Compiling a Basic Vocabulary for German Sign Language (DGS) – lexicographic issues with a focus on word senses

Gabriele Langer, Susanne König, Silke Matthes
Institute of German Sign Language and Communications of the Deaf, University of Hamburg
Gabriele.Langer@sign-lang.uni-hamburg.de,
Susanne.Koenig@sign-lang.uni-hamburg.de,
Silke.Matthes@sign-lang.uni-hamburg.de

Abstract

Nowadays lexicographic work such as lemma selection, identification of word senses and usage information are usually based on large corpora. Sign language lexicographers face the same difficulties as their colleagues of other less studied and previously not written minority languages. They cannot rely on written texts and large corpora. In addition sign language linguists have to cope with sign-language specific issues due to the visual-gestural modality, namely the iconicity of signs, a broad utilization of simultaneity of linguistic signals and the integration of lexical material of a spoken language by the way of mouthings.

Sign language specific lexicographic issues are discussed and exemplified with regard to the Basic Vocabulary of German Sign Language (DGS) that is compiled within the larger context of the DGS Corpus Project. The Basic Vocabulary is not corpus-based but based on previously published sign collections that are used as a starting point. The data is reviewed, sign senses are finer split and disambiguated and the data then undergo a validation process by the sign language community. This validation process is conducted through a feedback system especially designed for surveys involving sign language.

Keywords: lexicography of sign language; German Sign Language (DGS); word senses; feedback system; crowd sourcing; basic vocabulary; iconicity; mouthing

1   Background

1.1   Sign Language Dictionaries

Sign language lexicography is still in the process of striving for the best lexicographic methods to analyse signs, as well as solutions to document and present them in an adequate and user-friendly way. For meaning analyses and descriptions there are two aspects especially relevant in a dictionary pro-

ject: First, what kind of data is the given information based upon, and, second, how can the identified meanings of signs (senses) be made clear and best presented to the dictionary user.

In the past – due to a lack of an everyday writing system and written sign texts – sign dictionaries have hardly been in the position to base their sign selection and information on sign use on corpus analyses. Up to the present, sign dictionaries and sign collections have mainly been based on information drawn from earlier sign collections, word-to-sign elicitations, the introspection of single signers or groups of signers (and their metalinguistic discussions on signs' meanings and usage), on participant observation, or on a mixture of these methods (compare for example Johnston 1989: 8, Brien 1992: x-xi, Kristoffersen & Troelsgård 2010: 5). To our knowledge the first sign language dictionary largely and systematically based on a collection of filmed and transcribed sign language data was the ASL Dictionary published in 1965 (Stokoe, Casterline & Croneberg 1965). Considering the limited technical possibilities at the time (cf. Stokoe 1993) this work was outstanding in many regards and way ahead of its time.

Technical progress in recording equipment and transcription software now makes the collection of relatively large sign language corpora feasible. Sign corpora exist or are being collected in several countries such as Australia, Germany, Great Britain, Italy, the Netherlands, Poland, France and others. Most of these corpora are still in the stage of data collection or transcription and annotation. With the availability of larger annotated sign corpora the information in sign dictionaries – especially on the meanings and usage of signs – can reach new levels of quality.

Without a fully functioning writing system for signs (see below) some effort has to be put into the presentation of sign forms and how to provide a search by sign form in order to ensure bi-directional accessibility in sign dictionaries. However, how to best identify meaning(s) of signs and to make them accessible has not been discussed in great depth so far in sign language lexicography. Common to all general sign dictionaries that we are aware of is the bilingual solution of expressing the meanings of signs by matching translation equivalents (keywords) of the surrounding spoken language. In addition to that some dictionaries also provide meaning explanations (e.g. Johnston 1989), usage notes (e.g. Stokoe, Casterline & Croneberg 1965) or example-like contexts (e.g. Kennedy 2008, cf. esp. xvii) – all using the surrounding spoken language as meta-language – or signed examples (e.g. *Ordbog over Dansk Tegnsprog*).

In this paper we differentiate between sign dictionaries and more simple sign collections. A sign dictionary regards the signs as the linguistic units to be described with information about their form, grammar, meaning and usage. Sign language dictionaries of this kind have sign entries and – if necessary – include several senses of a given sign. Sign collections, on the other hand, are predominantly mono-directional word lists with the direction from spoken language words to signs. Usually, they contain no real sign entries and do not provide much information about the sign other than its form. For German Sign Language (DGS) there exist quite a number of different sign collections but not yet a corpus-based general dictionary.

1.2   The DGS Corpus Project

The DGS Corpus Project[1] is a long-term project of the Academy of Sciences and Humanities in Hamburg. The research is carried out at the Institute for German Sign Language and Communication of the Deaf (University of Hamburg). The project has started in 2009 and will continue until 2023. The project has two major goals: building a large general reference corpus of DGS and producing the first corpus-based general dictionary of DGS – German.

