Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2020

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2020

Selected methodological challenges of sign language lexicography

Lucia Vlášková*,†, Hana Strachoňová*,†

* Support Centre for Students with Special Needs (Teiresiás), Masaryk University
Komenského náměstí 2, 602 00 Brno, Czech Republic
† Department of Linguistics and Baltic Languages, Faculty of Arts, Masaryk University
Arna Nováka 1, 602 00 Brno, Czech Republic
l.vlaskova@gmail.com
hana.strachonova@gmail.com

Abstract

Sign language lexicography, as a growing field of study within sign language linguistics, faces many challenges that have already been answered for the audio-oral language material. In this paper, we present some of these challenges together with methods developed to help navigate the often problematic field of lexical classification. The described methods and strategies are implemented in the first Czech sign language online dictionary, a part of the platform *Dictio*. Among the covered topics are questions of lemmatisation and how to decide what constitutes a lexeme in sign language; what is the place of classifier constructions in a dictionary, given their peculiar semantic status; what is the proper treatment of mouthings and mouth gestures concerning citation forms and derivation; why is it difficult in sign language to distinguish synonyms from variants and how can our proposed phonological criteria help; and, finally, what kinds of utterances are considered good examples of use.

1. Introduction

Dictio is a multilingual online dictionary that includes multiple languages, both sign and spoken. This ongoing project is being realised at the Masaryk University in Brno, Czech Republic. At the moment, it includes entries for the following languages (ordered by the approximate number of entries): Czech (120 thousand), Czech Sign Language (13 thousand), Slovak Sign Language (7 thousand), Slovak (7 thousand), English (5 thousand), Austrian German (800), Austrian Sign Language (650), International Sign (170), and American Sign Language (120). Only a part of the entries is published, the rest is still a subject of the editing work of multiple working groups, including international teams of Deaf university employees. At the moment (August 2020), the numbers of the published entries for the sign languages are as follows: Czech Sign Language – 3670, Slovak Sign Language – 130, International Sign – 100, American Sign Language – 80, and Austrian Sign Language – 0.

The field of sign language lexicography is growing rapidly. Considering Stokoe's (1960/2005) description of ASL lexical units as the pioneering work, which respect the established linguistics principles, sixty years later, we make use of systematised databases for a whole range of sign languages in the form of printed books or off-line and on line databases (see the overview in McKee and Vale, 2017 or Fenlon et al., 2015). Since the seminal work of Johnston and Schembri (1999) on lemmatisation of the AUSLAN corpus (and closely connected AUSLAN lexical database), several researchers have published their experiences in the form of applicable universal guidelines for the lexicographic work on any sign language. At the moment, many topics concerning mainly the electronic lexical databases are addressed in the literature: e.g., history and options of the sign description and search (Zwitserlood, 2010, focusing on dictionaries of NGT), lexicographic specifics of sign languages compared to spoken languages (Kristoffersen and Troelsgård, 2012, with particular focus on the DTS lexical database), phonological and morphological variation in the process of lemmatisation (Fenlon et al., 2015, on the material of BSL), and others. At the beginning (around 2009), our project was inspired mainly by the work of Johnston and Schembri (1999) and on-line public dictionaries that included semantic definitions in sign language (e-LIS – the dictionary of LIS; and Elix – the dictionary of LFS).

In the absence of a representative Czech SL corpus, the linguistic material for the Czech SL part comes from two primary sources: published dictionaries and Czech SL informants. *Dictio* has the ambition to collect all the published Czech SL dictionaries and make them available in one database. That covers printed books (mainly Potměšil, 2002; 2004a; 2004b), CDs (Langer, 2005a; 2005b; 2008, a.o.), and other individual projects (e.g., diploma theses focusing on specific semantic fields, teaching materials for Czech SL commercial or university courses). The collection of previously published material is being edited, annotated and completed by the team of native signers of Czech SL. The substantial part of the team's work is to discuss synonyms and variants for the published items. This way, plenty of new material is being elicited for the *Dictio* database.

