EDictionary: the Good, the Bad and the Ugly

Janjić, Marijana; Poljak, Dario; Kocijan, Kristina

eDictionary: the Good, the Bad and the Ugly

Marijana Janjić, Dario Poljak, Kristina Kocijan

Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Ivana Lučića 3, 10 000 Zagreb (Croatia)
E-mail: marijanajanjic@yahoo.com, poljak5@gmail.com, krkocijan@ffzg.hr

Abstract

On its own, learning a new language is an inherently daunting task. Combined with lacking or simply non-existent language resources, the task itself seems almost impossible. For some languages, this scarcity of available resources is even more obvious and further complicates the issue.
With an interdisciplinary approach, a team of linguists, language teachers, information scientists, and students themselves undertook a task of developing a learner's dictionary of Asian languages. With a great deal of care and discussion, an online e-dictionary was chosen as a platform for its ease of use, accessibility, and expandability, in lieu of a traditional printed dictionary.
Since *eDictionary* is built as a website, it is established as a platform, agnostic and available to everyone with Internet access. Furthermore, such a design allows a link to resources hosted on other web portals. To that end, cooperation was initiated with *Croatian Language Portal* and their Croatian dictionary with the aim of hyperlinking all of our Croatian lemmas to their word definitions. With the added benefits of giving users the ability to request new resources while keeping track of the request internally and allowing the updates of the whole language database seamlessly, the proposed solution to *eDictionary* provides user engagement and continuous integration that should benefit us all.

Keywords: e-dictionary; learner's dictionary; user engagement; Asian languages; Croatian

1. Introduction

History suggests that dictionaries in the form of word lists are a very old invention. From ancient Akkadian times to today, dictionaries represent important and valuable achievements in various cultures. Over the centuries, man has created different types of dictionaries with different purposes in mind. Among them is the dictionary aimed at learners of foreign languages, which is a version of a learner's dictionary - smaller in size than a general-purpose dictionary, with elements that enhance the learner's knowledge and skills in the target language. Modern times and the rapid advance of technology have made the existence of an online dictionary possible, not as a replacement for a printed dictionary, but as an addition to it.

Aside from the monolingual dictionaries, the new technology has also been used to develop student-oriented online dictionaries. A number of e-dictionaries have become popular for several reasons (Heuberger, 2016: 41): 1. the **size** of an e-dictionary is greater than that of a printed dictionary, and it is easier to alter; 2. e-dictionaries, unlike printed dictionaries, can **function** in more than one direction, i.e. any of the

languages included in it can be taken as a starting point (L1) and learning goal (L2); and 3. the possibility to include multimedia **features** (ex. recorded pronunciations, pictures).

Some good examples of online dictionaries include a dictionary for students of translation studies with a focus on Internet-related vocabulary (Alipour, Robichaud & L'Homme, 2015), a dictionary for language learners and other users (Deksne et al., 2013), and a dictionary for learners of Spanish (Renau & Battaner, 2011). Another interesting project is a multilingual lexicographic project for immigrants (Vacalopoulou & Efthimiou, 2015).

After giving a short overview of related work, the structure of the paper will take a closer look at the main idea behind this project, taking into account its upsides (the Good), downsides (the Bad), but also those aspects that could have been done better (the Ugly). Before the concluding remarks, a short analysis of the analytics will be provided.

2. Related work

Looking from the perspective of Croatian students, the search for web-based dictionary resources of Asian languages with Croatian as either the source or target language is like a scene from *Mission Impossible*. Most of the available resources have English as a link-language to the meaning of words from the Asian continent. If we were to operate under the assumption that all students know English well (and very well), and that this should not be considered an obstacle in using it to learn a third language, as different as any of the Asian languages, then things are all well and we can conclude our paper at this point.

However, this is not a valid assumption to make. Not all Croatian students have the same knowledge of English when starting university. Also, some language nuances are surely lost in translation, and even more so if they need to make their journey via multiple language groups (Slavic -> Germanic -> Asian and back). The lack of resources in one's native language puts an additional burden on the student, as it forces them to become a learner of not just one foreign language, but two - the link language, as well as the target language. As our experience in learning and teaching Hindi and Sanskrit in Croatia shows, there are students who come into the classroom equipped with not just dictionaries that include entries in the target language and English, but also with English – Croatian and Croatian – English dictionaries. This means that they are familiar with English to some level, and that they are simultaneously tackling two foreign languages at different levels.

Our goal of building Croatian language resources for the benefit of students in Croatia stems from the question as to whether students would be more efficient and successful in mastering the target language if having to master the link language was removed from the equation. The overview of available literature on the use of dictionaries and

other linguistic resources in a foreign language classroom suggests that authors and teachers assume the presence and availability of foreign language learning resources in the students' native language. According to some, the importance of native language resources is particularly high at the beginning stages of learning a new language, as the role of context is negligible at that point (Pavičić Takač, 2008; Summers, 1988). This is precisely the situation that Croatian students face when they opt to learn one of the Asian languages. Thus, our decision to focus on the use of dictionaries in foreign language teaching and learning was supported by two facts: a) acquisition of new vocabulary presents an important part of language learning, especially at the beginner level; and b) some students have to overcome a considerable obstacle, which is mastering a link language.

Multiple experiments in a Croatian context (Dovedan et al., 2002; Družijanić Hajdarević et al., 2006; Lauc et al., 2006; Librenjak et al., 2012; Janjić et al., 2016a; Librenjak et al., 2016c) have reported on how well language learning and digital resources go together i.e. the learners were mostly positive about technology usage, which led to greater motivation and consequently to more frequent usage of resources, so resulting in better language acquisition and greater retention. Hence, it seemed reasonable to assume that an e-dictionary would be better received by students than its printed counterpart. The following section gives more information on the *eDictionary* project and its intended users.

3. The *eDictionary* Project

The idea of building the *eDictionary*[1] of Asian languages for the Croatian users emerged naturally during the work on the *MemAzija* project. The aim of the *MemAzija* project was to test the influence of technology in learning Asian languages (Librenjak et al., 2016c; Janjić et al., 2016a; 2016b; 2017b). In order to test its hypothesis, the research team developed a number of Croatian language resources for learning four Asian languages: Hindi, Korean, Japanese and Sanskrit. From that point on, it did not take long to see that a learner's dictionary aimed at learning Asian languages was long overdue.

