

Croatian Web Dictionary – Mrežnik vs. Croatian Linguistic Terminology – Jena

Lana Hudeček

Institute of Croatian Language and Linguistics, Zagreb, Croatia
lhudecek@ihjj.hr

Milica Mihaljević

Institute of Croatian Language and Linguistics, Zagreb, Croatia
mmihalj@ihjj.hr

Summary

The Croatian Web Dictionary – Mrežnik is a four-year project which started on the 1st of March 2017 and the duration of the project is four years. The main result of the project will be a free, monolingual, hypertext online dictionary consisting of three modules (the module for adult native speakers – 10,000 entries, the module for children aged 6 to 10 – 3,000 entries, and the module for non-native speakers learning Croatian – 1,000 entries). Mrežnik is based on two Croatian web corpora. Croatian Linguistic Terminology – Jena is a new terminological project conducted within the Struna program. The project started on the 24th of May 2019 and lasts until the 23rd of November 2020. The main result of the project will be a multilingual database consisting of 1,500 entries. As a specialized corpus of Croatian linguistic terminology doesn't exist, it is compiled in parallel with the database. Although Mrežnik and Jena differ in their basic goals and approach; one is monolingual and general and the other is multilingual and specialized (terminological), one is compiled from the existing corpora and the other is compiled in parallel with the corpus, they have two important meeting points: 1. General linguistic terminology is presented in Mrežnik (mostly but not exclusively in the module for adult native speakers) and 2. Within the Mrežnik project, Glossary of E-lexicographic terminology is compiled. These four parameters will be compared: 1. wordlist/termlist, 2. relation to the corpus, 3. giving normative information, 4. entry structure. The compilation process and the structure of entries for the same headword will be compared and the important similarities, as well as differences, will be shown. The reason for this comparison is that the two projects are conducted at the same time and strongly influence each other in many aspects.

Key words: Mrežnik, Croatian Linguistic Terminology – Jena, linguistic terminology, e-lexicography

Introduction

The project *Croatian Web Dictionary – Mrežnik*¹ aims at creating a free, monolingual, easily searchable hypertext online dictionary of standard Croatian. It will be the first web-born dictionary of the Croatian language. Entries, sub-entries, and meanings will be interconnected, as well as linked to entries in databases created within the framework of the project in parallel with the creation of the dictionary (language advice database, conjunction database with description of groups of conjunctions and their modifiers, database of explanations of the origin of idioms, database of ethnics and ktetics), as well as databases being created by project collaborators or other Institute members within the framework of other projects.² *Mrežnik* consists of three modules: the module for adult native speakers of Croatian which will have 10,000 entries, the module for school children which will have 3,000 entries, and the module for non-native speakers which will have 1,000 entries). The dictionary is written in the *TLex* program, which has been adapted to the needs of the project. The main goals of the project are: 1. to create the three dictionary modules, 2. to connect the dictionary with the databases created in parallel with the dictionary, 3. to connect the dictionary with other web sources currently being compiled at the Institute of Croatian Language and Linguistics, 4. to compile a reversed dictionary based on the *Mrežnik* wordlist, 5. to write a monograph on *Mrežnik*. The project

¹ More on *Mrežnik* see in Hudeček 2018; Hudeček, Mihaljević 2017a, 2017b; Hudeček, Mihaljević, 2018a, 2018b.

² More on this topic see in Hudeček, Mihaljević 2019a.

started on the 1st of March 2017, so at the moment we are in the second half of the project and more than 5,000 entries have been compiled.

The project *Croatian Linguistic Terminology – Jena* is conducted within the *Struna* program. *Struna* is a database of Croatian Special Field Terminology³ financed by the Croatian Science Foundation. *Jena* is a year-and-a-half project which started on the 24th of May 2019. The main goals of the project are: 1. to compile 1,500 entries with definitions, synonyms, antonyms, hyponyms and equivalents in English, German, French, Russian, and Swedish in the *Struna* database, 2. to collect works on linguistic terminology and present them on the *Jena* website (ihjj.hr/jena/), 3. to write a monograph on Croatian linguistic terminology. At the moment 1035 entries have been entered into the database. However, it is important to note that both *Croatian Web Dictionary – Mrežnik* and *Jena* are conceived as a dynamic dictionary/databases that will be further compiled and edited even after the formal end of the project funding and will not reach their full extent if they do not continue to grow and become an everlasting project of the Institute.

Hypothesis and reason for comparison

The hypothesis of this paper is that although a terminological database obviously differs from a general e-dictionary there are many similarities from which both projects can profit.