The DGS corpus is intended to be a linguistic and cultural resource for further research as well as the language community and other interested persons. It will also provide data for the first corpus-based dictionary DGS – German. For the corpus 330 signing informants from all over Germany were filmed in pairs at 12 different locations. The filmed material includes narrations of personal experiences, discussions, re-narrations, and other kinds of language use elicited in staged communicative events.[2] The raw data are comprised of 900 hours recording time with estimated 540 hours of signed activity containing estimated 3.5 million tokens of signs. A large portion of this data is being made accessible through segmentation, lemmatization, transcription, annotation and translation to become a searchable corpus. Corpus annotation work is carried out using iLex, an annotation environment that is linked to a lexical database (Hanke & Storz 2008). For lemmatization tokens are matched to types that are represented by unique glosses. The citation form of each sign type is described using the notation system HamNoSys (Hanke 2004). This is very time-consuming and cumbersome work, as none of these steps can be carried out in an automated way yet. Hence, the corpus will not have reached a relevant size for a number of years to go. This leads to the situation that at the present time DGS lexicographers have to resort to other sources and methods for dictionary compilation.

Within the DGS Corpus Project two electronic dictionaries of DGS will be produced in consecutive phases. First – while the corpus is still being built and the elicited data not yet fully accessible – a preliminary Basic Vocabulary with about 1500 entries is being produced. It is based on previously published sign collections and validated through feedback from the signing community. At a later stage the corpus data will be used as basis for the production of a general Dictionary of DGS – German with approximately 6000 sign entries to be published in 2023.

1.3   Sign Language as a Multi-channelled and Non-written Language

One challenge each sign dictionary project has to face is, that sign language is a visual-gestural language without written form. A sign stream involves signals produced by different parts of the upper body – two hands, mouth movements, facial expressions, head movements, posture, body shifts and eye-gaze. In the visual-gestural modality these articulatory features are easily produced and perceived simultaneously. Since the sign stream consists of several simultaneous but different signals it is diffi-

---

1   For more information on the project see: www.dgs-korpus.de.
2   For more detail see Nishio et al. (2010).

cult to devise a writing system that on the one hand can capture and represent the form and given complexity and on the other hand is simple enough to be written and read easily. In the research of spoken languages it is possible to adapt an existing writing system to a previously unwritten language – at least for a transition period – because the problem of how to write spoken languages has been solved. In sign language, however, there is no functioning writing system for any language that could simply be adapted to DGS for the purpose of dictionary making.[3] For sign language lexicography this absence of an operating writing system poses several challenges:

First, researchers cannot rely on already existing written texts to build a corpus of or to base their analyses on. Corpora have to work with data consisting of filmed signing. For corpus building such films are often recorded specifically for this purpose. Sign films recorded for other purposes are still limited to specific contexts (types of language use, genres). For most signers they are not part of everyday communication and can therefore cover only a small segment of sign language use. Second, for corpus-building and transcription purposes there is no fast and easy way of writing down signed text in a way that represents the form of the language units. More demanding and time consuming notation and transcription methods or indirect forms of annotations have to be used. This also makes it more difficult to analyse and automatically process corpus data. For instance searches on word forms have to be provided for by other means if there is no written representation of the forms available. Third, there is no direct and stable representation of form that enables the analyser to scan and review the sign and its context on a glance directly through their linguistic forms in the same way as a written text would do for example in a concordance line. Fourth, for dictionary production there is no practical way to use sign language as meta-language extensively. Film clips with signed information do not provide the same usability for scanning, browsing, searching and punctual access as does written information, since the content of a film is not permanently visible or machine-readable. Furthermore, films are very time-consuming in production and editing and therefore also for practical and financial reasons no real alternative to the use of written information.

Fifth, since there is no written form of the lemma sign, the question arises of what element to use as lemma. The lemma usually represents the form of the word, serves as an element for quick identification, sorting, searching and ordering and as the address for cross-references. All elements that could possibly be chosen (film, drawing, notation, gloss, number) do not fulfil all of these functions and have their drawbacks.[4] The best solution so far is to use a combination of one element to represent

---

3   There are some notation systems developed for academic purposes (e.g. HamNoSys, Stokoe-Notation). They are suited for detailed descriptions of sign forms. However, these systems do not handle facial expressions, body posture, eye gaze, mouthings or iconic locations well enough to completely rely on them in research. Also stretches of text are not easily written and read. Other systems, intended for general use such as SignWriting (Sutton et al. 2009), are not widely established and also have other drawbacks like not being machine-readable. For a further discussion of merits and shortcomings of the most common writing systems see König & Schmaling (2012).

4   For a more in-depth discussion of the advantages and disadvantages of these elements representing signs in dictionaries see Zwitserlood 2010: especially pp. 445-454 and Kristoffersen & Troelsgård, 2012a: 295-298.

the form of the lemma sign (e.g. drawing or film) and a second element for unique identification, sorting and address for cross-referencing (e.g. unique gloss or number).