The primary reason for building the dictionary was the lack of similar resources in Croatian. As a result, students used the available e-dictionaries that translated Asian languages to English, and vice versa. During the *MemAzija* project, the research team realized how useful e-tools were to new generations of students, as they used them frequently in order to learn new languages or further improve their language skills.

The dictionary was built mainly for Croatian students studying one of the included Asian languages. But as it turns out, students are a heterogeneous group, and therefore the focus was further narrowed down to those students just beginning their studies, as

[1] Dictionary is available at: http://erjecnik.ffzg.hr/.

it was deemed that the resources in Croatian would be most useful to them. This decision was beneficial for our project in two ways. Firstly, it allowed us to provide students with a tool that would let them study a new foreign language with more ease, as they would be able to focus fully on acquiring just the target language vocabulary. Secondly, it served as a clear starting point for what could have been a broad and aimless project. As there is no Croatian dictionary, or more specifically no e-dictionary, which targets students and combines different Asian languages, it seemed reasonable to start with a smaller project i.e. a dictionary for A1 – B1 learners and go on from there.

This focus has had an effect on the dictionary design and its information architecture in several different aspects:

a) **dictionary mode**, i.e. choice between printed and online form,
b) **the choice of lemmas** included in the dictionary, and
c) **the structure of lemmas**.

For a more in-depth look at the choice and structure of lemmas included in the dictionary, please refer to Section 5.

Prior to building *eDictionary*, the research team had taken steps which were considered very important for both the end and front design of the dictionary, i.e. for the database and administrative dashboard design, as well as the user interface design. These steps included the **analysis** of existing e-dictionaries, **consultation** with lexicography experts, and a **survey**[2] of students' opinions regarding the preferable form and structure of an e-dictionary. All steps were equally useful to the research team. The last one, however, proved to be crucial as it showed what the primary user demographic considered important, relevant and beneficial for a learning tool to have.

Issues that students considered relevant when it came to the use of e-dictionaries for learning were helpful for determining what an e-dictionary should and/or should not include. The five most prominent of those issues were:

a) different rules for typing in Asian alphabets often require downloading various additional programs;
b) lack of off-line availability;
c) lack of compatibility with non-desktop devices (tablet or mobile);
d) no available Croatian translation;
e) direct translations lack examples of usage.

[2] The survey was conducted in 2016 during the *MemAzija* project and it involved 82 learners of Asian languages, with 72 female and 11 male students. The larger group, 59 of them, spoke Croatian as L1, while other participants were native speakers of Serbian, Montenegrian, Bosnian, Slovenian, etc. In this paper, we will refer to this survey as the MAP Survey (*MemAzija Project Survey).*

Although there were some students who reported feeling content with the quality of the e-dictionaries they normally use, the majority reported feeling the opposite. Most students reported a lack of additional information (gender, part of speech, etc.) which they deem important for a dictionary aimed at students, an absence of particular context or even worse – the complete absence of some words. Another criticism that students had was that the dictionaries rarely function well in both directions (L1 – L2, L2 – L1). They also expressed a very clear dislike for advertisements on some websites that host e-dictionaries. Another criticism to note is that they pointed out that those dictionaries that translate one word into a number of L1 words make it difficult for them to discern the primary meaning of the word. And finally, according to students' opinions, e-dictionaries should be closer to paper dictionaries in the sense that words with similar spelling and roots - words that would be found next to each other in a printed dictionary, should also appear together in an e-dictionary.

4. Introducing the Platform

When building a web platform there is much to take into consideration: from laying out the database and data flow to the sort of visual representation of elements on the page that would suit the user's needs. As Heid et al. (2013:271) point out: "We have passed the stage of putting paper dictionaries on computer or simply designing electronic dictionaries in the same way as paper dictionaries". Taking that into account, as well as the fact that this dictionary has a notably niche audience, we decided to adopt an architect's approach to building it in order to satisfy our users' needs. In other words, every feature was thoroughly planned out from the ground up. This approach has allowed us to future-proof the application by putting minimal constraints on adding new features or more languages.

Right at the start, it was obvious that one of *eDictionary's* main features would have to be two-level expandability including a) **depth**, allowing for new words to be added to each existing language, but also b) **width**, allowing completely new language additions. To tackle that, we have designed the dictionary with Croatian words as meta words in a pivot table applicable to all languages. Another planned feature relying on that design element would allow users to compare their query in multiple languages at once, a feature specifically targeted towards students studying Sanskrit and Hindi.

In order to achieve our goals in a timely manner, we have designed a relational database model with Model-View-Controller (MVC) architecture in mind, and tried to delegate most of the heavy lifting to technology. According to Majeed and Rauf (2018) MVC provides three types of classes:

A. **Model:** Model classes are used to implement the logic of data domains. These classes are used to retrieve, insert or update the data into the database associated with our application.
B. **View**: Views are used to prepare the application interface through which users

interact with the application.

C. **Controller**: Controller classes are used to respond to and perform user-requested actions. These classes work with model classes and select the appropriate view that should be displayed to the user according to their requests.

Thanks to such a clear division between the MVC layers we were able to effectively break down the development requirements. That has in turn allowed us to focus completely on creating the first usable versions of *eDictionary* and getting it tested by students themselves. Because *eDictionary* uses a pivot language, the database must contain transfer tables with the pivot language and every other language. This kind of relational model requires smooth data manipulation using models and MVC architecture (Janjić et al., 2017a). Relying on open source technologies, we have opted for Laravel on the back-end and jQuery on the front-end, which has in turn allowed us to focus on the user experience. Additionally, *eDictionary* is entirely hosted on faculty servers, i.e. all the documentation, codebase, and the complete language database are securely backed up. All of that combined with the fact that technologies used in development of *eDictionary* are well-established and widespread, means that there are no technical obstacles for further development of the project.

When it came to user experience, the term "accessibility" came up most often. We wanted to focus on three device types for optimal accessibility: mobile phones, tablets and computers (Figure 1). Due to the sheer volume of data on display to our users, we had to ensure adequate accessibility for the smallest of devices from the very start. Using responsive design-driven methodology, we have managed to scale our design down to resolutions of 480x960 pixels, while still retaining all the features of the page.