In this paper, we introduce selected topics from sign language lexicography. The idea is to describe some linguistic issues we have encountered while working on the material for the Czech SL part of the dictionary and propose a guideline that could be applicable in the field of sign language lexicography in general. Czech SL was the first language introduced into the dictionary. Creating the linguistic methodology has been especially challenging since the original vision of the whole project was to build the first monolingual dictionary, in this case, a dictionary of Czech SL, where the meaning and the use of the signs are explained and illustrated solely in Czech SL. Such entries, then, include the links to the parts containing other languages (translations). However, this paper is focused on the internal classification of Czech SL material and does not cover translations between spoken and sign languages.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2020

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2020

2.   Lemmatisation and types of dictionary entries

The most fundamental question to be raised when composing a sign language dictionary is what kind of signs to include, i.e., what constitutes an entry in a dictionary. To answer this question, the following strategy has been developed: first, we take all the possible kinds of signs occurring in natural speech (lexeme, deixis, description, compound, collocation, set phrase) and divide them into two groups according to their complexity: the ones that do not consist of individual semantic units (lexeme, deixis) and the ones that do (description, collocation, compound, set phrase). Among the first group, we set aside the expression, the meaning of which changes according to the referent (deixis) and select the expression with a conventionally established meaning (lexeme). From the second group, we single out the expressions with a non-compositional meaning, i.e., the set phrase and the compound. Thus, this strategy leaves us with only three candidates for a dictionary entry: a traditional lexeme, a compound, and a set phrase, with each of their meanings conventionally established.

In our dictionary, however, we classify signs into four types of entries: simple signs, compounds, set phrases, and derivatives.

Simple signs are lexemes with one root, one handshape, and one place of articulation. Thus, the Selected Finger Constraint and Place Constraint hold (Sandler, 2006).[1]

Compounds are morphologically complex signs that originated by merging two independent signs, i.e., two free morphemes. From the semantic point of view, compounds are not bound to introduce a new meaning, as seen in the Czech SL example of SUNGLASSES = SUN^GLASSES. Nevertheless, it is possible, e.g., FLOWER^SPRING = MAY (Mladová, 2009). It is often difficult to distinguish compounds from set phrases, the third type of entries in our dictionary. Set phrases also consist of two (or more) free morphemes, but their meaning is not compositional, e.g., in Czech SL sign UNIVERSITY, which consists of the signs HIGH and SCHOOL. However, in the case of compounds it is not the semantic shift that classifies them as such but the phonological reduction/assimilation, as defined by Zeshan (2004): the first sign is shortened and loses stress, any repetitions and internal movements are deleted, handshape and location can be assimilated, and the passive hand can function as a place of articulation.[2] On the other hand, no such modification can be found in set phrases, where all constituting signs are fully realised.

The last type represented in our dictionary are the derivatives, defined as forms that have been derived from their respective motivating signs by means of adding or changing a non-manual component, specifically mouthing or mouth gesture. Typically, this process occurs while deriving a technical or more specific term from a general vocabulary sign. Take an example from Czech SL where SACCHARIDE is derived from SUGAR. These two signs have the same manual component but differ in mouthing.

SUGAR is standardly articulated without mouthing, and SACCHARIDE contains the mouthing of the Czech word *sacharid*, proving mouthing to be of a significant lexical role (Sandler, 2006).[3] We will cover mouthing and mouth gestures in more detail in Section 4.

Another important question is the choice of a citation form of each entry. Following Johnston and Schembri (1999), only the unmodified signs in their basic forms are present in the lexicon (and, therefore, the dictionary), inflection and modification is part of the grammar. Modification can take several forms, as defined in Zeshan (2003): (i) modified movement expresses the change in aspect, number, degree or directionality (verbal inflection encoding the subject and/or the object of the given verb); (ii) modified handshape signals classifier constructions and numeral incorporation; (iii) modified facial expressions distinguish between indicative, interrogative, negative and other clause types (Zeshan, 2004). In our dictionary, the information whether a sign can incorporate numerals, subject and/or object, and other modifiers is given in the grammatical part of the dictionary entry for the corresponding lemma in its basic form, i.e. singular, non-modified and non-intensified sign.[4] The basic form for signs that incorporate a numeral is the one with incorporated ONE. For directional signs, it is the form directed from the speaker to the addressee.

