A Call for a Corpus-Based Sign Language Dictionary: An Overview of Croatian Sign Language Lexicography in the Early 21st Century

Klara Majetić, Petra Bago
Faculty of Humanities and Social Sciences, Zagreb
E-mail: klaramajetic13@gmail.com, pbago@ffzg.hr

Abstract

Many sign languages today are still not standardized nor accessible to a wider audience. Sign languages with high quality dictionaries are scarce. In this paper we give a brief description of sign languages and how they differ from spoken ones, in order to better understand the issues lexicographers might face when compiling dictionaries for these languages. We focus on Croatian Sign Language (HZJ), giving an overview of the current situation in HZJ lexicography following some criteria we find relevant for both online and printed sign language dictionaries. The criteria have been classified into twenty-five categories and applied to create a model for an online HZJ dictionary, briefly presented in this paper. By presenting an unsatisfactory status of HZJ lexicography, we are issuing an urgent call for the compilation of a HZJ corpus as a basis for a high quality dictionary that could benefit both the potential hearing and deaf users.

Keywords: Croatian Sign Language (HZJ), sign language lexicography, dictionary evaluation, e-dictionary model

1   Introduction to Sign Languages and the Croatian Sign Language

Sign languages worldwide have not always been recognized as natural, genuine languages with an extensive vocabulary and complex grammar. It was only in the middle of the 20th century that the sign languages have been perceived to be as flexible and as expressive as spoken languages. The pioneer in this paradigm shift was William Stokoe and his linguistic research of the American Sign Language (ASL). His work has influenced other sign language researchers worldwide to investigate sign languages not as mere systems of gesture or simplified versions of their spoken counterparts, but as complex and independent natural languages in their own right. In order to better understand issues in lexicographical works on sign languages, a brief description of the Croatian Sign Language (HZJ) follows.

As with all sign languages when faced with their spoken counterparts, HZJ has a different modality from spoken Croatian – instead of an oral-auditory modality, sign languages have a spatial-visual modality (Kavčić 2012). A sign is a basic unit in sign languages, like a word in spoken ones. A sign consists of five elements: handshape, location, movement, orientation, and non-manual markers (mouthing, facial expressions, etc.). One of the first problems lexicographers face when approaching sign languages is the fact that they do not have a standard written form – there are a few notation systems (Stokoe notation[1], *HamNoSys*[2], and *SignWriting*[3] being some) and the option of glossing[4] (Filić

---

1   Stokoe 1960

2   Hanke 2004

3   URL: http://www.signwriting.org/

4   Glossing is a way of writing down signs using words – in HZJ glosses are verbs in infinitive form, nouns in singular and nominative case, adjectives in nominative case, singular and masculine grammatical gender.

2016). Notation systems use symbols and abstract pictures to describe a sign and all its elements, and are not written in a single line, but also use the vertical plane to add information (Stokoe 1960; Costa & Dimuro 2003). As a result of such writing requiring a lot of space on a page, traditional printed dictionaries had to have a rigorous lemma selection. Ordering of entries is an additional complication if lemmas are written in one of the mentioned notation systems, due to the lack of standardization. On the other hand, glossing conveys too little visual-spatial information – it gives very little information about sign elements because it assumes the user already knows what the sign looks like, and is therefore not intended for beginners. Sign languages provoke another problem by having several ways of modifying signs in different contexts. Those modifications provoke a question as to whether they should each be separate entries or all part of one entry. In recent years many researchers worldwide have been struggling with lexicographical issues regarding sign languages, such as Zwitserlood, Hedegaard Kristoffersen and Troelsgård (2013), Zwitserlood (2010), Capovilla (2003), Hanke and Storz (2008), König, Reiner and Langer (2004), among others. Most of the issues could be solved by utilizing new technologies rich with multimedia and thus moving on from the traditional printed dictionaries (Singleton 2000).

2   Current HZJ Lexicography

The current Croatian lexicography lacks a comprehensive sign language dictionary. In this section we describe existing dictionaries of HZJ using an instrument for evaluating (online) dictionaries of sign languages we have developed (Majetić & Bago 2017). The dictionaries are described according to the following criteria: intended users, subject field covered, type of dictionary according to the norm, number of languages in the dictionary, scope, function, direction, comprehensiveness, source and target language, data collection, data selection, multimedia, written form of the sign, other information regarding the sign, additional information in an entry, restrictions on content, content creators, content updates, content download feature, user interface design, searching and search results, browsing, and extra content.