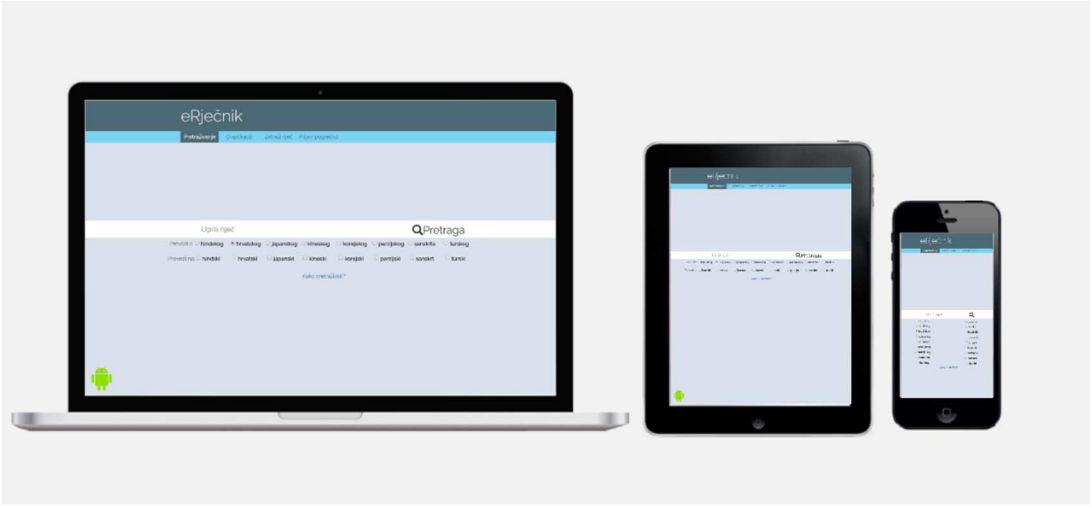

Figure 1: The *eDictionary* website as it appears on different resolution screens.

Further user experience improvements had to be made on the data querying front. Since we started working on the platform at the end of 2016, we had to take into consideration other non-dictionary web platforms which mainly use asynchronous calls

to the web server to deliver the data to the user without having to refresh the page. Thanks to our selection of development technologies, it was just a matter of agreeing on the data structure being sent to and from the server (Janjić et al., 2017a). In more technical terms, that means that we have opened up an API (Application Platform Interface) on our back-end Laravel server to receive data in JSON (JavaScript Object Notation) format via AJAX (Asynchronous JavaScript) calls and return the results in the same manner. That way we have achieved smooth and seamless data transmission between the user and the web server, therefore making *eDictionary* a full-fledged web application as defined by Paulson (2005).

Taking full advantage of the fact that the *eDictionary* is a website, all Croatian entries were made into hyperlinks. That means that each word in Croatian was connected to the well-established *Croatian Language Portal* (HJP) website, which already provides a single-language dictionary functionality for Croatian. This way, we made it possible for users to easily access any additional data they might find relevant without overwhelming them during the process (Heid et al., 2013).

Furthermore, due to the versatility of web 2.0 technologies *eDictionary's* expansion roadmap is not 'set in stone', but is rather expanding according to user needs. The way this works is through a *request system* that we have implemented, which makes it possible for users to request the needed resources for certain languages, as a team of language teachers or linguists develops requested resources for the next update of the internal database.

An internal monitoring tool was also prepared to provide administrators with analytics data on searched words, requested examples and reported errors, more or less covering the core functionality of *eDictionary*. Additionally, we are also using Google Analytics for data on user demographics, retention and bounce rate, as well as the type of platform users are accessing *eDictionary* on.

We have tried to cover as many fronts as possible while creating a web application that is both useful and not inherently limited in scope like a traditional dictionary. With prolonged use and administration, we have added new languages, new words and cooperated with domain experts to polish the core language learning functionalities.

4.1 Similar projects

Even though there are, at the time of writing, very few Croatian printed dictionaries targeting Asian languages[3], we can still examine how other similar Croatian projects were developed. In this chapter, we will evaluate four Croatian web portals that serve as different types of e-dictionaries. The four web portals discussed are: the *Croatian*

[3] Croatian-Japanese dictionary (2006) and Croatian-Turkish dictionary (2014).

Language Portal dictionary, which helped us with Croatian lemmas for *eDictionary*, *Croatian Encyclopedia*, *eGlava Online Valency Dictionary* and *Struna - Croatian National Termbank*. We are particularly interested in the technical implementations used for the four dictionaries, i.e. how they overcame some hurdles that we also faced, which features our dictionaries share and whether we are missing some features others consider crucial.

We chose these four projects in particular because of the shared similarity in niche target audiences, our affiliation with them (namely *Croatian Language Portal*), the fact that they came to be as a result of primarily academic efforts, and how well known they are.

4.1.1 Croatian Language Portal

Croatian Language Portal (cro. *Hrvatski jezični portal* or *HJP*) is a monolingual dictionary targeted at Croatian. It has emerged from collaboration between the publishing company *Novi Liber* and the *University Computing Centre – SRCE*. Armed with the prolific publication history of quality Croatian dictionaries by *Novi Liber*, the developers could easily construct a rich and detailed user interface for the dictionary.

Users of the *Portal* are given a plethora of information upon searching for a word, beginning with the word and its grammatical data, and followed by derived forms of the word in all cases/tenses/numbers, word definition(s) as found in printed dictionaries, in some cases even some example phrases, syntagmas, phraseology, onomastics, etymology and possibly even more. There is also a permalink feature for all searched words, which came in rather handy when we were connecting *eDictionary* Croatian entries with the existing definitions in *HJP*.

The *HJP* single language dictionary focuses on one thing and does it well, displaying all available language data for queried words. It is a synchronously loading site with a simple and straightforward design which translates well to mobile devices and computers alike.

The only downside to their design, one that we ran into during our own development process as well, is that it displays all of the language features even when there is no data associated with the searched word (for example, the title “Onomastics” appears to the user regardless of whether there is data to be shown or not). That, however, is almost a non-issue in contrast to the amount of presented and actually available information, since the user experience is not hampered in the least by this design “flaw”.