However, there are exceptional cases when the dictionary also covers other than basic forms of signs. Such instances include deixis with fixed initial position, e.g., the pronouns I and MY that are always signed facing the speaker, and, correspondingly, YOU and YOUR, always facing the addressee. Furthermore, lexicalised forms of different types have their place in the dictionary, e.g., lexicalised deixis. Take the Czech SL verb HEAR, which is realised by pointing to the speaker's ear with a crooked index finger. Normally, this would be a deictic sign meaning *ear*, which does not belong in the dictionary, but slight modifications (different facial components – raised eyebrows, added mouthing) indicate that it is lexicalised. A semantic shift occurred where the meaning is no longer the object at which the finger is pointing but the activity of hearing. Other forms of lexicalisation include lexicalised classifier constructions, which we will discuss in the following section, or lexicalised fingerspelling, as is the case with the sign for *engineer* – ING, fingerspelled with the letters of the Czech SL alphabet.

3.   Classifiers, specifiers and lexicalised constructions

Classifiers have repeatedly proven to be an exciting research topic among sign linguists. This section will focus on different types of classifiers, a closely related group of specifiers, and the ways of properly incorporating them into a dictionary.

Sign language classifiers are considered a special kind of morphemes representing nominals, the meaning of which is not precisely specified. They denote relevant

---

[1] Consequently, items that, synchronically, do not consist of two lexical roots but have two different places of articulation are classified as compounds. See, e.g., the sign POST_OFFICE in the appendix, with the movement realised from the head towards the non-dominant hand. Note that most of the following examples will also be provided in the appendix.

[2] At least one of the reduction/assimilation patterns must be present to classify the item as a compound.

[3] More precisely, the sign for SUGAR may be accompanied by the mouthing of the Czech word *cukr* (sugar), but the sign for SACCHARIDE must be articulated with the mouthing of *sacharid*.

[4] See, e.g., the entry for HOUR, unfortunately, with the commentary in Czech only.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2020

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2020

properties of entities via different configurations of the manual articulator (Zwitserlood, 2012), specify shapes and dimensions of objects, and denote spatial relations and motion events (Sandler and Lillo-Martin, 2006b). Such entities are then categorised according to the properties into groups, e.g., flat objects, long and thin objects, two-legged beings, etc. Classifiers have been attested in all known sign languages (Sandler and Lillo-Martin, 2006b), thus constituting a stable class with common general attributes, although the inventory of the particular classifiers differs from one language to another (Zwitserlood, 2012).

The categorisation of different types of classifiers has been a subject of much discussion. Earlier literature (Supalla, 1986, a.o.) had divided them into multiple classes based on various characteristics (e.g., semantics, shape, function, animacy) before currently stabilizing on two main ones: whole entity classifiers and handling classifiers, based more on their function in grammar rather than their semantic properties (Zwitserlood, 2012).

Whole entity classifiers denote their referents in their entirety. They are more abstract and "refer to general semantic classes rather than to visually perceived physical properties" (Sandler and Lillo-Martin, 2006b:77). However, various classifiers can be used to denote a single entity, each highlighting a different relevant aspect (Zwitserlood, 2012).

On the other hand, handling classifiers utilize iconicity on a larger scale; they indicate the shape of an entity as it is being held or manipulated with, while the manual articulator represents itself – a hand holding the entity. This strategy gives the speaker much more room to choose among different classifiers according to the situation in the actual world (Zwitserlood, 2012).

These types of classifiers, considered to be bound morphemes, must occur jointly with other expressions within so-called classifier constructions, within which they are incorporated mostly into classifier verbs, i.e. verbs denoting movement, position or existence of a referent in space or some kind of manipulation (Zwitserlood, 2012). Classifier constructions represent a very productive strategy in sign languages, and this unstable semantic and morphological status prevents them from being documented in a dictionary.

However, classifiers outside of classifier constructions (so-called classifier handshapes) can be documented. In our dictionary, classifier handshapes are registered in individual lexical entries if there is a (relatively neutral) stabilised representative form with (at least roughly) delimited meaning (e.g., via extensional definition by listing possible referents, see Section 5).

