Authentic Examples in a Corpus-Based Sign Language Dictionary – Why and How

Gabriele Langer, Anke Müller, Sabrina Wähl, Julian Bleicken
University of Hamburg
E-mail: gabriele.langer@uni-hamburg.de, anke.mueller@uni-hamburg.de, sabrina.waehl@uni-hamburg.de, julian.bleicken@uni-hamburg.de

Abstract

Within the DGS-Korpus Project, a corpus-based dictionary of German Sign Language (DGS) is compiled. Dictionary entries describe signs, their meanings and uses as they are reflected in the corpus. The dictionary entries include authentic examples taken directly from the original corpus recordings. Without a functional writing system for sign languages (SL), corpus building as well as SL usage examples in dictionaries have to resort to videos as representations of SL use. Examples can either be taken from the original corpus material (authentic examples) without the option of even mild editorial changes, or they can be re-recorded with a different signing model. While the latter allows for editing as well as constructed examples, it also entails very drastic changes to the appearance of the original authentic examples they are based on.

In the article our reasons for inclusion of authentic examples are discussed and criteria for example selection listed. To compensate for the challenges that authentic examples removed from their original contexts entail, translations and context information are added to the entries. The practical steps for example preparation, namely selecting the segment, providing context information and adjusting translations are described.

Keywords: authentic examples, sign language dictionary, spoken language (as opposed to written), corpus-based lexicography

1   Situation of the DGS Community

Sign languages are visual and spatial languages. They make use of the movement of both hands, the head and the upper body as well as facial expressions, movement of lips, tongue and cheeks, eye gaze and eye blinks. Spatial modifications of the sign form are used for semantic and syntactic purposes. Sign language is predominantly used in face-to-face interaction, that is in dialogic situations of a shared environment and the spatio-temporal co-presence of the interlocutors. Signed languages do not have a widely used functional writing system. This is mainly due to the difficulties of reducing a multi-modal language relying on situative embedding to a few abstract symbols, leaving iconic descriptions, facial gestures and indexicality open to the imagination of the addressee.[1]

Systematic situative embeddedness and the absence of a functional writing system make signed language in many respects comparable to spoken language (as opposed to written language)[2] that is used in conditions of everyday conversations. The similarities concern syntactic and semantic processes

---

[1]   There are quite successful initiatives to develop and use an iconic writing system known as Sutton SignWriting (cf. Sutton's SignWriting Site) with a growing international community of signers using it (for Germany see Wöhrmann 2003). This writing system might be an adequate system, but it is not yet commonly known. Therefore, it is not suitable to be used for a description of sign senses in the DGS dictionary.

[2]   In the field of sign language linguistics, the term spoken language is often used in contrast to sign language, irrespective of its medial appearance, which is of importance in this paper. For that reason we refer to this language type as vocal language. Vocal languages typically occur in two medial forms, spoken or written, which both have different properties.

including vagueness and fewer explicit references, and all aspects of on-line production of utterances, such as disfluencies and dynamic changes of plan (Ebbinghaus 2001, Schwitalla 1997).

Sign languages are used in communities of deaf, hard-of-hearing and associated hearing persons, only a minority being children of deaf parents. The majority of congenitally deaf people have hearing parents and are exposed to sign language only at preschool and school. Sign language users constitute a linguistic and cultural minority within a larger community of a vocal language with its spoken and written forms of communication. This has some impact on the social conditions of sign language use in a structural language contact situation. Signers are functionally bilingual "as a result of growing up in a hearing world and receiving education in (at least) spoken language" (Bank et al. 2016: 1284; cf. Ebbinghaus 2001 for DGS), using the surrounding vocal language for written communication. Also, in DGS and other SL spoken language elements, so-called mouthings are integrated into the signing. Mouthings are visible gestalts of German words, often accompanying content signs for denoting objects, events or concepts (cf. Ebbinghaus & Hessmann 2001). Many signs can be combined with several mouthings and mouth gestures, the denotation thus being specified in a dynamic way. This is facilitated by the underlying iconicity of signs, resulting in a lexicon with many highly polysemous signs.

Sign language can be regarded as a structurally spoken language with respect to the means of communication in the spatio-temporal situatedness of co-presence (cf. Fehrmann & Linz 2009). The structure of a sign language is largely determined by these medial contexts of its use, as is the case with spoken or written language. If texts are defined as the result of a temporal dislocation of utterance production and reception (Ehlich 1983), the textual situation leads to a de-contextualisation of language. The beginning of such a process of possible changes in the way language is used can also be observed within sign languages in planned recordings, addressed to a general audience and with editing techniques available (cf. Krentz 2007, Linz & Fehrmann 2009). Video enables the production of sign recordings as texts, and also the conversational use across places with camera devices in portable phones or the internet (cf. Keating et al. 2008 on medial factors influencing linguistic structure). Though we can talk of signed texts in the above definition, these are not written; and they are bound to bodily appearance.

2   Lexicography of Sign Languages

The lexicography of sign languages has been shaped by three major circumstances: First, SL are minority languages surrounded by and embedded in societies with at least one dominant vocal language that is also used for written communication. Second, due to their unique visual-spatial nature there is no functional writing system available to write signing. This directly results in a lack of written sources available to study sign language use, and also has major implications on how to represent sign language in a dictionary as well as to what extent sign language is chosen as the metalanguage in SL dictionaries – or not. Third, being a young field of study, general SL research has still not arrived at a general agreement on the basic categories, structures and properties of the languages.