4.1.2 Croatian Encyclopedia

Croatian Encyclopedia is, as the name suggests, an encyclopaedia of the Croatian language. Developed by the Miroslav Krleža Department of Lexicography, it is presented as a single language dictionary web application that focuses on content presentation and professional explanations. The website offers a deep search functionality that goes not only through the lemma itself, but also through the explanations for all the occurrences of searched term.

Search results are colour-coded depending on where in the lemma or explanation the searched term was found. It can range from a direct hit, represented by dark red, when the searched term is present as a singular explained lemma. But it can also be found as part of the explanation and is then bright red. It should be noted that direct hits usually lead directly to the explanation, but other search results can also be accessed by clicking on "*Search further"* (cro. "*Traži dalje*"). We have found that colour-coding search results is a design element that may prove useful to our own application, especially for showing search results depending on target language(s). We will thus consider adding this upgrade in the future expansion of *eDictionary.*

The webpage is well designed for single-handed use on mobile devices, with a pop-up menu available on the bottom of the page and the action button positioned on its right side. This is another feature which is well thought out and will surely influence any future design revisions of *eDictionary.*

All explanations also serve as jumping-off points for further research, as some of the words are also hyperlinks to other lemmas and their explanations. This is an approach that, similar to our own, provides a natural bridge to even more relevant information on the subject, as proposed by Heid et al. (2013).

During our research for this paper, the only issue we encountered was slow page performance, with wait times for search results of up to 30 seconds. We did not, however, inspect this matter further since this can be attributed to many factors that are not directly controlled by the maintainers of the *Croatian Encyclopedia.* Still, we believe this to be an important issue, since wait times have been proven to be quite an important factor for user retention rate.

4.1.3 e*Glava Online Valency Dictionary*

eGlava Online Valency Dictionary is an online valency dictionary of Croatian verbs, developed within the project "*Valency Database of Croatian verbs*" at the Institute of Croatian Language and Linguistics. It contains valency descriptions for 900 verbs specifically built for linguists, teachers and students of Croatian language (Baza hrvatskih glagolskih valencija, 2019).

The website's sole purpose is listing all of the verb valences available in the database. There is no conventional search functionality through an input form, but rather an alphabetized list that can be filtered out by clicking on a specific letter. It is accompanied by an effective, albeit simple mobile design which keeps all website features accessible on all device types.

Some of the problems we encountered during research were mostly to do with misplaced links (i.e. some clickable elements throw the user back to the homepage), but otherwise the data it holds is presented exquisitely and in great detail. Furthermore, the site is also available in English, which is a feature still missing from *eDictionary,* but something that we strongly consider adding since we hope that our target group might also include non-native speakers of Croatian language in the future.

4.1.4 *Struna – Croatian National Termbank*

The last of the websites that we will discuss in this segment was also created under the leadership of the Institute for Croatian language. *Struna* is a website that focuses on standardized Croatian terminology for all professional domains.

Even though it is similar to the previously mentioned *eGlava Verb Valency Dictionary* in its narrow field of interest, the difference in website functionality is quite apparent. It offers both simple and advanced search options, mixes in attachments for certain defined terms similar to the *Croatian Encyclopaedia* website, and offers origin of the source for all defined terms.

Even though the content side of the website is meticulously crafted, there are some technical issues present that hamper the user experience. At the heart of said technical issues is the option to view the page in English – a feature that would be immensely useful, if only it were functional. Instead, what happens when the option is selected is that it breaks most of the hyperlinks on the website and instead returns *404 error pages* to user queries. There is also no responsiveness to speak of, so mobile use is strenuous at best.

But it is worth mentioning that both *eGlava* and *Struna* projects will be included in the ongoing *Mrežnik – Croatian Online Dictionary* (*Mrežnik – Hrvatski Mrežni Rječnik*) project (Hudeček & Mihaljević, 2017). That way, the content of both platforms will be unified and presented through a similar user interface.

After this analysis of projects similar to our own, we can conclude that we all had similar problems that were handled in similar ways – no matter the solution, the main focus was always on the content rather than the platform. This should not come as a surprise since the linguistic substance is the main reason users are visiting these sites, and in that regard all of them are very well executed. In the following section we will discuss the architecture of our own content.

5. The Soft Side of *eDictionary* – the Content

The target audience of *eDictionary* are both Croatian and Asian students learning Asian languages and Croatian, respectively. Because of this, Croatian was used as a source language, providing within *eDictionary* resources for six Asian languages (Hindi, Japanese, Chinese, Korean, Persian and Sanskrit) in varying degrees of fidelity. At the most, the dictionary is supposed to provide several key attributes for mastering a language. Among these attributes are translation, transliteration to Latin alphabet, grammatical notation and examples of usage. In cases where not all attributes are available at the time of the query, the base information always includes a translation and the link to the definition of the queried term in Croatian.

According to the MAP Survey of students' needs, a perfect e-dictionary for learning a new language would consist of lemmas that include a number of elements that we list here in the order of how many students selected them, starting with the most common one:

a) translation
b) grammatical information
c) pronunciation
d) examples of usage, phrases or sentences
e) visual representation
f) links to other resources, such as a lexicon or encyclopaedia that incorporates more elaborate definitions of particular cultural elements, products, ideas, etc.; links to other bilingual e-dictionaries with more elaborate lemma structures or to other monolingual dictionaries with more information
g) orientation regarding the level at which a learner is supposed to master a particular entry (ex. A1 or B1).

We have tried to include as many of the listed elements as we could at the time of building *eDictionary*, while still maintaining the capability to include all of the suggested elements at some point in the future, depending on the availability of funds. The financial side of the project determined at an early stage that the pronunciation could not be included in *eDictionary* from the start due to high production costs. The same was concluded for visual representation.

Regarding the examples, however, the decision was made to provide them for a number of lemmas, with an open invitation extended to students and other learners and teachers of Asian languages in Croatia to send in their own examples. Their validity would then be evaluated by our project experts for each language and included in *eDictionary* if deemed valid. This decision was made with the intention of opening bidirectional communication between users and the authors. This would effectively result in expansion of the authors' roles, since they would now also serve as dictionary administrators as well, which could be seen as an opportunity for new classroom

activities where students are encouraged to look for new examples or new words that would be useful to them in their own studies.