The reason for such a comparison is that these two projects are conducted at the same time in the same institution, the head of *Mrežnik* (Lana Hudeček) is the collaborator of *Jena* while the head of *Jena* (Milica Mihaljević) is the collaborator of *Mrežnik*. Thus some results of one project can be applied to the other and vice versa. In the comparison all linguistic terms which appear in *Mrežnik*⁴ and the *Glossary of e-lexicographic terminology* compiled within the *Mrežnik* project are taken into account. As both projects are in progress the instructions for the respective team of lexicographers and terminographers can be modified according to new results. The basic points of comparison are: 1. the ways of compiling wordlist/termlist, 2. the approach to the corpus, 3. the approach to normativity, 4. the structure of dictionary entries.

Wordlist vs. termlist

To compile the *Mrežnik* wordlist the frequency lists of *hrWaC* (first 12,000 words) and the *Hrvatska jezična riznica* (first 10,000 words) were overlapped, all words present only in *Hrvatska jezična riznica* and not present in *hrWaC* were extracted, their frequency was multiplied by four, and they were added to the shared list. This wordlist (first 8,000 entries) was juxtaposed with two separate wordlists: the wordlist for the module for children (which was excerpted 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 taken from a list in textbooks for non-native speakers, to ensure that words found in both these lists (which partially overlap) appear in the list for adult native speakers. This wordlist was supplemented with male/female (in cooperation with the project *Male and Female in the Croatian Language*) and aspectual pairs, possessive and descriptive adjectives, adverbs derived from adjectives from the list, nouns ending in *-ost* derived from adjectives from the list, numerous grammatical and semantic groups, etc. This resulted in a wordlist of 10,000 words 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 and we first began compiling the entries for words from this list in order to make processing for the module for children as compatible as possible with that for adult native speakers (Hudeček, Mihaljević, 2018b).

³ The Institute of Croatian Language and Linguistics was chosen to serve as the national coordinator. The objective of the program in a broader sense is to lay the foundation for the development of national terminology policy, to establish various forms of more structured education in this field, and to intensify long-term cooperation with national and international academic and other institutions dealing with different aspects of terminology work, with the Croatian Standards Institute and with other interested parties. Within the program, a terminology database has been developed to store and terminographically manage standardized and harmonized Croatian terms from various subject fields and their equivalents in English and other languages. Experts from eighteen domains have so far joined the program with the aim of standardizing the terminology of their respective disciplines. <http://struna.ihjj.hr/en/about/>.

⁴ Of course some linguistic terms have a non-linguistic meaning which occurs in *Mrežnik* and doesn't occur in *Jena* but this was not the subject of our analysis.

The *Jena* termlist consisting of 1,500 terms was compiled by project collaborators divided into workgroups by subject fields: basic linguistic terminology, cognitive linguistics, contact linguistics, dialectology, e-lexicography and corpus linguistics, generative linguistics, glottodidactics, language history, lexicography, lexicology, onomastics, orthography, phraseology, pragmatics, sociolinguistics, terminology, translation theory, valency theory. Table 1 shows a small extract from the termlist divided by subject fields.

Table 1. Extraction of *Jena* termlist by specialized subject fields

1.1 Generative linguistics	1.2 Cognitive linguistics	1.3 Phraseology	1.4 Translation theory	1.5 Language history
1.6 E-jezik 1.7 generativna gramatika 1.8 I-jezik 1.9 jezična moć 1.10 jezična sposobnost 1.11 jezična uporaba 1.12 logički problem jezičnoga usvajanja 1.13 negativni dokazi 1.14 objasnidbena prikladnost 1.15 opisna prikladnost 1.16 oskudnost poticaja 1.17 pozitivni dokazi	1.18 apsolutni prostorni sustav 1.19 apstrahiranje 1.20 argumentna struktura 1.21 asimetrija izvornoga i ciljnoga okvira 1.22 automatsko prepoznavanje metafora 1.23 autonomistički gramatički pristup 1.24 konceptualne integracije 1.25 ciljna domena 1.26 dinamični razvojni model 1.27 dinamika sile 1.28 diskursna analiza vodena metaforom	1.29 frazeologija u užemu smislu 1.30 frazeologija u širem smislu 1.31 paremiologija 1.32 krilatologija 1.33 zoonimna frazeologija 1.34 somatska frazeologija 1.35 internacionalna frazeologija 1.36 nacionalna frazeologija 1.37 posuđena frazeologija 1.38 arhaična frazeologija 1.39 dijalektna frazeologija 1.40 regionalna frazeologija 1.41 frazeološki obrat	1.42 automatsko prevođenje 1.43 doslovno prevođenje 1.44 književno prevođenje 1.45 komunikacijski model prevođenja 1.46 ljudsko prevođenje 1.47 pismeno prevođenje 1.48 simultano prevođenje 1.49 slobodno prevođenje 1.50 strojno prevođenje računalo prevođenje 1.51 traduktologija 1.52 univerzalni prevodilac	1.53 starohrvatski jezik 1.54 filološke škole 1.55 zagrebačka filološka škola 1.56 zadarska filološka škola 1.57 riječka filološka škola 1.58 škola hrvatskih vukovaca 1.59 štokavski hrvatski književni jezik 1.60 čakavski hrvatski književni jezik 1.61 kajkavski hrvatski književni jezik 1.62 ozaljski književno-jezični krug