2.1   Sign Dictionaries

Serious sign dictionaries[3] contain sign entries that aim at a description of the sign's meanings and uses in their own right from a basically monolingual perspective. However, without the option to write sign language texts no sign dictionary – as far as we know – has tried to completely do

---

3   We do not discuss here sign dictionaries of the type known as sign collections. These are in essence simple bilingual but mostly uni-directional word-to-sign lists aimed at hearing users, and do not contain comprehensive sign entries which describe signs and their uses in detail. They also do not include signed example sentences.

without a vocal language as its metalanguage. At least front and back matter, headings and indexes such as subject fields for a thematic access are given in the surrounding vocal language. Almost always the written vocal language is also used as metalanguage for the description of meaning (dictionary definition), and further information on the signs' usage and grammar.[4] Also, all larger SL dictionaries include translational equivalents of the lemma sign into the vocal language, even if the product is not primarily designed to be a bilingual dictionary. Translational equivalents are either given as an indication of meaning or in addition to the dictionary definition. The rationale here is that not only learners but also native signers – who are functional bilinguals – prefer to have an access via the surrounding vocal language.

Sign dictionaries also tend to provide a way to find signs by searching for their form. As far as we know there are no dictionaries of a signed and vocal language pair that are truly bilingual in the sense that both languages receive the same amount of attention and display the same complexity, thoroughness and depth of description for both languages. As such, existing sign language dictionaries are always some form hybrid dictionary between monolingual (with the focus of description on the signs their meanings and uses in the entry structure) and bilingual (with the inclusion of translational equivalents and bidirectional access). One example of such a hybrid dictionary is the ODT-SL, described by Kristoffersen and Troelsgard (2010: 1550) as follows: "As a result of this decision, the Danish Sign Language Dictionary could be described as monolingual dictionary, which instead of definitions has (searchable) equivalents in another language, Danish, which is also the general metalanguage of the dictionary."

Because of the lack of a functional writing system it is very difficult to present complete sign utterances on paper (as can be seen looking at example 4). Therefore, printed sign language dictionaries usually do not include real SL example utterances. Sometimes they include information on typical semantic contexts of a sign's use – as a substitute for usage examples – in the written vocal language (cf. for example D-SAS). In electronic or online SL dictionaries example sentences are usually presented in the form of video clips, and almost always an additional translation into the vocal language is provided. Sometimes either a gloss transcription and or a gloss-like literal translation of each sign of the example utterance is added in order to visualize the sequential sign order and to support easy, clickable cross-referencing to the entries of the other signs in the example sentence (see for example ODT-SL for glosses and ONZSL for a gloss-like literal translation of each sign). Normally the usage examples in sign dictionaries are studio recordings, and as far as we know we are the first dictionary project to include example sentences directly taken from the filmed original corpus data.

2.2   Corpus-based Lexicography of Sign Language

Without the availability of written texts in sign language, research has to rely on recordings of signing as permanent representations of language use to be investigated and analyzed. Only recently has technological progress in recording and storing signed data, and in annotating and retrieving this data more efficiently by the way of annotation tools, made the collection and use of sign corpora of a considerable size feasible. However, the idea of corpus-based SL lexicography has been around for quite a while. The first, very impressive attempt to collect and use a corpus for SL lexicography was made in the D-ASL despite technical limitations of that time (cf. Stokoe 1993). Other general SL dictionary projects, such as the D-NZSL and the ODT-SL, included data collection sessions. The recordings were reviewed and selectively tagged – in the case of D-NZSL for signs expressing concepts of a pre-defined list (cf. Kennedy 1996), and in the case of the ODT-SL for sentences to be used as models for re-recorded examples (cf. Kristoffersen 2010). Both projects combined the analysis of

---

4   One exception is the D-LSFB. This online dictionary includes videos with signed texts for etymology explanation and dictionary definitions, and it also includes recorded competence examples.

some corpus data with the intuition of consultant groups and editors for the description of lemma signs in dictionary entries.

The KS-PJM is based on a large corpus of Polish Sign Language (PJM) collected between 2011 and 2016. The corpus was used for lemma selection and corpus data was used in the preparation of the entries. Gaps of signs missing the corpus were filled and additional unattested meanings/uses of corpus signs were added by the editors. Examples were drawn from the corpus and re-recorded in the studio (cf. Linde-Usiekniewicz & Rutkowski 2016: 377-379).

Up to now all SL dictionaries that worked with corpus data also drew extensively on the intuitions of the editors and consultant groups to supplement the findings. With SL corpus data just becoming available in considerable sizes the field of corpus-based lexicography of SL is very new and remains in the process of forming. Methods, processes and presentational formats are largely still experimental or only just developing, and standards have not yet established in the field. As more and larger SL corpora become available, corpus data is increasingly used in dictionary projects for lemma selection, the discovery and description of sign meanings and uses, and as a source of usage examples.

3   DGS-Korpus Project

The DGS-Korpus Project is a long-term project (2009-2023) of the Academy of Sciences and Humanities in Hamburg, Germany. Its central aims are the collection of a corpus, making parts of it available as the Public DGS Corpus to the language community and researchers alike, and the development of a corpus-based dictionary for DGS-German.

3.1   Data and Annotation

The filming sessions for the corpus data collection took place between 2010 and 2012 in 12 different German cities. The sample is balanced for region, age and gender. A mobile studio was used to film the 330 informants participating in this project. Informants from the same region were filmed in pairs with a moderator leading through the sessions. The data collection sessions were designed to cover a wide range of different topics. Tasks to initiate signing on different topics included open conversation, narrations of life experiences or events witnessed and retellings of stories (Nishio et al. 2010). The data collected reflects the properties of sign languages in typical face-to-face interactions, and is of structurally spoken nature. In the corpus, there are no discourse or text types implying registers of social distance, as would be seen in public talk, working instructions or legal texts. And there are no signed texts in the sense of spatio-temporal dislocation or signing through media for bridging spatial distance.