The first version of *eDictionary* consisted of 5,953 Croatian entries. Currently, as a result of newly added lemmas, this number has increased to 6,172. However, this number is not evenly distributed among all included languages. The languages with the most entries at the moment are Hindi (2,232) and Japanese (1,330), while Persian has the least (156). In-between these two groups are Sanskrit (1,028), Chinese (762) and Korean (668). Some of the words are unique to one Asian language, while some exist in more than one. All listed languages can be compared among themselves, which learners of similar languages, like Hindi and Sanskrit, could potentially find useful.

The lemmas included in the first version of *eDictionary* are based on the learning/teaching programs used in Croatia, as well as the authors' experience as teachers and learners of the included Asian languages. With every new word request from the user, *eDictionary* becomes that much better of a learning tool that mirrors not only the teachers' perspective, but the students' as well. Hence, the decision about which lemmas to include has been greatly affected by practical experience and focused on the learner's perspective.

This, however, is not true for Sanskrit. Sanskrit is one language that stands out in *eDictionary* in terms of the methodology used for choosing the initial set of words. For all modern Asian languages, the vocabulary is similar to any other learner's dictionary in the world. But, since Sanskrit is not used for everyday communication, the same rules do not apply here. It is a classical language that Croatian students study at the Indology Department so that they could successfully read and analyse old Sanskrit documents (literature, philosophy, etc.). In other words, Sanskrit vocabulary does not contain those elements that would allow one to easily order from a restaurant, but it does contain elements relevant for the study of ancient Sanskrit texts. For that reason, the Sanskrit lemmas found in *eDictionary* are based on the frequency of their usage in texts that students often work with as they learn to master Sanskrit. The frequencies were based on the lexicographic work by Oliver Hellwig (2016) and can be found online as *Digital Corpus of Sanskrit*, hosted by the Cluster of Excellence "Asia and Europe in a Global Context" research facility.

eDictionary can be used in several directions, i.e. the Asian languages which were considered target languages (L2) in the project can also be used as L1 languages, i.e. source languages. In that sense, *eDictionary* can also be useful to students or other interested parties for a comparative search - for example from Hindi to other languages present in the database: Croatian, Chinese, Sanskrit, Japanese, etc. (Figure 2). With that in mind, *eDictionary* has the potential to become a multi-source online project for all the Asian languages concerned.

However, at this point, the *eDictionary* database is not suitable for learners of Croatian as a second language (L2), but only for native speakers where Croatian is the source language (L1). This is due to the fact that information that would be useful to L2 learners still needs to be added to our database, either as a direct entry or as a link to other existing projects that would serve this same purpose. One such project, for example, is *Mrežnik* - an online Croatian dictionary project that contains a separate module focused on learners of Croatian with 1,000 entries (Hudeček & Mihaljević, 2017). Although the *eDictionary* team had the intention of making the project accessible to L2 learners, the decision was made to use Asian languages only in the sense mentioned in Hannesdóttir (2015: 245-247), i.e. that lexical descriptions of languages in online dictionaries should be based on multiple accessibility rather than on the tradition of printed dictionaries.

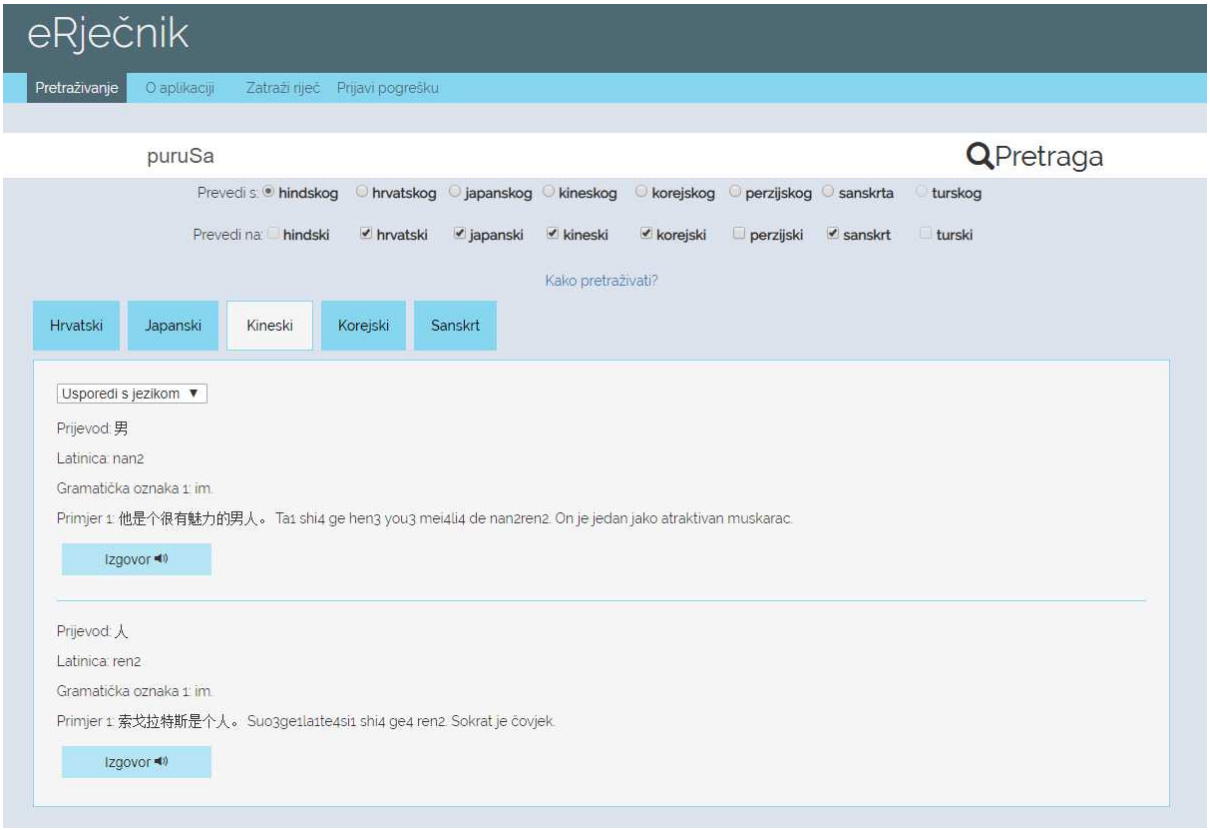

Figure 2: Example of a comparative search in *eDictionary* with Hindi as a source language.