These terms will not appear as headwords of entries or subentries in *Mrežnik*. However, terms belonging to basic linguistic terminology and orthography will appear in *Mrežnik* as well as in *Jena*. Some terms belonging to basic linguistic terminology are shown in the text below. Figure 1 shows an extraction of the wordlist in *Jena*.

infinitiv	jezikoslovje	filologija	opće jezikoslovje (lingvistika)
imperfekt	jezikoslovje	filologija	opće jezikoslovje (lingvistika)
imperativ	jezikoslovje	filologija	opće jezikoslovje (lingvistika)
imenica	jezikoslovje	filologija	opće jezikoslovje (lingvistika)

Figure 1. General linguistic terminology from *Jena*

These linguistic terms will also be entries in *Mrežnik*.

Corpus-based

Both *Mrežnik* and *Jena* are corpus-based, and not corpus-driven. This means that the corpus and all data extracted from it serve only as guidelines. The *Glossary of E-lexicographic Terminology* 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. The reason for this approach is that neither of the corpora on which *Mrežnik* is based (*Croatian Web Repository* online corpus (riznica.ihjj.hr/index.hr.html) and *Croatian web corpus – hrWaC* (nlp.ffzg.hr/resources/corpora/hrwac/) are representative of the Croatian language (*hrWaC* is primarily based on the colloquial and journalist style and *Croatian Web Repository* on the literary style), they are not corpora of the standard language nor are they balanced corpora. It follows that, in composing an entry, lexicographers can add meanings to a particular entry or to the collocation field even if they do not appear in the corpus.

Data extraction from the corpora for *Mrežnik* as well as for *Jena* is performed with the SketchEngine web tool, which allows the display of lexeme context through WordSketches, the most common collocations sorted into syntactic categories and the discovery of good examples of word usage or collocations. After lexicographic processing of *Mrežnik* is completed, the data will be exported from TLex to 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.

Jena is based on the corpus *Jezikoslovlje* composed specially for the needs of the project. It consists of a corpus of linguistic papers and monographs compiled under SketchEngine. The *Jena* corpus which has been compiled by project members and collaborators is the corpus of standard language (in the field of linguistics) but it is as yet not representative enough. Moreover, on many modern linguistics topics there are not many texts in Croatian and many Croatian terms have to be coined by the authors (specialists of the particular linguistic field) themselves. From this corpus a term list has been compiled which contrasted the words appearing in the corpus with the words from *hrWaC*. The basic term list is still the one created by project collaborators but it will be checked against the one created by Sketch Engine from the corpus so *Jena* will also be corpus-based. The *Jena* corpus is also helpful when creating definitions and deciding on the normative status of synonymous words.

Normativity

The Croatian Web Dictionary – Mrežnik is a normative dictionary and *Jena* is a normative terminological database. The normative nature of *Mrežnik* is apparent in the following: 1. the selection of entry-words, 2. giving normative advice in all three modules, 3. the selection of forms acceptable by the standard language norm in the grammatical block, 4. the selection of examples (the dictionary collaborators try to select examples with no language errors while examples with language errors are edited), 5. the accentuation of entry-words and forms in the grammatical block according to the standard language norm.

The most important normative aspect of *Jena* is differentiating between the preferred, allowed, non-preferred, obsolete, and jargon terms (as will be shown in the examples below). If needed normative advice is given in the field note, e.g. why the preferred term is *točka sa zarezom* and not *točka-zarez* as shown in table 2.

Word entries vs. term entries

Two important meeting points of *Mrežnik* and *Jena* are 1. General linguistic and orthographic terminology is presented in *Mrežnik* and 2. Within the *Mrežnik* project, a *Glossary of E-lexicographic Terminology* is compiled (ihjj.hr/mreznik/page/pojmovnik/6/).