Data collection resulted in nearly 560 hours of signed material. The data is tokenized, lemmatized and annotated. For sign languages this has to be done manually. Lemmatization is achieved by matching tokens to types that are identified by unique glosses[5] and HamNoSys notations describing the signs' forms. Lemmatization and annotation is still ongoing. Already available for lexicographic analysis are 66 hours of the material with a completed continuous basic lemmatization[6] and annotation (May 2018). A very large part of the corpus has been translated. Translations are time-aligned with the signing, and thus can be searched for specific concepts even in parts that have not yet been lemmatized. In some cases, relevant signs, chunks or smaller passages have been selectively token-type-matched and annotated in videos that remain unlemmatized (so-called spot transcriptions). The current corpus size,

---

5   In the underlying database types are identified internally by an unchangeable ID-number. Glosses are bound to that ID. Any change concerning the gloss or the HamNoSys notation of the type will automatically affect all tokens that have been linked to that type.

6   Lemmatization here is token-type-matching and an important part of the basic annotation.

including all tokens, is approximately 480,000 tokens (May 2018). The DGS corpus is the basis for the work on the corpus-based dictionary. Compared to the size of written language corpora containing billions of tokens, the DGS corpus is fairly small. Nevertheless, the token count is sufficient to start with the description of the more frequent signs.

3.2   The DGS Dictionary

The DGS dictionary is one of the first corpus-based dictionaries of an SL. The DGS corpus has been collected to serve as the basis for lexicographic description of signs. The corpus and dictionary are to be interconnected. The dictionary is descriptive, with no intention of standardization.

One of the guiding principles at the present time is that dictionary entries contain only sign forms and sign meanings/uses that are attested in the corpus.[7] These are found in the corpus by summarizing views on the corpus data and looking at the original occurrences of a sign in context (cf. Langer et al. 2018). Also, with a few very particular exceptions, all examples shown in the dictionary are taken directly from the corpus. This approach can be described as corpus-based, but at the same time also as corpus-bound. We consider it rewarding to explore the possibilities and limits of this approach to base all essential information on signs on "objective evidence of language in use", as Atkins and Rundell have put it (2008: 53). However, considering the relatively small size of our corpus we are aware of the limitations of this approach, and might decide to move beyond a strict corpus-bound method at a later stage. If we do, we feel it would be necessary to clearly mark information that has been added from other sources than the corpus.

As a consequence of the corpus-bound approach, we have corresponding corpus evidence for all variant forms and senses of a sign included in the entry, and thus can retrace the abstracted information to the original data. It is planned to link and cross-reference between the dictionary and the Public DGS Corpus, where the authentic usage examples can be viewed in a wider context and further occurrences of the lemma signs can be found.

Just like other sign language dictionaries, the DGS dictionary will also be a hybrid type with some properties of a monolingual and some properties of a bilingual dictionary.[8] The DGS dictionary will have to serve very different user groups, such as native signers, advanced learners and beginning learners. The different information types in the entries have different functions for these groups, and not all information types are addressed to all groups in the same way. It will be a monolingually oriented dictionary in that it primarily describes the language units of only one language, that is DGS. Full entries with a description of forms, meanings, uses and distributions will only be listed for DGS signs. The listing and description of the senses of a sign is in the first step described as independently as possible from the structure of the second language (German). However, written German is used as a metalanguage for any written information, such as the description of the senses. In practice, the dictionary can be used as a bilingual dictionary. For each sense listed there are one or several German equivalents shown serving the direction of use from the source language DGS to target language German. The dictionary does not have detailed information on the German equivalents. However, for bilingual native signers providing a simple equivalent will often be enough to fulfill their informational needs and, if not, it can serve as a starting point for the search in a monolingual German dictionary.

---

7   In other dictionary projects corpus data has been supplemented by the intuition of the editors and SL consultants for missing signs and senses (cf. for example Stokoe 1993: 132; Kennedy 1996: 37-38; McKee & McKee 2013: 517-518; Linde-Usiekniewicz & Rutkowski 2016: 377). For the user of these dictionaries it remains unclear which parts and information of the resulting dictionary entries are actually derived from corpus data evidence, and which parts are additions made on the basis of intuition.

8   Erlenkamp has mentioned that for an adequate description of a language a monolingual dictionary is the better choice, while for a minority language in a majority society it is useful and politically more supportive to produce a bilingual dictionary (1998: 102). A hybrid form combines both requirements in a useful compromise.

The German equivalents will be searchable. Access by German equivalents is helpful for learners of DGS, but also for bilingual native signers to search for sign entries via vocal language words (source language German to target language DGS). Also, a search function via the sign form will be provided, as well as a thematic access to the sign entries.

Entries will be continuously published until 2023, with revisions published as necessary. At this point in time, structure, labeling and layout of the entries is still experimental and allows users to view and discuss the information types and contents of the first entries.

One important decision on layout function is to not represent lemma signs by glosses.[9] While glosses are very convenient for internal workflow processes, they can also lead to a confusion of languages and misleading assumptions on the signs' properties as inferred from the gloss words' properties by non-expert users. All DGS elements in the dictionary entries are represented as video (or a play button indicating that a video is available), or as a micon (moving icon) in combination with an identification number. The micon represents the lemma, while the number provides a unique sortable and searchable label in cross-references, and supports the discrimination of signs with a similar form.