6. Analysing the Analytics

Most of the *eDictionary* website functionality was designed in-house, including both the database model and request system. This approach lets us examine the data being recorded in the database at any point in time, namely the amount of word searches per language, which language pairs users are searching for and in which directions, as well as requested sound examples and words.

On a more global scope, we are relying on Google Analytics tracking to acquire broader data about our users. This data includes information of the country users are visiting from, number of visits, the average visit duration, and whether they are a returning or a new user.

This mixed approach to analytics allows for a more detailed overview of how the users are accessing and using our site. This method has already been described as effective by Lorentzen and Theilgaard (2012) in assessing user needs and planning on future improvements.

6.1 The Words

Looking into the global search history details, *eDictionary* users are most frequently searching with Croatian as the source language and Hindi as the target language, as presented in Tables 1 and 2 respectively.

SOURCE	COUNT	FOUND
CROATIAN	3052	1838
HINDI	521	252
SANSKRIT	200	81
JAPANESE	35	13
CHINESE	20	11
KOREAN	15	3
PERSIAN	10	5

Table 1: Count of searches, grouped by the source language

TARGET	COUNT	FOUND
HINDI	1135	929
NONE	711	710
JAPANESE	493	312
SANSKRIT	462	311
CHINESE	392	230
KOREAN	321	151
CROATIAN	257	257

Table 2: Count of searches grouped by the target language

Furthermore, we can also glean some more useful information from the search history of *eDictionary* (Table 3). Results are sorted alphabetically by the source language, i.e. the language of the queried word. The results per source language are sorted by the number of searches for each target language. Croatian is the most queried target language for all the other source languages. It is also the only source language that is paired with all the target language combinations, and is at the same time the most queried source language. These results are not surprising, since the expected dictionary users are dominantly Croatian students.

Source Language	***Target Language***	***Searches***	***Found***	***Total Searches***	***Total Found***
Chinese	Hindi	1	1		
	Persian	1	0		
	Japanese	2	2		
	Korean	2	2		
	Croatian	**6**	**6**		
	none	8	0	20	11
Croatian	Persian	77	13		
	Korean	295	132		
	none	307	0		
	Chinese	362	205		
	Sanskrit	420	276		
	Japanese	469	294		
	Hindi	**1,122**	**918**	3,052	1,838
Hindi	Persian	4	0		
	Korean	17	14		
	Japanese	18	15		
	Chinese	24	20		
	Sanskrit	42	35		
	Croatian	**168**	**168**		
	none	248	0	521	252
Japanese	Korean	3	2		
	Croatian	**11**	**11**		
	none	21	0	35	13
Korean	Japanese	2	0		
	Croatian	**3**	**3**		
	none	10	0	15	3
Persian	Japanese	1	1		
	Chinese	1	1		
	Hindi	1	1		
	Croatian	**2**	**2**		
	none	5	0	10	5
Sanskrit	Japanese	1	0		
	Korean	4	1		
	Chinese	5	4		
	Hindi	11	9		
	Croatian	**67**	**67**		
	none	112	0	200	81
Total		**3,853**	**2,203**		

Table 3: Detailed look into search history data.

The second most queried target language is **Hindi**. It is by far the most searched for target language with Croatian as a source language, and is also found as a target language for a small number of the source languages (Chinese, Persian, Sanskrit) and as a source language for most of the target languages (Chinese, Croatian, Japanese, Korean, Persian, Sanskrit). Similarly, despite the low number of total searches, **Chinese** as a source language is paired with five other target languages (Croatian, Hindi, Japanese, Korean and Persian) and is found as a target language for three (Hindi, Persian and Sanskrit). **Japanese** as a source language is only requested with Croatian and Korean as the target languages, but is found as a target language for all the other language combinations. Something similar is true for **Korean**, which is never requested as a target language except for Persian. **Persian**, with so far the lowest count of requests as a source language, was queried with Chinese, Croatian, Hindi and Japanese as target languages, but was also requested as a target language for Chinese, Croatian and Hindi source languages. Finally, **Sanskrit**, as a second most searched source language, was paired with Chinese, Croatian, Hindi, Japanese and Korean as target languages, and as a target language for Croatian and Hindi as source languages.

The detailed search history of source and target languages (Table 3) shows that there is quite a large number of searches with the target language "none". This means that the users tried searching without any specified target language, and that could have happened in two use cases. The first is the possibility of unticking all the target language boxes and thus hitting search without choosing a target language, either on purpose or by mistake. The second entails users changing the target and request languages to the same value (i.e. Source: Hindi, Target: Hindi) where the Target language checkbox gets automatically unticked. The number of such requests (711) accounts for 18.45% of the total (3,853). This could indicate that the users want single language search functionality in conjunction with the existing multiple language translation functionalities. Since we as the designers of the platform are not its core users (Nielsen, 2008), that hypothesis will require further verification in the form of a user questionnaire before any further development.

However, a look at the *eDictionary* word request system (Table 4) shows some rather unfortunate results. The total of only 26 requested words, unevenly divided between Japanese, Hindi and Sanskrit may be a sign of either a job done well on the designers' part when choosing those words to include in the dictionary, or a students' lack of interest in actively helping the dictionary to develop.

LANGUAGE	REQUESTS
JAPANESE	19
HINDI	4
SANSKRIT	3

Table 4: Requested words per language.

Since the project was done in cooperation with students that will use it the most, we believe that the answer lies in our first assumption. This would mean that we have already covered the most common words that appear in curricular activities and the usual learning materials. And in spite of the fact that not many words were requested, they were all promptly added through the Admin panel of the website by language experts cooperating on the project.

6.2 The Users

Taking a look into the Google Analytics webpage, we can get some insight into probable explanations as to why there were not as many visits as we anticipated. The *eDictionary* website has had only a handful of active users during the past two years since it was published, the only exception being the first month of its publication (Figure 3).

The data on visitor nationality is in accordance with our expectations. The great majority of visitors, an overwhelming 80.42%, are users from Croatia (Figure 4). Surprisingly enough, the number of visits from the USA ranks it as second, with 3.11% or 28 unique users. Visits from around the region are expected due to the similarities in language and cooperation with colleagues from the region.