An example of such a classifier handshape from Czech SL is a basic, unmarked handshape – an open palm with all fingers stretched out. In the grammar part of this entry, the sign is categorised into its classifier group, whole entity classifiers, below which two meanings are listed: a denotation of either flat objects or four tired vehicles. Consequently, there are definitions and examples of use listed for each meaning separately; in this case a sentence where the classifier denotes a book in the former, and a car in the latter meaning.

Furthermore, there is another type of expression formerly thought of as classifiers: tracing Size and Shape Specifiers (SASSes), "which have a movement of the hand(s) that outlines an entity's size/shape" (Supalla, 1986). Although some linguists still include them into the

group of classifiers, the standard conclusion is to keep them separate, based on their multiple differing traits (presence of movement, not just a handshape; no need for antecedent, syntactically can function as nominals, adjectives or adverbs) (Zwitserlood, 2012). Specifiers are not incorporated into verbal predicates; instead, they compound with nominals and usually serve a modifying function (Sandler and Lillo-Martin, 2006b:77).

For a specifier to be registered as a separate entry in our dictionary, the same criteria apply as those for classifiers; a stabilised representative form with a roughly delimited meaning has to be attested.[5]

However, when classifiers and specifiers are part of lexicalised classifier constructions, they are automatically included in a dictionary and are treated as lexemes. In these structures, the otherwise productive forms become "frozen", their individual features (handshape, movement, location) no longer contribute morphological content to the given expression but bear only phonological status (Sandler and Lillo-Martin, 2006a). Among such constructions in Czech SL are, e.g., signs BOW ($\approx$ ARCHERY) and TREE, which originated by lexicalising a classifier, or YOGURT and OMELETTE, in which a former specifier is lexicalised.

4. Mouth patterns accompanying signs

Non-manual components of signs defined as "all linguistically significant elements that are not expressed by the hands" (Pfau and Quer, 2010) are equally important for speech comprehension and production as the manual articulators. These components can take the form of head and body movements, facial expressions, or mouth patterns. In this section, we will focus on the last type and assess which mouth patterns should and should not be documented in a dictionary.

Mouth patterns are commonly divided into mouth gestures and mouthings, differing in their relationship to the surrounding spoken language.

Mouthings (or spoken components) are either influenced or directly derived from the corresponding word in the spoken language; they are silent articulations of the whole word or a part of it, usually its first syllable (Pfau and Quer, 2010). Mouthings are understood as cross-modal borrowings (Sandler and Lillo-Martin, 2006a; Mareš, 2011), where it is possible to observe a gradual change and adaptation to the "host" language, a process typical for borrowings observed among spoken languages.

Mouth gestures (or oral components), on the other hand, are defined as "all motions/positions of the mouth that are not derived from a spoken language and contribute to the speech structure" (e.g., Mareš, 2011:8) and are therefore considered a native component of sign language.

In order for mouth patterns to be included in our dictionary, they need to satisfy two conditions: (i) they are obligatory for the given sign and (ii) they do not introduce additional meaning, they do not modify the sign in terms of intensification, adjectival or adverbial modification, nor do they express speaker's attitude (Mareš, 2011:24; Pfau and Quer, 2010:385). Condition (ii) does not apply to the mouth patterns (whether mouth gestures or mouthings) that have a disambiguating function between signs with identical or similar manual components, e.g., in the case of (a) two general signs: SALT vs. PEPPER, each with different mouthing of the corresponding Czech word *sůl* and *pepř*, respectively; (b) a general sign and a specific term: SUGAR vs. SACCHARIDE covered above in Section 2; and

---

[5] See the appendix for an example.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2020

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2020

(c) a noun and a verb: CAR vs. DRIVE, where CAR is articulated with the mouthing of the corresponding Czech word *auto*, while DRIVE is performed with a mouth gesture.

In the case of a single sign (conveying a single meaning) with variable mouth patterns available, the citation form is accompanied by the most neutral one, usually a mouth gesture. The other options are classified as variants of that sign. Any obligatory mouth patterns are given in the grammatical description for each meaning of the lexical entry (corresponding Czech word for mouthings and specialised symbols for different mouth gestures), and information about the association with another sign in a noun-verb pair.

5. Strategies of semantic definitions

So far, we have discussed what kinds of lexemes are eligible to be listed in a dictionary, but let us now turn to each lexical entry structure with a particular focus on their definitions.