General orthographic terminology in *Mrežnik* and *Jena*

Table 2 illustrates the structure of the entries *točka* (period) and *točka sa zarezom* (semi-colon) in *Jena* and compares them to the respective entry or subentry in *Mrežnik*:

Table 2. Entries *točka* (period) and *točka sa zarezom* (semicolon) in *Jena* and *Mrežnik*

1.63 <i>Jena</i>	1.64 <i>Mrežnik</i>
<p>1.65 točka 1.66 unesen: 01.08.2019, 20:57 faza obradbe: urednik pregledao status naziva: preporučeni naziv definicija: pravopisni znak koji stoji na kraju rečenice te iza kratica i rednih brojeva vrelo definicije: Jozić, Željko i dr. 2013. Hrvatski pravopis. Institut za hrvatski jezik i jezikoslovlje. Zagreb. područje: jezikoslovlje potpodručje: pravopis jezična odrednica: imenica rod: ženski broj: jednina 1.67 istovrijednica - engleski: period; full stop 1.68 njemački: Punkt 1.69 francuski: point 1.70 ruski: точка 1.71 švedski: punkt 1.72 simbol: . poveznica: http://pravopis.hr/pravilo/tocka/55/</p>	<p>1.73 točka 1.74 <i>pravop.</i> Točka je pravopisni znak (.) koji stoji na kraju rečenice te iza kratica i rednih brojeva. 1.75 - Definicija mora počinjati malim slovom i nema točku na kraju. 1.76 - Argument koji govori u prilog tomu da se parataksička rečenica ne razlikuje samo formalno od dviju rečenica, tj. da se ne može promatrati kao dvije rečenice koje su odijeljene točkom, odnosno koje se od dvorečeničnog ustrojstva razlikuju samo formalno. 1.77 Koordinacija: točka i crtica, točka i uskličnik, točka i zarez 1.78 Poveznica Hrvatski pravopis: http://pravopis.hr/pravilo/tocka/55/</p>
<p>1.79 Comparison: <i>Mrežnik</i> and <i>Jena</i> have <i>točka</i> as a headword of the entry. <i>Jena</i> has only one meaning of <i>točka</i>, while <i>Mrežnik</i> has many meanings only one of which is the meaning in the orthographic sense. <i>Mrežnik</i> also has many subentries of <i>točka</i> one of which is <i>točka sa zarezom</i> (semicolon). They have similar definitions, but while <i>Jena</i> gives the source of the definition <i>Mrežnik</i> gives examples and collocations (coordination). These examples are taken from the <i>Jena</i> corpus as it was difficult to find adequate examples from two corpora on which <i>Mrežnik</i> is primarily based. Both are connected to the same paragraph from <i>Croatian Orthography Manual</i>. In <i>Jena</i> equivalents in English, German, French, Russian, and Swedish are given. While in <i>Mrežnik</i> the sign (.) is a part of the definition given in brackets due to the very strict structure of the terminological database (brackets cannot be included in the definition) in <i>Jena</i> it is included in a special symbol field.</p>	
<p>1.80 točka sa zarezom 1.81 unesen: 01.08.2019, 20:58 faza obradbe: urednik pregledao status naziva: preporučeni naziv definicija: pravopisni znak koji se piše pri jačemu odvajanju od onoga koje označuje zarez, a slabijemu od onoga koje označuje točka vrelo definicije: Jozić, Željko i dr. 2013. <i>Hrvatski pravopis</i>. Institut za hrvatski jezik i jezikoslovlje. Zagreb. područje: jezikoslovlje potpodručje: pravopis jezična odrednica: višerječni naziv 1.82 istovrijednica - engleski: semicolon 1.83 njemački: Semikolon 1.84 francuski: point-virgule 1.85 ruski: точка с запятой 1.86 švedski: semikolon 1.87 simbol: ; 1.88 napomena: U hrvatskome pravopisnom nazivlju u istome se značenju upotrebljavaju nazivi <i>točka-zarez</i> i <i>točka sa zarezom</i>. Budući da u nazivlju istoznačenice nisu poželjne, a polysloženice se ne uklapaju u strukturu hrvatskoga jezika te ih je, kad je to moguće, bolje zamijeniti istoznačnim nazivom drukčije strukture, prednost se daje nazivu <i>točka sa zarezom</i>. 1.89 nepreporučeni naziv: točka-zarez poveznica: http://pravopis.hr/pravilo/tocka-sa-zarezom/62/</p>	<p>1.90 točka sa zarezom pravop. 1.91 Točka sa zarezom pravopisni je znak (;) koji se piše pri jačemu odvajanju od onoga koje označuje zarez, a slabijemu od onoga koje označuje točka 1.92 - Definicije su u kurzivu i međusobno su odvojene zarezom, a sinonim koji nije u kurzivu odvojen je točkom sa zarezom. 1.93 - Veoma je često u engleskome tekstu uz veliko slovo u okomitome nabranjanju i točka sa zarezom. 1.94 •normativna napomena: U hrvatskome pravopisnom nazivlju u istome se značenju upotrebljavaju nazivi <i>točka-zarez</i> i <i>točka sa zarezom</i>. Budući da u nazivlju istoznačenice nisu poželjne, a polysloženice se ne uklapaju u strukturu hrvatskoga jezika te ih je, kad je to moguće, bolje zamijeniti istoznačnim nazivom drukčije strukture, prednost se daje nazivu <i>točka sa zarezom</i>. 1.95 Mrtvi sinonim: točka-zarez 1.96 Poveznica: <i>Hrvatski pravopis:</i> http://pravopis.hr/pravilo/tocka-sa-zarezom/62/ 1.97</p>