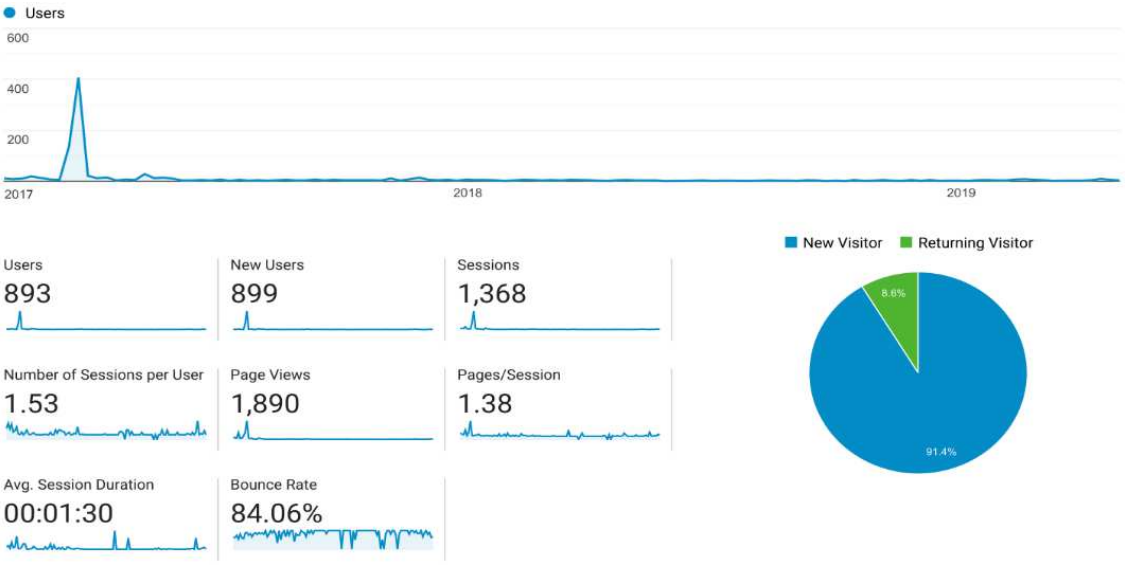

Figure 3: Number of users since 23.01.2017. to 25.05.2019.

Even though the technical side of *eDictionary* has been well thought out and technically polished, the analytics do not speak in favour of site usage. A thorough look through our e-mail system and integrated error reporting tool shows no indications of users having a buggy experience or requesting more materials. Still, somehow, the *eDictionary* project has not been able to collect more than a handful of returning users, which, considering the niche target audience of our project, though slightly discouraging, may not be that unexpected.

	Country	Users	Users
		894 % of Total: 100.00% (894)	894 % of Total: 100.00% (894)
1.	Croatia	723	80.42%
2.	United States	28	3.11%
3.	Serbia	26	2.89%
4.	Bosnia & Herzegovina	22	2.45%
5.	Brazil	18	2.00%
6.	Austria	11	1.22%
7.	Germany	8	0.89%
8.	China	5	0.56%
9.	Ireland	5	0.56%
10.	Italy	5	0.56%

Figure 4: Country of users since 23.01.2017. to 25.05.2019.

7. Conclusion and Future Work

Considering everything that we have learned over the course of this project, the research team came to several conclusions that could help with future work. The creation of *eDictionary* was a rich learning experience for everyone involved in the process, from linguists and programmers to participating students. Students' input on their needs served as an important guideline for the project. However, the MAP students' survey covered just one small group of learners active at that particular moment, and their needs should be revisited and checked against the new generations of students.

Part of the job that was not covered well and should be altered in the future (the part that certainly falls into "the Ugly" category) has to do with the promotion of active use of *eDictionary* as a learning tool among new generations of students. The active role that was envisioned for students (word and pronunciation requests, sending in examples in target languages) turned out to be not so inviting for them. We believe that this could be changed through cooperation with teachers and active integration of *eDictionary* into curricula and lesson plans.

Future work would also entail further strengthening of the *eDictionary* database, including examples and grammatical information in coordination with users' observations. At this stage, it is only accessible as L1 material, understandable and manageable by native speakers. However, since the Croatian language is also used as a second language, one further step would be to make the Croatian database appropriate for such use as well.

8. Acknowledgements

This research was supported in part by the University of Zagreb Research Grant (43-917-1030). *eDictionary* would not be possible without the help of our students, both new and old, as well as their teachers. Their help is gratefully acknowledged. We thank Mr. Navaey for giving us guidance on the Persian resources.

9. References

Alipour, M., Robichaud, B. & L'Homme M.-C. (2015). Towards an Electronic Specialized Dictionary for Learners. In I. Kosem, M. Jakubíček, J. Kallas & S. Krek (eds.) *Electronic lexicography in the 21st century: linking lexical data in the digital age.* Proceedings of the eLex 2015 conference, Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., pp. 51–69.

Deksne, D., Skadiņa, I. & Vasiļjevs, A. (2013). The modern electronic dictionary that always provides an answer. In I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets & M. Tuulik (eds.) *Electronic lexicography in the 21st century: thinking outside the paper.* Proceedings of the eLex 2013 conference. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, pp. 421-434.

Dovedan, Z., Seljan, S. & Vučković, K. (2002). Multimedia in Foreign Language Learning. In P. Biljanović & K. Skala (eds.) *Proceedings of the 25th International Convention MIPRO 2002: MEET + MHS.* Rijeka: Liniavera, pp. 72-75.

Družijanić Hajdarević, E., Vučković, K. & Dovedan, Z. (2006). Računalo ili raćunalo uz pomoć računala. In *Proceedings of the 29th International Convention MIPRO, Rijeka*, pp. 283–287.

Hannesdóttir, A. H. (2015). What is a target language in an Electronic Dictionary? In I. Kosem, M. Jakubíček, J. Kallas, & S. Krek (eds.) *Electronic lexicography in the 21st century: linking lexical data in the digital age*. Proceedings of the eLex 2015 conference. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., pp. 236-249.

Heid, U., Prinsloo, D. & Bothma, T. (2013). Dictionary and corpus data in a common portal: state of the art and requirements for the future. *Lexicographica - International Annual for Lexicography* / Internationales Jahrbuch für Lexikographie. 28, 10.1515/lexi.2012–0014, pp. 269-291.