At the moment HZJ has just one small traditional printed dictionary, *Hrvatski znakovni jezik*[5], published in 2015. All other dictionaries are not standalone works but a part of textbooks, such as *Znak po znak*[6] *1-3* (ZPZ) that was published in 2006 and 2007, and *Gluhi i znakovno medicinsko nazivlje: kako komunicirati s gluhim pacijentom*[7], published in 2010. The above-mentioned works are all alphabetical lists of words translated into HZJ by a photograph or an ilustration. *Znak po znak* coursebooks are each accompanied by a DVD with video content, but it loads slowly and is somewhat difficult to navigate. HZJ used to have only one online dictionary, which is no longer fully usable, called *CroDeafWeb* – but the project has been abandoned and is no longer up to date with today's Internet browsers. The multilingual online dictionary *Spread the Sign* partnered up with a Croatian team in December 2015. To date they have added almost 10,000 entries into the dictionary out of the planned 15,000. A comparison of the number of entries in these works is given in Table 1.

In the following Sections 2.1, 2.2. and 2.3 we give an overview and a description of *Hrvatski znakovni jezik*, *CroDeafWeb* and *Spread the Sign* according to the abovementioned criteria. These three works have been chosen because they are not a part of textbooks, but instead are standalone dictionaries.

---

5   Eng. Croatian Sign Language.

6   Eng. Sign by Sign.

7   Eng. The Deaf and the Medical Sign Terminology: How to Communicate with a Deaf Patient.

Table 1: Approximate number of entries in various dictionaries of HZJ

Dictionary:	Number of entries:
Znak po znak	4,500
Hrvatski znakovni jezik	1,200
CroDeafWeb	500
Gluhi i znakovno medicinsko nazivlje	250
Spread the Sign	10,000

2.1   Hrvatski znakovni jezik

The second unaltered edition of this general traditional printed dictionary was published in 2015. The dictionary was created through an EU project "Poticanje zapošljavanja mladih jačanjem njihovih kompetencija za stjecanje boljeg položaja na tržištu rada"[8] to aid the education of future communication intermediaries in education for the deaf. It has been intended for hearing users as a general normative dictionary, with a primary function of helping in sign production. The dictionary is bilingual, monoscopal and monodirectional – the source language is Croatian and the target language is HZJ. The entries are listed in the alphabetical order of Croatian words. There are approximately 1,200 entries which is not enough for a comprehensive dictionary, especially considering that they have been chosen with no regard to the frequency of words nor signs. For example, one can find the word *žuboriti* (to murmur, to babble), but not *žvakati* (to chew). It is not known from where the data was collected, and there is no additional data within entries besides the translation. A word in Croatian is translated to HZJ by a photograph of a person signing. The editors are not professional lexicographers. As a first standalone HZJ dictionary of its kind this project is praiseworthy, and it was to be expected for a pioneer project to have some shortcomings. However, owing to the editors not being professional lexicographers, an inexpert approach is unfortunately evident in all aspects of the dictionary. Nevertheless, due to the commendable work done by the editors by compiling the first standalone printed Croatian dictionary of sign language, it is now possible to analyze the pros and cons of the work, and use these insights to compile a new and improved dictionary of HZJ.

2.2   CroDeafWeb

CroDeafWeb was the first online dictionary of HZJ (http://www.crodeafweb.org/rjecnik/index.html) published around the year 2000 as a part of a same named portal (http://crodeafweb.org/). Unfortunately, it has not been updated for a long time, and the video content is no longer supported by today's Internet browsers. The dictionary has a section where it uses gif format files, which still work but are not very detailed. Besides gifs, there is a short description of each sign's elements, such as handshape and hand movement. It was probably intended for all groups of users as a normative dictionary with a primary function of helping in sign production. This dictionary seems to have been intended to be general, but it contains a special section Liturgy in Sign Language, and liturgical words are a big part of the dictionary. The dictionary is multilingual, monoscopal and monodirectional – the source languages are Croatian and English and the target languages are HZJ and Croatian. There is no search function. It is only possible to browse the entries, which is why they are listed in alphabetical order of Croatian and English words. There are also a few topics (e.g. food, days of the week) one can use to filter entries. The user interface design is simple, with a good

---

8   Eng. Promoting youth employment by strengthening their competencies to gain a better position in the labour market.

overview of the page. There are no options for typographic adaptions[9]. The home page contains some basic instructions on how to browse the dictionary, and the necessary system requirements needed to use it. The dictionary contains approximately 500 entries, which is again not sufficient for the average user. There are no restrictions prohibiting the users from saving the entries, but there is no explicit download button for this action. It is not known from where the data was collected and who the compilers were. There is no functioning extra content, but there are links to other sources that no longer work.