The definition of a lexical entry is a crucial part of any monolingual dictionary. Thus it is of great importance to develop a firmly established method before beginning any lexicographic work, and to adhere to it throughout the process of compiling a dictionary. It can be especially challenging in the case of sign language dictionaries, where there is very little prior work to build on, and one may encounter several unprecedented issues.

In *Dictio*, we face these challenges with the help of precisely determining processes for forming each definition.

There are two main strategies for defining the meaning – intensional and extensional definition (Filipec, 1995). To define a lexeme intensionally means to specify necessary and sufficient conditions for the usage of a given lexeme. Such intensional definition has the following structure: first, the closest general term, a hypernym, is posited to categorise the lexeme into a broader semantic class; the next step is to list necessary distinguishing properties in order to differentiate the lexeme from other elements of the same semantic class. This way, we delimit all potential occurrences while ruling out other cases.[6] A nice example of the application of this general lexicographic strategy is the definition of the sign CD-ROM, which is given here in glosses and can be seen under the link: CD-ROM THAT CLF_ROUND_OBJECT SPC_THIN_OBJECT THAT SAVE DATA HOW CLF:DRAW_CIRCLES APPEAR SPC_LITTLE_HILLS 0 1 0 1.

Extensional definitions employ a different strategy, namely that they specify an extension of a given lexeme, e.g., by naming a typical representative or several objects that are members of a specific set, requiring the reader to extract the properties common to all listed examples and compile the meaning of the lexeme from them. Such a definition can be accompanied by qualitative or circumstantial properties of a concept, e.g., size, color, or application. An example is the semantic definition of the sign BLACK, which is given here in glosses and can be seen under the link: COLOR THAT LOOK_LIKE SUN GO_DOWN GET_DARK THAT.

Between the two strategies, it is always preferred in our dictionary to use the intensional definition. However, in sporadic cases, the meaning can be determined extensionally or by combining the two, i.e., by specifying a superordinate concept followed by several examples of referents.

6. Multiple meanings and semantic relations

In each lexical entry, the field of semantic relations includes both the intra-language relations (synonyms, antonyms), and the inter-language relations (translations). We will comment in detail on the first type, leaving the latter for another occasion. However, let us first consider the cases of polysemy.

In our dictionary, we follow the traditional practice of listing every meaning of a polysemous word under one lexical entry. These individual meanings differ, and therefore separate definitions, examples (and translations) are needed for them.

In principle, we have encountered three types of situations: (i) a general term with multiple meanings (e.g., GERMAN, which may stand for the country or a citizen of the country); (ii) a technical term with different meanings for their respective semantic fields of use (e.g., the sign BASIS with three different meanings – for the field of informatics, mathematics, and chemistry); and (iii) a sign with general and technical use. If the two forms are entirely identical – including the non-manual component – two meanings can be defined with the general one listed as first. However, more often, new mouthing is added during the creation of the technical term. In this case, we understand the non-manual component as a phoneme, and we register each sign under a separate entry.[7]

In *Dictio*, we register synonyms (expressions with identical or nearly identical meaning) and antonyms (expressions with the opposite meaning). A question closely tied to synonyms is how to distinguish them from variants (because, after all, both pairs share the same meaning) and classify synonyms and variants according to their semantic relationship to a given lexical entry.

For audio-oral languages, a dictionary entry standardly contains the citation form of a lexeme and all the variants (Čermák, 1995), e.g., the gender variants in Czech: *brambor* (potato-masculine) vs. *brambor-a* (potato-feminine). However, two (or more) expressions of a different word-forming nature are not considered variants but synonyms (Filipec, 1995), e.g., the Czech pair: *jazykověda* (linguistics; Czech origin) vs. *lingvistika* (linguistics; foreign origin).

What seems like a simple task for spoken languages (basically, common root signals variants, different roots – synonyms) becomes a challenge for sign languages while the discussion about the definition of morphemes and lexical roots is still open-ended (Zwitserlood, 2012). The lexicographic processing of the variants in sign languages has been addressed in the canonical work of Johnston and Schembri (1999) for AUSLAN, although the topic of synonyms is not elaborated.