3.3   Structure of Preliminary Entries

Currently a preliminary entry page (see Figure 1) is divided into two main areas.[10] A fixed display window is placed on the left side. It is used to display any kind of signed information located at different places within the entry. On the right side all information on the lemma sign is displayed. At the moment the entry is organized in the form of a table. At the top a unique number identifies the entry and the lemma sign. The head of the entry contains information on the sign form, possible sign variants and further comments on the sign and its grammar. Play buttons in the row labelled "form" can be used to start the video showing the chosen form. The main body of the entries is a list describing the different senses. Here you find sense-related information on, for example, mouthings, definitions, German translational equivalents, signed usage examples and possible synonyms and antonyms and collocational patterns.

Authentic examples taken from the DGS corpus illustrate each sense. By using the play button the DGS utterances may be viewed in the display window. Next to the play button, context information (text in square brackets) and a German translation of the example is given (with the sign's translational equivalent in bold face).
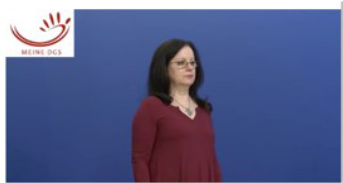
All cross-reference signs (e.g. synonyms and antonyms) are represented by micons and identification numbers. When the mouse is placed over an icon it starts to move. A click onto the micon will show the sign in the display window. If the entry for the cross-reference already exists, the number background is red and a click on the number changes the screen to the corresponding entry. Information on sign combinations and related or similar signs is given at the bottom of each entry below the sense listing.

In the preliminary entries, information given on DGS in DGS is the lemma sign and its variants, all synonyms, antonyms and other cross-references, collocational patterns and the signed examples. Usually two examples per sense are given. Thus the user is given the possibility to draw conclusions

---

9   A gloss in the world of sign language research and teaching is a vocal language word used as a label for a sign. Such labels – usually written in capitalized letters – make it easy to speak about signs, represent signed utterances in a written form, to sort and search for sign entries electronically, and to order them alphabetically according to their gloss. Often numbers or other markers are added to the gloss word in order to distinguish different signs and/or to mark morphological features. Usually the gloss name is a word that corresponds in one of its meanings to a core meaning of the sign, and can therefore be mistaken for a translation. In annotation sign glosses are used as unique labels for signs in the token-type-matching.

10   The entry in Figure 1 does not include all of the following information types.

from the examples. All these information types support the differentiation of senses. Translational equivalents and the translations of the signed examples are bilingual elements of the dictionary, and are in the target language, German. German is used as metalanguage for the context on examples as well as the definitions.[11]

Figure 1: Structure of a preliminary entry

4   Authentic Examples

Atkins and Rundell (2008: 453-355) list the following most basic functions of examples in a dictionary entry: to attest the existence of a word or meaning, to elucidate the meaning by complementing the definition and clarifying the sense, and to illustrate contextual features such as syntactical patterns, collocation, and the like.

One of the basic decisions when planning a dictionary is whether to work with constructed or authentic examples. Prinsloo and Gouw (2000) discuss advantages of each of these and sketch a continuum between constructed editorial examples, heavily or slightly altered examples from a corpus, and unaltered authentic examples. Atkins and Rundell (2008: 457) note that in practice examples are usually abstractions from recurrent patterns found in a large number of corpus data, with adjustments to fulfil

---

11   Though definitions could be given in sign as videos, this will not be done in the DGS dictionary. Signed definitions as a text type are not established yet, and developing and recording them would require a lot of resources. Second, as videos progress in time they vanish quickly and cannot easily be surveyed. The mnemonic function of writing is not fulfilled by filmed definitions. Moreover, video clips take up a lot of space. The same considerations hold for the description of example contexts.

the criteria of good examples. For written languages constructing or editing examples is therefore more a matter of degree than a clear-cut choice.

In the case of sign language examples, even a small editing adjustment results in a full re-recording, and thus a completely new appearance of the example and the loss of much of its original character. So for sign languages the question of authentic vs. edited or constructed examples is a fundamental decision. Here sign language and spoken language lexicography face similar challenges with respect to some problems and difficulties arising from their structural characteristics as face-to-face communication systems (Verdonik & Sepesy Maučec 2017). There are initiatives in spoken-language research to build corpus-based dictionaries or to add spoken-language specifics to existing dictionaries that use written texts as their base (cf. Verdonik & Sepesy Maučec 2017: 147-148). Auditory examples along with a written transcription and contextual information can be found occasionally in the ODT-VL on spoken language elements (Hansen & Hansen 2012: 930). Möhrs et al. (2017: 282) intend the integration of "multimodal information, such as corpus-based audio-examples and transcriptions for each entry" in their planned lexicographic resource of spoken German. However, these cases remain exceptional. Trap-Jensen (2004) also argues for taking authentic examples of spoken language, even if they are transformed into the written modality as is the case with the Danish dictionary. Though it may be "impossible to excerpt readable quotations in unaltered form from the spoken language corpus" (Trap-Jensen 2004: 316), because features intrinsically bound to the spoken medium are lost by that transformation, it may be worthwhile including authentic spoken examples. To convey the original quality as far as possible, all editorial changes to "authentic quotations with the source of the origin" like omissions, additions and reformulations shall be indicated as in a critical edition (Trap-Jensen 2004: 317).