1.98 **Comparison:** In *Jena* *točka sa zarezom* (semicolon) is an entry while in *Mrežnik* it is a subentry of the entry *točka*. The reason for this is that a multiword term has the same terminological status as a single word term. They have similar definitions but only *Jena* states that the source of the definition is the *Croatian Orthographic Manual*. *Mrežnik* gives examples from the corpus while *Jena* has no examples. *Mrežnik* gives another synonymous term *točka-zarez* as a „dead synonym” which means it is not an entry in *Mrežnik*. In *Jena* there are no synonymous entries and *točka-zarez* is given as a non-preferred term. Both sources give the same explanation why *točka sa zarezom* is preferred to *točka-zarez* but this explanation occurs in the note field in *Jena* and in the field normative advice in *Mrežnik*. However, both are connected to the paragraph on semicolon from the *Croatian Orthographic Manual*. Both *Mrežnik* and *Jena* state that this term belongs to the field of orthography. In *Jena* equivalents in English, German, French, Russian, and Swedish are given. Examples in *Mrežnik* are taken from the *Jena* korpus, as it was difficult to find an adequate example in the two corpora on which *Mrežnik* is primarily based.

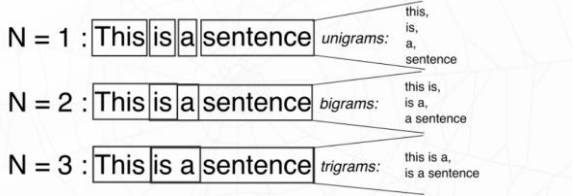
Similar results could be shown when comparing some other entries of general linguistic terms, e.g. *imenica* (noun), *padež* (case), *sklonidba* (declension), *sintaksa* (syntax).

Glossary of e-lexicography and Jena

Glossary of E-lexicography compiled within the *Mrežnik* project and in collaboration with the *Jena* project consists of names and terms relevant for e-lexicography. This *Glossary* is an important source for *Jena* as from it most of the terms (not names) are taken over into the *Jena* database. In table 3 the comparisons of the entries *odostražni rječnik* (reversed dictionary) and *n-gram* is shown in *Jena* and the *Glossary of e-lexicography*.

Table 3. Entries *odostražni rječnik* (reverse dictionary) and *n-gram* (n-gram) in *Jena* and *Glossary of e-lexicography*

