. Accessed at http://ihjj.hr/metafore/ (18 May 2017)

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

Štrkalj Despot, K. (2014). Croatian Metaphor Repository. *Proceedings of the 2nd COST ENeL Working Group 3 („Innovative e-dictionaries")* Meeting: „Workflow of Corpus-based Lexicography". Available at http://www.elexicography.eu/wp-content/uploads/2014/07/Strkalj-Despot_2014_COST_Bolzano.pdf

Töpel, A. (2014). Wörterbücher im muttersprachlichen Deutschunterricht der Sekundarstufe II – Zur Nutzbarkeit der Bedeutungserläuterungen von *elexiko*. In: A. Klosa (ed.), *Themenheft "Dateninterpretation und -präsentation in Onlinewörterbüchern am Beispiel von elexiko*. Deutsche Sprache 4/2014, pp. 291-308.

Wječnik. Accessed at https://hr.wiktionary.org/wiki/Glavna_stranica (30 April 2017)

13. Appendix

Field	Sub-field and/or comment
accentuated headword	headword is a direct link to the type in the corpus
homonym mark	
grammatical information (part of speech code, e.g. im. m. noun, masculine)	
accentuated inflectional forms (selected forms)	
link to the inflectional forms (all forms)	
masculine/feminine pair	
perfect/imperfect pair	
cross-reference to another entry (*v.* see)	
accentuated sub-entry - reflexive verb	
grammatical label *povr.* - reflexive	
style and usage label	
field label	
differentiation of meanings 1., 2., 3. …	
stylistic label	
field label	
grammatical restriction	
definition	
examples from the corpus	
link to collocations	
link to pragmatic comment	
link to semantic relations	synonyms

	antonyms
	hyponyms
	co-hyponyms
phrase	
style label	
field label	
definition	
examples from the corpus	
link to collocations	
link to pragmatic comments	
link to semantic relations	synonyms
	antonyms
	hyponyms
	co-hyponyms
idiom	
link to the explanation of the idiom	
definition	
examples from the corpus	
link to semantic relations	synonyms, antonyms
link to pragmatic information	
word formation analysis of the headword	
link to derivatives and compounds from the corpus	

Table 10: Dictionary fields in the module for adult native speakers

Field	Sub-field and/or comment
headword with marked place of the accent	link to the audio recording of the pronunciation
homonym mark	
grammatical information (full words)	
masculine/feminine pair	
cross-reference to another entry	
inflectional forms	
differentiation of meanings 1., 2., 3. …	
stylistic label	
field label	
grammatical restriction	
definition	
examples	
collocations	
usage	
semantic relations	synonyms
	antonyms
phrase	
style label	
field label	
definition of meaning	
collocations	list of most frequent collocations
link to pragmatic comments	

link to semantic relations	synonyms
	antonyms
Idiom	link to the explanation of idioms
definition	
semantic relations	synonyms
	antonyms
headword with marked place of the accent	link to the audio recording of the pronunciation
homonym mark	
grammatical information (full words)	
division into syllables	
masculine/feminine pair	
cross-reference to another entry	
inflectional forms	
subentry reflexive verb	
differentiation of meanings	
grammatical restriction	
definition	
examples	most common collocations
synonyms – introduced into the definition by the formula: the same meaning have the words...	
antonyms – introduced into the definition by the formula: the opposite meaning have the words	
phrase	
definition	definition contains information on style and field (in spoken language, when speaking to friends, in mathematics...) as well as on

	synonyms and antonyms
examples	
synonyms – introduced into the definition by the formula: the same meaning have the words…	
antonyms – introduced into the definition by the formula: the opposite meaning have the words	
idiom	link to the explanation of the meaning of idioms for children
definition	
synonyms – introduced by the formula: the same meaning have the words…	
antonyms – introduced into the definition by the formula: the opposite meaning have the words	

Table 11: Dictionary fields in the module for school children

Field	Sub-field and/or comment
headword with marked place of the accent	link to the audio recording of the pronunciation
homonym mark	
grammatical information (full words)	
masculine/feminine pair	
cross-reference to another entry	
inflectional forms	
differentiation of meanings 1., 2., 3. …	
stylistic label	
field label	
grammatical restriction	
definition	

examples		
collocations		
usage		
semantic relations	synonyms	
	antonyms	
phrase		
style label		
field label		
definition		
collocations	list of most frequent collocations	
link to pragmatic comments		
link to semantic relations	synonyms	
	antonyms	
idiom	link to the explanation of idioms	
definition		
semantic relations	synonyms	
	antonyms	

Table 12: Dictionary fields in the module for foreigners