Editing examples is useful for the reason of intelligibility, and distracting performance factors in particular can be eliminated with this. However, the DGS dictionary will include authentic examples taken from the corpus. As our data is filmed signed interactions, the options for presenting usage examples are, in principle, to use a segment of the original data recording, to re-record examples with a signing model, or to produce an animation to render the corpus example. Re-recording and animation[12] both allow for editing – with respect to content as well as to form aspects, while anonymization is also facilitated. However, we have good reasons to give priority to authentic examples as signed in the corpus, as outlined below:

- Re-recording an example – with or without editing – is very labor-intensive and the result is prone to lose the performative aspects of the original. Signed language production is not detachable from the body. Gestural and expressive aspects form integral parts not only of the utterance, but also of the language system (Fehrmann & Linz 2009). Performativity is an essential feature of SL and is, as such, not fully reproducible. Re-recordings are thus bound to lack important information.
- It is not entirely clear which aspects of signed utterances are to be preserved in reproduction; see for example the difficulty in separating affective and grammatical functions of facial expressions (cf. De Vos et al. 2009 on eyebrow movement).
- DGS exhibits a lot of regional variance. When re-producing examples from different regions a signing model would have to reproduce the signing in the examples very closely to ensure consistency with regard to regionality. The model would thus have to execute many signs that they might not use themselves, a task which is very difficult to do naturally, especially in a studio setting.
- The dictionary will be closely linked to the freely accessible Public DGS Corpus (cf. Jahn et al.

---

12   Animations that can be generated automatically from signed input data may be a future device to compensate for the lack of a writing system, for granting anonymity and the possibility of editing, but are to date not yet available. We do not dig further into this issue, as all arguments listed here are independent of the techniques of how occurrence examples are reproduced.

2018), which is a valuable language and cultural resource of DGS. Using corpus examples in the dictionary is a way of appreciating this resource as well as appreciating the contribution of the informants to the project, and exploring further the possibilities of its use. The dictionary can thus contribute to the further recognition of DGS within the surrounding German society (cf. Erlenkamp 1998: 99). From the entries there will be a direct access to the source data. We regard it as an advantage that, in general, examples can be traced back to their source conversation with all background context to be viewed. Because the informants have given their informed consent, there is no general need to re-record all examples for reasons of anonymization.

- Including authentic examples signed by many members of the signing community may boost the interest in and the acceptance of the dictionary, as people can browse through the product and recognize some of their relatives, friends or other people they might know. Authentic examples of many signers are certainly livelier and show a larger range of signing styles than plain studio recordings of only a few signing models.
- In a poly-functional dictionary different user groups should be served. Authentic examples are directed more at native signers and advanced learners than novices.

We want to stress the central argument for including authentic examples. The DGS-Korpus project aims at language documentation, and this should also especially apply to the choice and properties of the examples. Editing examples means to basically change the language found in the corpus, which is shaped by face-to-face interaction. This means making utterances normally embedded in a discourse context self-contained, making the arguments of the verb explicit, avoiding digressions and – for the sake of having as clear example – removing superfluous information and redundancy. Although the reception of examples may be improved, the editing process results in the well-known text type called an example sentence, not a product of natural conversations – and it is natural conversation we want to document.[13]

5   Authentic Examples in the DGS Dictionary: Practical Experiences

Several dictionary projects include editorial examples based on corpus examples. Their editing process follows guidelines that have been created to produce examples suitable for the dictionary user (e.g. McKee & McKee 2013, Kristoffersen 2010). Editing includes actions such as changing the corpus utterance into a self-contained sentence by clarifying unresolved references, removing distracting elements and modifying the content for reasons of political correctness or balancing subjects.

In the DGS dictionary we include authentic examples from the original recording. This entails a number of consequences, because editing is not possible at all. Perceived shortcomings in the original recording have to be balanced by additional information, e.g. context information to ensure understanding of the example even in isolation from its context. Disturbing elements of performance (e.g. lax, disrupted, too fast signing) have either to be tolerated or lead to non-selection. As the choice of possible examples is limited to the corpus, occurrences using authentic examples is always a compromise between naturalness and accepting perceived imperfections. However, by means of careful selection, choosing the most suitable segment of the utterance and adding information (context, translation), it is possible to provide the user with useful and interesting examples.

---

13   One of our elicitation tasks asked for isolated signs known for regional variation, the use of which should be exemplified. When asked, the informants produced sentence-like, self-contained examples, often also useful evidence of signs other than the target sign. These sentences may fulfil the properties of a good example, but they may also appear artificial and similar to constructed examples.

5.1   Selection of Examples

Authentic examples are selected to best support the dictionary users' understanding of the senses. The following criteria have proven useful to achieve a good selection from the possible example candidates.

5.1.1 Clear Illustration of Sense

Atkins and Rundell (2008: 454) state that one important function of examples is to "illustrate usage", and thus complement the more abstract definitions and help "clarify sense distinctions in a polysemous word". Sometimes one utterance may cover more than one sense – depending on the interpretation – while another example illustrates only one specific sense clearly in contrast to the others. Examples that unambiguously illustrate one specific sense should be chosen. Moreover, examples should complement and support the definition by providing at least some semantic information on the sign's meaning in its surrounding context. The user should be able to infer from the examples some information about the meaning of the sign and typical contexts of its use in this specific sense. Synonyms or antonyms in the example utterance provide important clues on the sign's contextual meaning, and help to distinguish the sense more clearly. Utterances containing synonyms or antonyms are preferred to be chosen as examples, if available.

5.1.2 Clarity of Sign Execution

For the dictionary user the examples need to be understandable, and should be easy to perceive. Examples taken directly from a corpus may contain unclear and partially superfluous signing, such as false starts, or signs that are only indicated or executed in an unclear or idiosyncratic way. There may also be side remarks that distract from the relevant parts of the example. This and disfluencies in signing, e.g. breaks or searching for signs, can make it difficult to understand the content easily. Moreover, signing that is too quick can disturb the dictionary user, as the target sign may be seen too briefly to perceive it easily. All the previously mentioned difficulties should be avoided as much as possible, and examples with clear and undisturbed signing are preferred.