2.3   Spread the Sign

The online dictionary *Spread the Sign* (https://www.spreadthesign.com/) started in 2006 as a Swedish project that was initially funded by the European Commission and the Leonardo program under the slogan Life Long Learning. Today it is a general normative dictionary intended for all groups of users, with the function of helping in both sign reception and sign production. It contains 35 sign languages and their spoken counterparts, which makes it multilingual. The dictionary is monoscopal as it translates from a spoken into a sign language, but there are links within the entries that enable translation from one sign language into another. The dictionary is bidirectional inasmuch as it is intended for native users of any spoken or sign language. At the moment, there are over 380,000 signs from various sign languages – the goal is to have 15,000 entries in the dictionary for every sign language, which should satisfy the needs of the majority of users and thereby "make sign language available all over the world".[10] Each language represented in the dictionary has a team in its country that works on compiling the dictionary and creating video content. These teams consist of sign language researchers and native speakers. The signs seem to be representative and authentic. The Croatian team joined the project in 2015, and have so far added 10,565 entries. Not all entries have gotten their HZJ translations, but it is possible to translate Croatian into other signed languages. Only the spoken languages are source languages, while the sign languages are target languages. It is not described what the first source for the entries list was, but we know other languages that were added over time followed the existing ones and added their translations, both in the spoken and the sign language. The only data all entries consist of is a gloss in a spoken language, a detailed video of signing and links to other sign languages that have added the same word and its translation of a sign. Some entries also have an illustration or a picture related to the meaning of the entry. Entries can be divided by topic such as: profanities, nouns, numbers, colors, music, business and so on. There is a special category with whole sentences ranging from *Are you deaf?* to *Do you want mustard, mayo, or both?* which are very useful for a user who is just starting to learn the sign language. All content is free, and even though there is no explicit download button within the entry, video can be saved to a user's device. Users cannot create new content, but they can contact each country's team to let them know if there are any mistakes in the translations. The content seems to be frequently added to and updated. The user interface is simply designed, and one can easily navigate between connected information, but there are no options of typographic adaptivity nor a user's guide. The search function is monodirectional – users can only search from a spoken language to signed languages. The user has to first choose the language they want at the top of the page. The search result list contains all words that could match the searched word and the list of signed languages that the word can be translated to. The user can browse the content of the dictionary alphabetically and/or thematically. There is no extra content. This dictionary offers a lot to the average user, and with time might add a lot more information to its entries.

---

9   The deafblind use sign language too. Therefore, an e-dictionary should allow typographic adaptions (e.g. font size and type, contrast) to allow easier use for such users.

10   European Sign Language Centre (ESLC) Accessed at: https://www.signlanguage.eu/en/about-us/ (10.3.2018.)

3   Model of an Online HZJ Dictionary

In this section we present the ideal online HZJ-Croatian dictionary following the evaluation criteria mentioned earlier. An ideal HZJ-Croatian dictionary is an online, general, descriptive, bidirectional and bilingual dictionary that takes the needs of the community into consideration. It is biscopal and intended for all groups of potential users. Its functions are directed to both text and sign reception and production. The ideal dictionary is based on a corpus that at the moment does not exist. Using such a corpus, the lexicographers can determine the authenticity and frequency of signs, and thereby choose the initial 5,000 entries that we see as the required minimum for the first publication of the dictionary. The corpus is the source for the multimedia. Each entry contains a video of the sign, a gloss, a textual description of the sign's elements, information on context or the use of the sign with examples, a definition of the meaning, ID number and the topic of the entry. The website clearly states how the content may be used and how it is protected. The ideal dictionary is free and available to all users. Distribution for non-commercial purposes is allowed, and so is downloading of the content. The main authors of the first edition are professional lexicographers. The users have a possibility to send feedback, corrections, and suggestions. Updates and revision of the existing content and adding of new content is done on a regular basis. The dictionary is easy to use and intuitive. The user interface design is simple. with options for typographic adaptivity. A user's guide, instructions for entry content, key to symbols, notations system and abbreviations are always accessible. All relevant content within the dictionary is connected by hyperlinks to make it easily and quickly accessible. The user has the option to search the dictionary by a Croatian word (spellcheck is an automatic part of the search tool) or by choosing elements of a sign and topic it might fall into. The list of search results contains the video of the sign, the gloss and the ID of the entry. The user is allowed to narrow down the search by adding new search criteria. It is possible to browse the entries alphabetically, thematically or by the sign elements. The necessary extra content are texts that describe HZJ, its grammar and the orthography used in the ideal dictionary.

In later editions the dictionary is expanded with less frequent entries, new entry content (e.g. geographical aspect), written form of signs using one of the notation systems, some more extra content such as words or signs of the day, educational games and other languages.