---

[6] Since the key to intensional definition is to capture the internal hierarchy of a given semantic area, the work of Půlpánová (2017) on Czech SL becomes useful. In her thesis, she investigated the functional signs used for categorisation in Czech SL. Under *categorisation*, she understands the expression of hyper-

hyponymic relations in the lexicon. Such functional signs are, e.g., TYPE and GROUP in her elicited Czech SL expression ANIMAL + TYPE + GROUP + HOME (in the meaning of *pet*).

[7] See Section 4 above, namely examples SUGAR and SACCHARIDE.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2020

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2020

In *Dictio*, a method has been developed (and is now being applied) to distinguish variants from synonyms in Czech SL (with possible extension to other sign languages).

We propose to classify a pair of lexemes as variants in case their (possibly multiple) differing phonological features fall within only one of the three main categories (as defined in Sandler, 2006): handshape, place of articulation or movement. Variants altering within the place of articulation (location) are exemplified on a pair of lexemes COFFEE#1 and COFFEE#2, whereas PRAGUE#1 and PRAGUE#2 demonstrate variants with different handshape. On the other hand, the two signs for *kitchen* (see examples KITCHEN#1 and KITCHEN#2) present an apparent synonymy; the lexemes differ in all three main categories, and there is no doubt they do not share a morphological root. However, examples similar to MAY#1 and MAY#2 (which represent two forms from several variants and synonyms for *May*) are challenging since they present two morphologically related forms. Nonetheless, given that they differ in two of three main categories, namely handshape and movement, we conclude that they should be classified as synonyms. More complicated cases show that we are working with a scale rather than a binary distinction. However, the proposed solution appears to be extremely useful for the lexicographic task at hand.[8]

From a formal point of view, variants are presented in the grammatical or stylistic part of an entry, according to their status. A grammatical variant is a lexeme that is freely interchangeable with the citation form which does not add any extra information about the speaker. On the other hand, a stylistic variant adds such information about, e.g., social status, regional categorisation or a generation. However, synonyms are linked to individual meanings and not to the whole lexical entry because they may differ according to their semantic content.

7. Examples of use

The final section discusses examples, namely what kinds of expressions are appropriate for an example and what guidelines need to be followed when adding an example to an entry.

In the absence of Czech SL representative corpus, the examples of use are not elicited but created by the team of native signers, forming a small corpus by itself.

It is desirable to include at least one, but ideally, several examples are listed in each lexical entry, demonstrating the use of a given lemma in different communicative situations. An example could be an expression (two or more signs), a sentence, or an utterance (several sentences) illustrating the use of the lemma or its variants.

The fundamental idea of examples is to portray how lexemes are used in natural language. Therefore, it is not unusual to exemplify modification where possible, such as numeral and classifier incorporation, inflection of directional verbs, aspectual modification, and plural and negated forms.

As an illustration of the strategy described above, consider two examples for MONTH. The first example contains the simple citation form, the second one – a pluralized form with an incorporated numeral:
(i) TOMORROW MONTH MAY (video under the link),
(ii) SUMMER IN_THAT YEAR PERIOD HAVE FOUR

MONTH_THREE-redupl THAT SEGMENT IN_THAT JUNE 21 UNTIL SEPTEMBER 22 (video under the link).

8. Conclusion

Dictio is a work in progress, similar to any other dictionary trying to capture and describe natural language. However, even now, in its developing stages, it already serves multiple functions. *Dictio* has been used in Czech SL and linguistic education, providing valuable examples of signs and their categorisation. Moreover, it represents the biggest Czech SL material collection to date, making it the closest equivalent to Czech SL corpus available. Finally, *Dictio* poses many lexicographic challenges, and solving them brings us closer to understanding the nature of Czech sign language (among others) and its phenomena.

9. References

František Čermák. 1995. Paradigmatika a syntagmatika slovníku: možnosti a výhledy. In: F. Čermák in R. Blatná, eds., *Manuál lexikografie*, Vol. 1, pages 14–49. H&H, Jinočany.

Dictio: Multilingual Online Dictionary [online], Masaryk University, Brno [cit. 2020-03-27]. Available at: www.dictio.info.