5.1.3 Inclusion of a Wide Variety of Informants: Balancing for Region, Age, and Gender

Since the corpus contains a lot of different peoples' signing this diversity should be reflected in the selection of examples. In the dictionary, the overall selection of examples is aimed to include all regions, age groups, both genders and as many informants as possible. When a particular sign's use is evidenced only for one region or mainly for a certain age group, the example selected for that sign will reflect this restricted use. No group should be overrepresented without need, as the dictionary is to be seen as a mirror of the data collected. Showing original informants avoids having to consider which sign model can best represent a certain region or group.[14]

5.1.4 No Stereotypes, Discriminating Content and Avoiding Personal Information

The contents of an authentic example should not be discriminating or reflecting stereotypes or prejudices. The non-selection of an example could also be a way of supporting a stereotype; for example, when only selecting examples occurring in the context of heterosexual partnerships for the sense being described as a "male partner in a relationship" when there is also a good example in the context of a homosexual relationship, like example (1).

---

14   The ONZSL dictionary project used a group of eight sign models. They took special care to match the model's social factors to the language use of the original example (McKee & McKee 2013).

(1) ▶ [Über Gunther Trube] Seine Mutter und sein **Lebensgefährte** haben einige Leute eingeladen.

Translation: [About Gunther Trube, a Deaf celebrity] His mom and his **husband** had invited some people.

Some informants reveal personal information about their lives or talked about people they know. Even though the informants have consented to the publication of their signing, we protect their and other people's privacy when needed. Examples containing very personal information are usually not selected. If it is not possible to do without a particular example with personal information, we actually re-record the example to achieve anonymization. This is one of the very rare cases where we do not show the original material. In example (2) we also replaced the real name with a very common German family name in the re-recording. This is a different case to that shown in example (1) with Gunther Trube, who is a well-known and even famous figure within the deaf community.

(2) ▶ [Über eine Lehrerin] Ich mochte **Frau** Meyer, weil sie gebärden kann.

Translation: [About a teacher] I liked **Mrs.** Meyer, because she could sign.

5.1.5 Handling of Citation Form vs. Contextually Modified Forms

In use, signs do not always appear in their citation form, but are modified according to syntactic and morphologic rules, and thus occur as different word forms of the sign. Many signs can also be modified according to their iconic characteristics – a type of modification where the forms the sign can possibly take are not fully predictable. The sign form shown in the examples need not necessarily be of the citation form, on the contrary – a variety of forms that are typical of the lemma sign are preferred, as that shows the patterns of its use. This is valuable information for L2-learners of the language, to see how a sign can be spatially or iconically modified. Sign 7 in example (3) is a directional verb indicating the arguments of the verb by its directional execution. In citation form, the sign is signed from the signer's front towards themselves (in the sense of "someone gives back to me"); in example (3) it shows a modified form meaning "I give it back to *previous location [here: the shop]*".

(3)

	1	2	3	4
contextual meaning	mostly	I	glas	bottle
mouthing	--------------meist--------------		-------------glas--------------	

	5	6	7	6	9
contextual meaning	*pointing*	can	return	can	*gesture*
mouthing		glas kann	wieder zurück	--------------kann--------------	

▣ [Thema: Umweltschutz und Pfandflaschen. Die Erzählerin kauft lieber Getränke in Glasflaschen als in Tetrapackverpackungen.] Ich nehme meistens Glasflaschen, die kann ich wieder **zurückbringen**.

Translation: [Subject: environmental protection and returnable bottles. The signer prefers glass bottles to Tetrapack carton packages when buying beverages.] Usually I take glass bottles because I can **return** them.

5.1.6 Treatment of Variants

DGS is an SL that shows many phonological variants. Usually such variants are described in one entry, thus one variant needs to be chosen to function as the lemma. This does not mean that examples containing other variants are avoided. It is instead the case that we aim to show a balanced selection of examples according to the variants' frequency. Sometimes authentic examples containing less frequent variants show the sense more clearly than possible examples containing the main variant of the lemma sign. In this case they are preferred over examples containing the main variant. Showing different variants also corresponds to the aim of documenting DGS and its range of variation.

5.2   Preparing Examples

After selecting a particular example, it is prepared for the inclusion in the dictionary entry. Preparation includes selecting the segment to be shown, adjusting the translation and adding a written context, if necessary.

5.2.1 Selecting the sequence to be shown

When deciding on which part of the utterance is to be shown it is important to consider the structures of DGS. Prosodic units should be completely included, hence utterance boundaries have to be considered. In example (3), theoretically sign 9 does not contribute to the meaning, and might have been cut off as it is a gesture mostly used to indicate that an utterance is finished. However, the prosodic phrase is not finished without sign 9, and the mouthing 'kann' stretches over signs 8 and 9. As such, leaving it out is not appropriate.

As for the content, a set of rules were established. The sequence needs to be long enough to properly illustrate the sign sense, but still short enough to be easily understood and perceived. When the shown segment is too long the user might lose focus, not recognize the sign and be distracted by other information that is not central in the signing.

5.2.2 Context

It is well known that genuine discourse examples taken out of context and shown in isolation are often hard to understand and not ideal in form and content, especially if they are taken from a spoken – or signed – conversation. Such authentic examples are difficult to comprehend in isolation, and need additional context information.

(4) ▣ Dort waren nur wenige **Besucher**.

Translation: There were only few **guests**.

In example (4) is not clear what 'dort' (pointing sign) refers to. Here, context information is needed to understand the reference of the pointing sign and to comprehend the sense-specific semantic conditions of the sign's use. Thus the example would occur with a context (within square brackets) in the dictionary, as follows:

(5) ▣ [Thema: Festveranstaltung eines Basketballvereins] Dort waren nur wenige **Besucher**.