1.99 <i>Jena</i>	1.100 <i>Glossary</i>
<p>1.101 odostražni rječnik 1.102 unesen: 04.08.2019, 18:12 faza obradbe: urednik uređuje status naziva: preporučeni naziv definicija: rječnik u kojemu su riječi abecedirane od kraja riječi vrelo definicije: Pojmovnik, Mrežnik. http://ihjj.hr/mreznik/page/pojmovnik/6/ područje: jezikoslovlje potpodručje: e-leksikografija i korpusno jezikoslovlje dopušteni naziv: odostražnik jezična odrednica: višerječni naziv istovrijednica - engleski: reverse dictionary 1.103 njemački: rüchläufiges Wörterbuch 1.104 francuski: dictionnaire inverse 1.105 ruski: Обратный словарь 1.106 švedski: baklängesordbok; finalalfabetisk ordbok 1.107 poveznica: https://borna12.gitlab.io/odostraznji/ 1.108 napomena: Nazivu <i>odostražni rječnik</i> daje se prednost pred nazivom <i>odostražnik</i> zbog sustavnoga odnosa s nazivljem ostalih vrsta <i>rječnika</i> (<i>opći rječnik, posebni rječnik, abecedni rječnik, normativni rječnik, deskriptivni rječnik, terminološki rječnik</i> itd.)</p>	<p>1.109 odostražni rječnik (engl. reverse dictionary) rječnik u kojemu su riječi abecedirane od kraja riječi Rückläufiges Wörterbuch des Serbokroatischen (1965. – 1967.) mrežno je dostupan na https://www.uibk.ac.at/slalistik/institut/matesic.html. Demoinačica odostražnoga rječnika naziva za vršitelje/vršiteljice radnje (https://borna12.gitlab.io/odostraznji-mz/, izradio Josip Mihaljević): 1.110</p>  <p>1.111</p> <p>1.112</p> <p>1.113 odostražnik v. odostražni rječnik</p>
<p>1.114 Comparison: <i>Jena</i> has only a definition and additional information is given in the note section. In the note the reasons for selecting <i>odostražni rječnik</i> as the preferred term are explained. In <i>Jena</i> equivalents in English, German, French, and Russian are given.</p>	
<p>1.115 n-gram 1.116 unesen: 04.08.2019, 16:41 faza obradbe: urednik pregledao status naziva: preporučeni naziv definicija: sekvencija određene duljine koju sačinjavaju znakovi ili riječi koje se</p>	<p>1.118 n-gram sekvencija određene duljine koju sačinjavaju znakovi ili riječi koje se pojavljuju unutar teksta; pri radu s korpusima n-grami se odnose na sekvencije riječi; unigram je jedna riječ, bigram je sekvencija od dvije riječi, trigram je sekvencija od tri riječi itd.</p>

pojavljuju unutar teksta korpusa vrelo definicije: Pojmovnik, Mrežnik. http://ihj.hr/mreznik/page/pojmovnik/6/ . područje: jezikoslovlje potpodručje: e-leksikografija i korpusno jezikoslovlje podređeni pojam: bigram; trigram; unigram jezična odrednica: imenica rod: muški broj: jednina istovrijednica - engleski: n-gram istovrijednica - njemački: N-Gramme istovrijednica - francuski: n-gramme istovrijednica - ruski: N-грамма 1.117 švedski: n.gram	1.119 
1.120 Comparison: In <i>Jena</i> unigram, bigram, and trigram are added as subordinate terms and they have a separate definition. In the <i>Glossary</i> they are explained under <i>n-gram</i> but also have separate definitions in the glossary. The <i>Glossary</i> is added as a source in <i>Jena</i> . In <i>Jena</i> equivalents in English, German, French, Russian, and Swedish are given.	

Jena vs. *Mrežnik*: entry structure

From the general structure of *Struna* these fields have been activated for *Jena* (Table 4).

Table 4. Fields in *Jena*

1.121 Field	1.122 Explanation
1.123 entry word	1.124 can be a multiword entry
1.125 grammatical data	1.126 only word class and gender and number for nouns
1.127 definition	1.128 with <i>genus proximum</i> and <i>differentia specifica</i> , not a whole sentence, starts with the small letter and does not end with a period; one term can have only one definition.
1.129 field, discipline	1.130 subfields of linguistics, e.g. generative linguistics, cognitive linguistics, pragmatics, etc.
1.131 synonyms	1.132 divided into preferred terms, allowed terms, depicted terms, obsolete terms, and jargon terms
1.133 antonyms (added for the purpose of this project)	1.134 defined in <i>Jena</i> by a similar definition
1.135 subordinate terms	1.136 defined in <i>Jena</i>
1.137 source of the definition	1.138 if the definition was taken over from a source and not formed by the compiler the source should be stated
1.139 equivalents in English, Russian, French, German	1.140 written (or checked) by experts for each of the languages
1.141 abbreviation or acronym	1.142 given if any
1.143 connected to other sources	1.144 the entries are often connected to the <i>Croatian Orthographic Manual</i> or <i>Croatian School Grammar</i> , sometimes they are connected to other sources, e.g. articles in the journal <i>Hrvatski jezik</i>
1.145 note	1.146 in the note relevant additional information is given
1.147 phase in the compilation process are recorded	1.148 different phases are: written by the author, checked by the editor, checked by the terminologist, checked by the language editor, finished

The diagram of the structure of *Mrežnik* is described in detail in Hudeček, Mihaljević, 2019a. In table 5 the main differences between the structure of *Jena* and *Mrežnik* are shown:

Table 5. Differences between the structure of *Jena* and *Mrežnik*

1.149 <i>Jena</i>	1.150 <i>Mrežnik</i>
1.151 headword – can be multiword, not accentuated	1.152 no multiword headwords, headwords and forms are accentuated
1.153 grammatical data, word class or multiword – for nouns gender and number	1.154 gives much more grammatical data as well as accentuated forms
1.155 definition – one headword has only one definition, if needed the same headword has multiple entries	1.156 a headword can have multiple senses and definitions. very often the same entry has many meanings and only one belongs to the field of linguistics (e.g. <i>crtica</i> , <i>točka</i> , <i>atribut</i>) and sometimes the same word has more than one meaning in the field of linguistics (e.g. <i>pravopis</i> , <i>rječnik</i> , <i>fonologija</i>)

1.157 field, discipline, has a list of disciplines and sub-disciplines	1.158 differentiates between linguistics, grammar and orthography, doesn't differentiate between sub-disciplines
1.159 synonyms – differentiates between the preferred term, allowed term, non-preferred term, obsolete term, and jargon term	1.160 gives synonyms, differentiates between synonyms that are dictionary entries and that are not (synonyms and dead synonyms), doesn't differentiate between the status of synonyms
1.161 antonyms – all given antonyms are dictionary entries	1.162 differentiates between antonyms which are dictionary entries and which are not (dead antonyms)
1.163 subordinate terms – a very important field for building the terminological system	1.164 sometimes gives subordinate terms
1.165 source of the definition	1.166 doesn't give data on the source of definition
1.167 equivalents in English, Russian, French, German, and Swedish	1.168 doesn't have equivalents in foreign languages
1.169 abbreviation or acronym given in a separate field	1.170 abbreviations and acronyms are given as synonyms and not in a separate field
1.171 symbols are given in a separate field	1.172 symbols, if needed, are included in the definition
1.173 connected to other sources – mostly connected to <i>Croatian School Grammar</i> , articles from the journal <i>Hrvatski jezik</i> and <i>Croatian Orthography Manual</i> ⁵	1.174 connected to a number of sources ⁶
1.175 additional information given in the note	1.176 differentiates between the pragmatic note and the normative note (language advice)
1.177 records the phase in the compilation process	1.178 doesn't state explicitly the phase in the compilation process
1.179 context	1.180 examples and collocations

An important difference between *Mrežnik* and *Jena* is the approach to collocations. In *Mrežnik* they have a separate field where they are introduced by questions, e.g. *What is xxx like?*, *What does xx do?*, *What can we do with xxx?*, *Coordination*, *What is mentioned in connection with xxx?* In *Jena* there is no special collocation field and they can be either introduced as subordinate terms which then have separate entries, explained in the note or ignored.

Results of the comparison

The results of the comparison prove the hypothesis that two such projects as *Jena* and *Mrežnik* can be compared and that they can mutually profit from each other and such a comparison. The results of the comparison are shown in table 6.

Table 6. Comparison of *Jena* and *Mrežnik*

Points of comparison	<i>Mrežnik</i>	<i>Jena</i>	Benefits
wordlist/termlist	terms extracted from the two corpora and supplemented by the lexicographers	termlist compiled by field specialists and supplemented by corpus data	<i>Mrežnik</i> wordlist was checked against the <i>Jena</i> termlist and all terms belonging to general vocabulary are included into <i>Mrežnik</i> . <i>Jena</i> termlist was supplemented by the terms from the <i>Glossary of e-lexicography</i>
corpus	corpus-based	corpus-based (but the role of the corpus is not as important as in <i>Mrežnik</i>)	for linguistic terms in <i>Mrežnik</i> examples can be taken from the <i>Jena</i> corpus
normativity	normative dictionary, normativity expressed in the normative advice note	differentiates between preferred, allowed, non-preferred, obsolete, and jargon terms	normative advice given in <i>Mrežnik</i> and in <i>Jena</i> is always the same (it can be explained differently), i.e. once the normative status of a term is determined within one project the same normative status is given to the term in the other

⁵ See in Hudeček, Mihaljević 2017c; Jozić et al., 2013.

⁶ See Hudeček and Mihaljević in print.

			project
the structure of dictionary entries	The structure of dictionary entries is determined by the project head in collaboration with project members	the structure is only partially flexible as it is limited by the <i>Struna</i> database	projects have different structure but all relevant data determined by the research on either of the projects could be included into the other project; the experience on <i>Mrežnik</i> resulted in the inclusion of the field antonyms in <i>Jena</i>

Conclusion

The aim of this paper is to show how two projects conducted at the same time in the same institution can influence each other and how experience with one project as well as the corpus and data from one project can help the other project. As we started working on this paper at the very beginning of *Jena*, this gave us the perfect opportunity to test the hypotheses that the work on a terminological project can help the work on a general e-lexicographic project as well as benefit from it. Despite many differences, *Mrežnik* and *Jena* are closely connected as the *Glossary of E-lexicography* serves as one of the sources for *Jena* and as some definitions of general linguistic and orthographic terms from *Mrežnik* can serve as a starting point for composing the entry of the same headword in *Jena* (as shown above). We also found the corpus compiled for *Jena* very useful for collocations and examples in *Mrežnik* as with some frequently used words which have many meanings in the general language (as for example *točka* shown above as well as *čestica*, *prilog*, *prijedlog*) it was difficult to find the adequate example in the general corpus.