4   Conclusion

Croatian Sign Language (HZJ), like all sign languages and their spoken counterparts, has not yet been as thoroughly researched and described as Croatian has. There is a need for new learning resources in an online format rich with multimedia, as this is the only way to convey enough visual-spatial information of a sign language. Such a format also allows changes to be made if needed, easier additions to the dictionary and access for all potential users.

Sign languages that have built a corpus from which they can choose entries are very rare. Some existing corpora are Netherlands Sign Language (NGT) corpus, German Sign Language (DGS) corpus, Australian Sign Language corpus and a few others (Crasborn 2010). Having a corpus allows lexicographers to research the frequency of signs (which is not equal to the frequency of words in spoken languages), as well as research the use of each sign, and use the resulting discoveries to produce a better dictionary. HZJ has no existing corpus, and, to the best of our knowledge, no corpus is currently being compiled. Most of the lexicographical issues with sign languages can be handled by moving away from the traditional printed formats towards electronic dictionaries – images and videos can provide enough visual-spatial information, and such dictionaries are easier to use for all groups of users. By presenting the unsatisfactory current state of Croatian sign language

lexicography, we hereby issue a call for a much-needed corpus-based online dictionary of HZJ that will satisfy the needs of all user groups.

References

Capovilla, F. et al. (2003). Brazilian sign language lexicography and technology: Dictionary, digital encyclopedia, chereme-based sign retrieval, and quadriplegic deaf communication systems. In *Sign Language Studies*, 3(4), pp. 393-430.

Crasborn, O. (2010). The Sign Linguistics Corpora Network: towards standards for signed language resources.

CroDeafWeb: rječnik hrvatskog znakovnog jezika. Accessed at: http://www.crodeafweb.org/rjecnik/index.html [3.4.2017.]

da Rocha Costa, A. C., & Dimuro, G. P. (2003). *SignWriting and SWML: Paving the way to sign language processing*. Atelier Traitement Automatique des Langues des Signes, TALN 2003.

Filić, M. *Izražavanje količine u hrvatskom znakovnom jeziku.* Nacionalni repozitorij završnih radova (ZIR). Accessed at: https://urn.nsk.hr/urn:nbn:hr:158:362319 [10.8.2017.]

Hanke, T. (2004). HamNoSys-representing sign language data in language resources and language processing contexts. In *LREC* (Vol. 4).

Hanke, T., Storz, J. (2008). iLex–A database tool for integrating sign language corpus linguistics and sign language lexicography. In *LREC 2008 Workshop Proceedings*. W 25: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora. Paris: ELRA pp. 64–67.

Ristić, M., Baštijan, Z., Biškupić Andolšek, T. (Eds.). (2015). *Hrvatski znakovni jezik*. Zagreb: Hrvatski savez gluhih i nagluhih.

Kavčić, D. (2012). *Hrvatski znakovni jezik: pregled opisanih jezičnih elemenata*. Diplomski rad. Filozofski fakultet Sveučilišta u Zagrebu. Accessed at: www.darhiv.ffzg.unizg.hr/4454/ [10.8.2017.]

König, S., Reiner K., Langer, G. (2004). What's in a sign? Theoretical lessons from practical sign language lexicography. Signs of the time. In *Selected papers from TISLR,* pp. 379-404.

Majetić, K., Bago, P. (2017). Proposing an Instrument for Evaluation of Online Dictionaries of Sign Languages. In *INFuture 2017 – Integrating ICT in Society*, pp. 189.

Singleton, D. (2000). *Language and the lexicon. An introduction*. Arnold.

Stokoe, W. C. (1960). Sign Language Structure: An Outline of the Visual Communication System of the American Deaf. In *Studies in Linguistics: Occasional Papers 8.* Linstock Press.

Šegota, I., Šendula-Jengić, V., Herega, D., Petaros, A., Conar, J. (2010). *Gluhi i znakovno medicinsko nazivlje: kako komunicirati s gluhim pacijentom*. Zagreb. Medicinska naklada.

Tarczay, S., et al. (2006–2007). *Znak po znak 1, 2, 3: udžbenik za učenje hrvatskog znakovnog jezika.* Zagreb: Hrvatska udruga gluhoslijepih osoba „Dodir".

Zwitserlood, I. (2010). Sign language lexicography in the early 21st century and a recently published dictionary of Sign Language of the Netherlands. In *International Journal of Lexicography* 23.4, pp. 443-476.

Zwitserlood, I., et al. (2013). Issues in sign language lexicography. In *The Bloomsbury Companion to Lexicography*. London: Bloomsbury, pp. 259-283.