    [Subject: festivity of a basketball club] There were only few **guests**.

Also without a larger context some examples would create a false impression of the informant (see example (6), being released from prison vs. being released from political imprisonment):

(6) ▣ [Der Erzähler wurde nach einem missglückten Fluchtversuch aus der DDR inhaftiert.] Als ich **aus** dem Gefängnis **kam**, war es mir wichtig, meine Ausbildung zum Tischler weiterzumachen. Die Prüfung habe ich dann auch bestanden.

    Translation: [The signer was detained after an attempted escape from the GDR.] When I **got out** of prison, I concentrated on my training to become a carpenter. I passed the final exam.

For most examples we provide written information to promote the understanding of the selected examples. Necessary information is given, but should be as short and precise as possible. Whether a context is necessary or not may also depend on the user group. Native signers usually need less additional information than learners.

5.2.3 Translation

Each authentic example is presented in the entry along with a translation. For large parts of the corpus, translations are already available and time-aligned to the videos. These translations aid annotations and are also shown in the Public DGS Corpus. For the examples these translations can be used as the starting point for the translation shown with the example in the entry. However, as we mostly focus on a small segment it is necessary to check whether the translation of that part corresponds to the signing in the chosen segment. Additionally we adjust translations to be understandable in isolation. The translational equivalent of the sign is highlighted in bold face in the translation (cf. Figure 1).

6   Conclusion

In the special circumstances SL lexicography has to deal with, we found from our practical work that using authentic examples taken from original corpus recordings can be done and is a choice worthy of further consideration in corpus-based SL lexicography. This option was made possible by a large lemmatized and annotated corpus of SL with good technical quality of the original recordings and the informed consent of the informants for public use of their data. We consider the authentic examples to be valuable information to illustrate the usage of SL in a very typical, natural and lively – that is authentic – way. We are aware of the drawbacks of authentic examples, but also found that these can largely be compensated for by careful selection, preparation and additional information given in the related entries. As corpus-based SL lexicography is still in its first stages of development, and thus it remains to be seen whether other SL dictionary projects follow along this line, and how the best practices and standards will be developed in the future.

References

Dictionaries

(D-ASL) Stokoe, W.C., Casterline, D.C. & Croneberg, C.G. (1965). *A dictionary of American Sign Language on linguistic principles*. Washington, DC: Gallaudet College Press.

(D-LSFB) *Dictionnaire de LSFB en ligne et la grammaire de LSFB*. Accessed at: http://dicto.lsfb.be [29/03/2018].

(D-NZSL) Kennedy, G., Arnold, R., Dugdale, P. & Moskovitz, D. (1998). *A dictionary of New Zealand Sign Language*. Auckland: Auckland University Press.

(D-SAS) Penn, C. (1992-1994). *Dictionary of Southern African signs for communicating with the deaf*. Pretoria: Joint Project of the Human Sciences Research Council and the South African National Council for the Deaf.

(KS-PJM) Łacheta, J., Czajkowska-Kisil, M., Linde-Usiekniewicz, J. & Rutkowski, P. (2016). *Korpusowy słownik polskiego języka migowego/Corpus-based Dictionary of Polish Sign Language*. Warsaw: Faculty of Polish Studies, University of Warsaw. Accessed at: http://www.slownikpjm.uw.edu.pl/en [29/03/2018].

(ODT-SL) *Ordbog over Dansk Tegnsprog*. Center for Tegnsprog (2008-2016). Accessed at: http://www.tegnsprok.dk [29/03/2018].

(ODT-VL) *Ordbog over Dansk Talesprog*. Københavns Universitet. Accessed at: http://odt.hum.ku.dk [14/05/2018].

(ONZSL) McKee, D., McKee, R., Pivac Alexander, S., Pivac, L. & Vale, M. (2011). *Online dictionary of New Zealand Sign Language*. Wellington: Deaf Studies Research Unit, Victoria University of Wellington. Accessed at: https://nzsl.vuw.ac.nz [29/03/2018].

Literature

Atkins, S.B.T., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Bank, R., Crasborn, O. & van Hout, R. (2016) The prominence of spoken language elements in a sign language. In *Linguistics*, 54 (6), pp. 1281-1305.

De Vos, C.,Van der Kooij, E. & Crasborn, O. (2009). Mixed Signals: Combining Linguistic and Affective Functions of Eyebrows in Questions in Sign Language of the Netherlands. In *Language and Speech*, 52 (2-3), pp. 315-339.

Ebbinghaus, H. (2001). Wird sich die Sprache der Gehörlosen 'vertikalisieren'? Möglichkeiten und Grenzen der Registerdifferenzierung in der Deutschen Gebärdensprache. In G. List, G. List (eds.) *Quersprachigkeit. Zum transkulturellen Registergebrauch in Laut- und Gebärdensprachen*. Tübingen: Stauffenburg Verlag, pp. 187-199.

Ebbinghaus, H., Hessmann, J. (2001). Sign language as multidimensional communication: Why manual signs, mouthings, and mouth gestures are three different things. In P. Boyes Braem, R. Sutton-Spence (eds.) *The hands are the head of the mouth. The mouth as articulator in sign languages*. Hamburg: Signum-Verlag, pp. 133-151.

Ehlich, K. (1983). Text und sprachliches Handeln. Die Entstehung von Texten aus dem Bedürfnis nach Überlieferung. In A. Assmann, J. Assmann & C. Hardmeier (eds.) *Schrift und Gedächtnis. Beiträge zur Archäologie der literarischen Kommunikation*. München: Wilhelm Fink Verlag, pp. 24-43.