Jordan Fenlon et al. 2015. The Lemma Dilemma Revisited. *International Journal of Lexicography*, 28(2):169–206.

Josef Filipec. 1995. Teorie a praxe jednojazyčného slovníku výkladového. In: F. Čermák in R. Blatná, eds., *Manuál lexikografie*, Vol. 1, pages 14–49. H&H, Jinočany.

Trevor Johnston in Adam Schembri. 1999. On Defining Lexeme in a Signed Language. *Sign Language & Linguistics*. 2(2):115–185.

Jette Hedegaard Kristoffersen in Thomas Troelsgård. 2012. The Electronic Lexicographical Treatment of Sign Languages: The Danish Sign Language Dictionary. In: S. Granger in M. Paquot, eds., *Electronic Lexicography*, pages 293–318. Oxford University Press, Oxford.

Jiří Langer et al. 2005a. *Znaková zásoba českého znakového jazyka k rozšiřujícímu studiu surdopedie se zaměřením na znakový jazyk (I, II)*. Palacký University, Olomouc.

Jiří Langer et al. 2005b. *Znaková zásoba českého znakového jazyka k rozšiřujícímu studiu surdopedie se zaměřením na znakový jazyk (III, IV)*. Palacký University, Olomouc.

Jiří Langer et al. 2008. *Slovník vybraných pojmů znakového jazyka pro oblast biologie člověka a zdravovědy*. Fortuna, Praha.

Jindřich Mareš. 2001. *Orální komponenty v českém znakovém jazyce*. Bc. thesis, Charles University.

Rachel McKee in Mireille Vale. 2017. Sign language lexicography. In: P. Hanks in G. M. de Schryver, eds., *International Handbook of Modern Lexis and Lexicography*. Springer, Berlin/Heidelberg.

Pavlína Mladová. 2009. *Kompozita v českém znakovém jazyce*. Bc. thesis, Charles University.

Roland Pfau in Josep Quer. 2010. Nonmanuals: their grammatical and prosodic roles. In: D. Brentari, ed., *Sign Languages*, pages 381–402. Cambridge University Press, New York.

---

[8] Our approach builds on the phonological Hand-Tier model of Sandler (2006) and contributes a set of clear criteria for distinguishing variants from synonyms. However, a more detailed discussion and formal analysis of the Hand-Tier model and its modification are beyond this paper's scope.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2020

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2020

Miloň Potměšil et al. 2002. *Všeobecný slovník českého znakového jazyka, A-N.* Fortuna, Praha.

Miloň Potměšil et al. 2004a. *Všeobecný slovník českého znakového jazyka, O-Ž.* Fortuna, Praha.

Miloň Potměšil et al. 2004b. *Všeobecný slovník českého znakového jazyka, O-Ž - doplněk.* Fortuna, Praha.

Lucie Půlpánová. 2007. *Kategorizace v českém znakovém jazyce.* Mgr. thesis, Charles University.

Wendy Sandler. 2006. Phonology. In: W. Sandler in D. Lillo-Martin, eds., *Sign Language and Linguistic Universals*, Vol. 1, pages 111–271. Cambridge University Press, New York.

Wendy Sandler in Diane Lillo-Martin. 2006a. Entering the lexicon: lexicalization, backformation and cross-modal borrowing. In: W. Sandler in D. Lillo-Martin, eds., *Sign Language and Linguistic Universals*, Vol. 1, pages 111–271. Cambridge University Press, New York.

Wendy Sandler in Diane Lillo-Martin. 2006b. Classifier constructions. In: W. Sandler in D. Lillo-Martin, eds., *Sign Language and Linguistic Universals*, Vol. 1, pages 111–271. Cambridge University Press, New York.

William C. Stokoe. 1960/2005. Sign language structure: an outline of the visual communication systems of the American deaf. Reprinted in *Journal of Deaf Studies and Deaf Education*, 10:3–37.

Ted Supalla. 1986. The classifier system in American sign language. *Noun classes and categorization*, 7:181–214.

Ulrike Zeshan. 2003. Towards a notion of 'word' in sign languages. In: R. Dixon in A. Aikhenvald, eds., *Word: A Cross-linguistic Typology*, pages 153–179. Cambridge University Press, Cambridge.