At the end of *Jena*, entries from *Mrežnik* will be connected with external links to entries in *Jena* as *Mrežnik* already connected to some other finished *Struna* projects. Some other projects conducted at the Institute are connected to one or both of the analysed projects, e.g. the project *Orthographic Manual of Religious Terminology* is connected with *Mrežnik* as religious terms also form an important part of *Mrežnik*. The project *Male and Female in the Croatian Language* is strongly connected to both of the analysed projects as in *Mrežnik* each noun denoting a male person is connected to the noun denoting a female person⁷ and in *Jena* special attention is paid to professional nouns in the field of linguistics (male and female).

Acknowledgments

This paper is written within the research project *Croatian Web Dictionary – MREŽNIK* (IP-2016-06-2141) and the project *Croatian Linguistic Terminology – Jena* (Struna-2017-09-05), both financed by the Croatian Science Foundation.

References

- Hudeček, L. (2018). Izazovi leksikografske obrade u jednojezičnome mrežnom rječniku (na primjeru Hrvatskoga mrežnog rječnika – Mrežnika). // *Visnyk of Lviv University: Series Philology* 69, 29-38
- Hudeček, L., Mihaljević, M. (2017a). A New Project – Croatian Web Dictionary MREŽNIK. // *The Future of Information Sciences. INFUTURE2017, Integrating ICT in Society / Atanassova, I. et al. (eds.)*. Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, 205-213
- Hudeček, L., Mihaljević, M. (2017b). Hrvatski mrežni rječnik – Mrežnik. // *Hrvatski jezik* 4, 4, 1-7
- Hudeček, L., Mihaljević, M. (2017c). Školska gramatika hrvatskoga jezika. Zagreb: Institut za hrvatski jezik i jezikoslovlje
- Hudeček, L., Mihaljević, M. (2018a). Croatian Web Dictionary Mrežnik: One year later – What is different? // *Proceedings of the Conference on Language Technologies & Digital Humanities / Fišer, D., Pančur, A. (eds.)*. Ljubljana: Oddelek za prevajalstvo, Inštitut za novejšo zgodovino, 106-113
- Hudeček, L., Mihaljević, M. (2018b). Hrvatski mrežni rječnik – Mrežnik: Upute za obrađivače. <http://ihjj.hr/mreznik/uploads/upute.pdf> (27.10.2019)
- Hudeček, L., Mihaljević, M. (2019a). Croatian Web Dictionary – Mrežnik – Linking with Other Language Resources. // *Electronic lexicography in the 21st century – Proceedings of the eLex 2019 conference / Kosem, I. et al. (eds.)*. Leiden: Lexical Computing CZ s.r.o, 72-98
- Hudeček, L., Mihaljević, M. (2019b). Profesijski nazivi u hrvatskoj e-terminografiji i e-leksikografiji. // *Studia lexicographica* 13, 24, 75-95
- Institut za hrvatski jezik i jezikoslovlje. *Glossary of E-lexicographic Terminology*. 14.10.2019. <http://ihjj.hr/mreznik/page/pojmovnik/6/> (30.08.2019)
- Institut za hrvatski jezik i jezikoslovlje. (2019). Hrvatsko jezikoslovno nazivlje – Jena. 23.10.2019. <http://ihjj.hr/jena/> (30.08.2019)

⁷ More on the topic see in Hudeček, Mihaljević, 2019b.

- Institut za hrvatski jezik i jezikoslovlje. (2010). Hrvatsko strukovno nazivlje – Struna. 16.09.2010. <http://struna.ihjj.hr/> (30.08.2019)
- Institut za hrvatski jezik i jezikoslovlje (2017). Muško i žensko u hrvatskome jeziku. 1.11.2017. <http://ihjj.hr/projekt/musko-i-zensko-u-hrvatskome-jeziku/72/> (20.08.2019)
- Institut za hrvatski jezik i jezikoslovlje. Religijski pravopis. 5.11.2015. <http://ihjj.hr/projekt/religijski-pravopis/23/> (20.08.2019)
- Jozić, Ž. et al. (2013). Hrvatski pravopis. Zagreb: Institut za hrvatski jezik i jezikoslovlje