Erlenkamp, S. (1998). Lexikographie als Teil einer Minderheitenpolitik – Minderheitenpolitik als Teil einer Lexikographie. In *Das Zeichen*, 12 (43), pp. 98-105.

Fehrmann, G., Linz, E. (2009). Eine Medientheorie ohne Medien? Zur Unterscheidung von konzeptioneller und medialer Mündlichkeit und Schriftlichkeit. In E. Birk, J.G. Schneider (eds.) *Philosophie der Schrift*. Tübingen: Max Niemeyer, pp. 123-143.

Hansen, C., Hansen, M.H. (2012). A Dictionary of spoken Danish. In R. Vatvedt Fjeld, J.M. Torjusen (eds.) *Proceedings of the 15th EURALEX International Congress, 7.-11. August 2012*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, pp. 929-935.

Jahn, E., Konrad, R., Langer, G., Wagner, S. & Hanke, T. (2018). DGS-Korpus Project: Different Formats for Different Needs. In M. Bono, E. Efthimiou, S.E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, J. Mesch & Y. Osugi (eds.) *Involving the Language Community. Proceedings of the 8th Workshop on the Representation and Processing of Sign Language. 11th International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan.* Paris: Elra, pp. 83-90.

Keating, E., Edwards, T. & Mirus, G. (2008). Cybersign: Impacts of New Communication Technologies on Space and Language. In *Journal of Pragmatics*, 40 (6), pp. 1067-1081.

Kennedy, G. (1996). Designing a dictionary of New Zealand Sign Language. In G. Kennedy (ed.) *Topics in New Zealand Sign Language Studies*. Wellington: Victoria University of Wellington, pp. 27-48.

Kristoffersen, J.H. (2010). *From utterance to example sentence – excerpting and adapting corpus sentences for use in the Danish Sign Language Dictionary*. [Poster from the Sign Linguistics Corpora Network. 4. Workshop: Exploitation, Berlin, 3.-4. December 2010.] Accessed at: http://www.tegnsprog.dk/hovedside/litteratur/SLCN_Berlin_2010_poster.pdf [29/03/2018].

Kristoffersen, J.H., Troelsgård, T. (2010). The Danish Sign Language Dictionary. In A. Dykstra, T. Schoonheim, (eds.) *Proceedings of the 14th EURALEX International Congress, 6.-10. July 2010.* Leeuwarden/Ljouwert: Fryske Akademy, pp. 1549-1554.

Krentz, C.B. (2007). The Camera as Printing Press: How Film has Influenced ASL Literature. In H-D.L. Bauman, J.L. Nelson & H.M. Rose (eds.) *Signing the Body Poetic.* Berkeley: University of California Press, pp. 51-70.

Langer, G., Müller, A., Wähl, S. (2018). Queries and Views in iLex to Support Corpus-based Lexicographic Work on German Sign Language (DGS). In M. Bono, E. Efthimiou, S.E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, J. Mesch & Y. Osugi (eds.) *Involving the Language Community. Proceedings of the 8th Workshop on the Representation and Processing of Sign Language. 11th International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan.* Paris: Elra, pp. 107-114.

Linde-Usiekniewicz, J., Rutkowski, P. (2016). The division into parts of speech in the Corpus-based Dictionary of Polish Sign Language. In T. Margalitadze, G. Meladze (eds.) *Proceedings of the 17th EURALEX International Congress, 6-10 September 2016.* Tbilisi: Ivane Javakhishvili Tbilisi University Press, pp. 375-388.

McKee, R., McKee, D. (2013). Making an Online Dictionary of New Zealand Sign Language. In *Lexikos*, 23, pp. 500-531. Accessed at: http://dx.doi.org/10.5788/23-1-1227 [16/05/2018].

Möhrs, C., Meliss, M. & Batinić, D. (2017). LeGeDe - towards a corpus-based lexical resource of spoken German. In I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek & V. Baisa (eds.) *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference. Leiden, the Netherlands, 19.-21. September 2017.* Brno: Lexical Computing CZ s.r.o., pp. 281-298.

Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke, T. & Rathmann, C. (2010). Elicitation methods in the DGS (German Sign Language) Corpus Project. In P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martínez Ruiz & A. Schembri (eds.) *Corpora and Sign Language Technologies. Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages. 7th International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta.* Paris: ELRA, pp. 178-185.

Prinsloo, D.J., Gouws, R.H. (2000). The Use of Examples in Polyfunctional Dictionaries. In *Lexikos*, 10, pp. 138–156. Accessed at: http://dx.doi.org/10.5788/10-0-891 [29/03/2018].

Schwitalla, J. (1997). *Gesprochenes Deutsch. Eine Einführung.* Berlin: Schmidt.

Stokoe, W.C. (1993). Dictionary making, then and now. In *Sign Language Studies*, 22 (79), pp. 126-146.

Sutton's SignWriting Site. Accessed at: http://www.signwriting.org [29/03/2018].

Trap-Jensen, L. (2004). Spoken language in dictionaries: Does it really matter? In G. Williams, S. Vessiers (eds.) *Proceedings of the Eleventh EURALEX International Congress 2004.* Lorient: Faculté des lettres et des sciences humaines, Université de Bretagne-Sud, pp. 311-318.

Verdonik, D., Sepesy Maučec, M. (2017). A Speech Corpus as a Source of Lexical Information. In *International Journal of Lexicography*, 30 (2), pp. 143–166.

Wöhrmann, S. (2003). GebärdenSchrift lesen lernen. In *Das Zeichen*, 17 (65), pp. 364–374.

Acknowledgements