Ulrike Zeshan. 2004. Interrogative and Negative Construction in Sign Languages. *Language*, 80(1):7–39.

Inge Zwitserlood. 2010. Sign Language Lexicography in the Early 21st Century and a Recently Published Dictionary of Sign Language of the Netherlands. *International Journal of Lexicography*, 23(4):443–476.

Inge Zwitserlood. 2012. Classifiers. In: R. Pfau, M. Steinbach in B. Woll, eds., *Sign Language. An International Handbook*, pages 158–186. De Gruyter Mouton, Berlin/Boston.

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2020

Conference on
Language Technologies & Digital Humanities
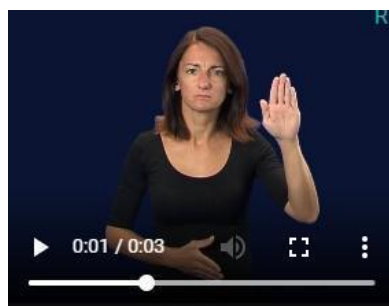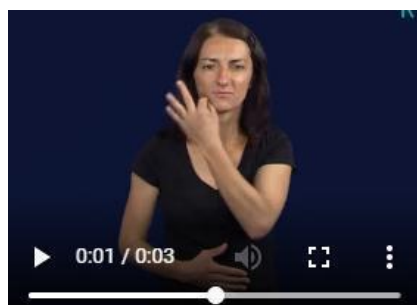Ljubljana, 2020

10. Appendix

BASIS

BLACK

BOW / ARCHERY

CD-ROM

classifier

COFFEE#1

COFFEE#2
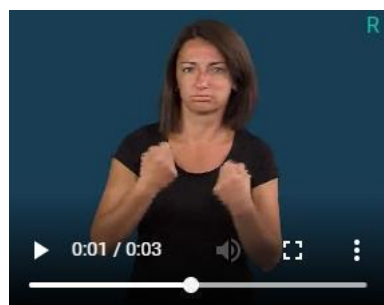
DRIVE / CAR
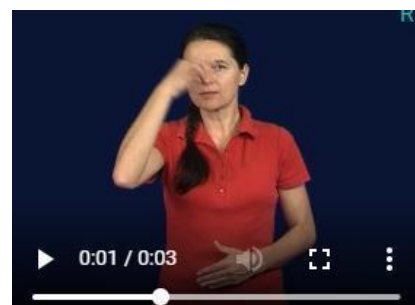
GERMAN / GERMANY

HEAR

HOUR

ING

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2020

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2020

KITCHEN#1
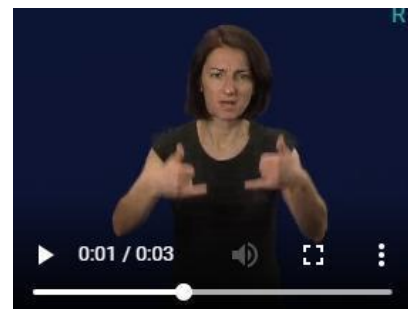
KITCHEN#2

MAY#1

MAY#2

OMELETTE

POST_OFFICE (initial position)

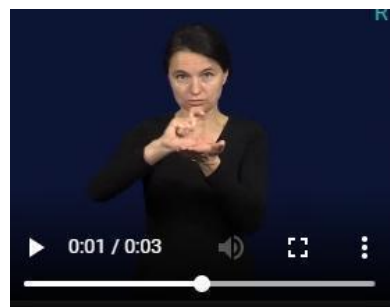POST_OFFICE (final position)
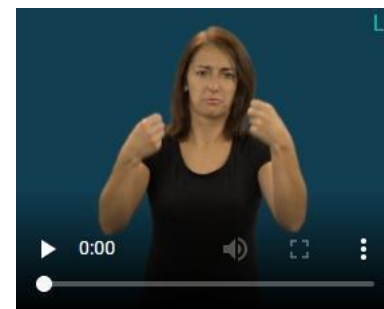
PRAGUE#1

PRAGUE#2

specifier

SPICE / SALT / PEPPER

SUGAR → SACCHARIDE

TREE

YOGHURT