Heuberger, R. (2016). Learners' Dictionaries: History and Development; Current Issues. In P. Durkin (ed.) *The Oxford Handbook of Lexicography.* Oxford/New York: Oxford University Press, pp. 25-43.

Hudeček, L. & Mihaljević, M. (2017). The Croatian Web Dictionary Project – Mrežnik. In I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek & V. Baisa (eds.) *Electronic lexicography in the 21st century*: Proceedings of the eLex 2017 conference. Brno: Lexical Computing CZ s.r.o., pp. 172-192.

Janjić, M., Librenjak, S. & Kocijan, K. (2016a). Asian language teaching and learning - the influence of technology on students' skills in SL classroom. In T. Erjavec &

D. Fišer (eds.) *Language Technologies & Digital Humanities* 2016, Ljubljana: Ljubljana University Press, Faculty of Arts, pp. 196-197.

Janjić, M., Librenjak, S. & Kocijan, K. (2016b). Croatian Students' Attitudes Towards Technology Usage in Teaching Asian Languages - a Field Research, In *MIPRO 2016*, Rijeka, pp. 1051-1056.

Janjić, M., Požega, M., Poljak, D., Librenjak, S. & Kocijan, K. (2017a). E-dictionary for Asian Languages. In I. Atanassova, W. Zaghouani, B. Kragić, K. Aas, H. Stančić, H. & S. Seljan (eds.), *INFuture2017 Proceedings: The Future of Information Sciences*, Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, pp. 213-216.

Janjić, M., Librenjak, S. & Kocijan, K. (2017b). Nastava stranih jezika: upotreba tehnologije. *Strani jezici: časopis za unapređenje nastave stranih jezika*, 44(4), pp. 232–243.

Lauc, T., Matić, S. & Mikelić, N. (2006). Educational multimedia software for English language vocabulary. In *Proceedings of the 1st International Conference on Multidisciplinary Information Sciences and Technologies: InSciT2006, Vol. I: Current Research in Information Sciences and Technologies Multidisciplinary approaches to global information systems*, Merida, pp. 117–121.

Librenjak, S., Janjić, M. & Kocijan, K. (2016a). Sustainable vocabulary acquisition in Japanese classroom with the help of Memrise. In M. Janesova, H. Kratochvilova, I.G. Rotaru, S. Pal, S. Anjali, R. Kratochvil, J. Grover & O. Beyhan (eds.) *Proceedings of International Academic Conference on Global Education, Teaching and Learning in Budapest.* Vestec: Czech Republic: Czech Institute of Academic Education z.s., pp. 54-61.

Librenjak, S., Janjić, M. & Kocijan, K. (2016b). Computer assisted learning of Japanese verbs - Analysis of errors in usage by Croatian students. In J. Vopava, V. Douda, R. Kratochvil & M. Konecki (eds.) *Proceedings of MAC-ETL 2016.* Prague: Academic Conferences Association, pp. 262-273.

Librenjak, S., Kocijan, K. & Janjić, M. (2016c). Improving Students' Language Performance Through Consistent Use of E-Learning: An Empirical Study in Japanese, Korean, Hindi and Sanskrit. *Acta Linguistica Asiatica*, 6(2), pp. 79-94.

Librenjak, S., Vučković, K. & Dovedan Han, Z. (2012). Multimedia assisted learning of Japanese kanji characters. In P. Bijanović, Ž. Butković, K. Skala, S. Golubić, N. Bogunović, S. Ribarić, M. Čičin-Šain, D. Cisić, Ž. Hutinski, M. Baranović, M. Mauher & J. Ulemek (eds.) *Proceedings of 35. jubilee international convention on information and communication technology, electronics and microelectronics*, Rijeka, Croatian Society for Information and Communication Technology, Electronics and Microelectronics - MIPRO, pp. 1284-1289.

Lorentzen, H. & Theilgaard, L. (2012). Online dictionaries – how do users find them and what do they do once they have. *Proceedings of the 15th EURALEX International Congress*, pp. 654-660.

Majeed, A. & Rauf, I. (2018). MVC Architecture: A Detailed Insight to the Modern

Web Applications Development, *Peer Review Journal of Solar & Photoenergy Systems*, vol 1, No 1, PRSP.000505.

Nielsen, J. (17.3.2008). *Bridging The Designer – User Gap.* Accessed at: https://www.nngroup.com/articles/bridging-the-designer-user-gap/ (June 1, 2019).

Paulson, L. D. (2005). Building rich web applications with Ajax. In *Computer*, 38(10), pp. 14-17.

Pavičić Takač, V. (2008). *Vocabulary Learning Strategies and Foreign Language Acquisition.* Clevedon/Buffalo/Toronto: Multilingual Matters LTD.

Renau, I. & Battaner, P. (2011). The Spanish Learner's Dictionary DAELE on the Panorama of the Spanish E-lexicography. In I. Kosem & K. Kosem (eds.) *Electronic lexicography in the 21st century: new applications for new users.* Proceedings of the eLex 2011 conference, Ljubljana: Trojina, Institute for Applied Slovene Studies, pp. 221-226.

Summers, D. (1988). The role of dictionaries in language learning. In R. Carter & M. McCarthy (eds.) *Vocabulary and Language Teaching.* London/New York: Routledge, pp. 111-125.

Vacalopoulou, A. & Efthimiou, E. (2015). Multilingual lexicography for adult immigrant groups: bringing strange bedfellows together. In I. Kosem, M. Jakubíček, J. Kallas & S. Krek (eds.) *Electronic lexicography in the 21st century: linking lexical data in the digital age*. Proceedings of the eLex 2015 conference. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd. pp. 315-326.

Websites:

Baza hrvatskih glagolskih valencija. Accessed at: http://valencije.ihjj.hr (May 25, 2019)

Hrvatska Enciklopedija. Accessed at: http://www.enciklopedija.hr/ (May 20, 2019)

Hrvatski Jezični Portal. Accessed at: http://hjp.znanje.hr (May 17, 2019)

Struna | Hrvatsko strukovno nazivlje. Accessed at: http://struna.ihjj.hr/ (May 12, 2019)

Hellwig, O. (2010-2016). *DCS - The Digital Corpus of Sanskrit*. Berlin. Accessed at: http://www.sanskrit-linguistics.org/dcs/ (July 15, 2019)