2   The Basic Vocabulary of DGS

2.1   The Basic Vocabulary within the Context of the DGS Corpus Project

The Basic Vocabulary is intended to be a basic learners' vocabulary and to cover the most central signs of DGS and their core meanings. It is based on information about DGS signs contained in different sign collections and learning materials. The information is lemmatized, reviewed, edited and prepared for the signing community's feedback by which it will be validated and supplemented. The Basic Vocabulary is only a first and preliminary step in the lexicographic work within the project. It serves a number of purposes within the overall project design:

It is a means to round up and review information already published on DGS signs and put some of it to the test. This allows for a critical review and evaluation of previous lexicographic work with the chance of identifying and eliminating artefacts and to move towards a more complete and adequate description of signs and their meanings. With the Basic Vocabulary we can also develop and test lexicographic solutions to sign language specific problems that result from the visual modality as well as from the lack of a written form.

In addition, the publication of the Basic Vocabulary is a way to give something back to the signing community for their previous participation in the project, i.e. as informants in the corpus data collection, much earlier than 2023. The feedback process is also a way to involve the language community and let them further participate in this project on their language. Through the feedback they gain some influence and control over how their language is represented in the resulting dictionaries.

For the compilation of the Basic Vocabulary a set of question types to validate lexical items are developed, tested and used to collect data on signers' intuitions about signs and their meanings. Also a feedback system is specifically designed to conduct this survey via sign language. In the context of the Basic Vocabulary we can develop, test and improve both feedback tool and procedures. The gained experience and know-how can at a later stage be utilized for supplementing corpus data in the production of the Dictionary DGS – German.[5]

---

5   With estimated 2.5 million tokens the DGS corpus will still be relatively small in comparison to corpora of well-studied written languages. Complementary methods to gain additional insights on signs and their meanings and uses will be useful.

2.2   Properties of the Basic Vocabulary

The Basic Vocabulary is developed with a primarily monolingual oriented perspective: It focuses on the signs and aims at an adequate description of their variants and meanings. The meta-language of the dictionary is written German. Following this approach the dictionary will contain only one type of entries, that is, sign entries. Nearly all information included in the Basic Vocabulary will concern DGS signs while at the same time there is neither intention nor need to reproduce information on German words that can be obtained easily from existing dictionaries of German. Nevertheless, the actual product will exhibit some features of a bilingual dictionary: The signs' senses will be identified and expressed by German translation equivalents and disambiguating contexts. Furthermore there will be an alphabetical index of the German equivalents, making the sign entries accessible also via German (as source language) and thus providing also for a bilingual access.

The sign entries will be ordered by formational parameters of the signs.[6] However, since the Basic Vocabulary will be produced in an electronic form, the primary ordering of signs (macrostructure) is not crucial. The product will include a search function accessing the entries via sign form. For a search the user can select different combinations of manifestations of the sign form parameters as presented in a series of menus.[7]

2.3   Sign Collections: Sources for Lemma Selection and Spoken Language Equivalents

For the compilation of the Basic Vocabulary nine previously produced sign collections of DGS are taken as a basis for the selection and analysis of basic DGS signs.[8] The sign collections are used in two ways: for lemma selection and as a starting point for word sense disambiguation.[9]

---

6   Formational parameters are e.g. handshape, location, number of hands, movement types. There is no agreed upon order available and existing sign dictionaries have arrived at different solutions. For the current project the final order has not been decided yet.

7   This search function will be similar to the one implemented for the *Fachgebärdenlexikon Gesundheit und Pflege* ([5e]). To go to the sign form search function click on "Suche über Gebärdenform" at http://www.sign-lang.uni-hamburg.de/glex/.

8   Some of these collections consist of several publications that have been produced by the same team or institution and are follow ups of each other or a series of complementary works. For an overview see separate list in the reference section. The list of sign collections also include some teaching materials with vocabulary lists.

9   *Word sense* here in the case of sign language means "sign sense". In linguistics many technical terms exhibit a spoken language bias. Examples are: speaker, word, phoneme, phonology and oral. However, in a more general sense they can be used to describe the same abstract or corresponding phenomenon, category or role in both language modalities. Other technical terms with a spoken language bias are applied to sign language accordingly in this paper.

2.3.1 Lemma Selection

The Basic Vocabulary aims at including signs that are most basic with regard to everyday communication needs and that are frequently used and widely known. With frequency lists still unavailable for DGS the second best way to approximate this aim was to look for signs that had been listed in several previously published sign collections. Following this, we based the lemma selection process primarily on a comparison of these signs.

Without a written representation of the sign forms in the products the only way to identify and compare the signs was to look at each visual representation (photos with arrows or films) of the given sign, sign combination or sign sentence, break sign combinations and sentences down into single signs (tokenisation) and lemmatize them consistently. This was done in our working environment iLex following the annotation rules of the DGS Corpus Project. The nine sign collections contained about 40.800 tokens (single signs) that have been processed. Not counting uses of the manual alphabet, pointing signs, number signs and productive signs about 9350 different sign forms of 5440 basic conventional signs haven been documented. All basic signs – about 1050 – that were found in at least five of the nine sign collections have been selected for further analysis and inclusion into the Basic Vocabulary.[10]

2.3.2 Spoken Language Equivalents

All of the nine sign collections contain to some extent bilingual information matching signs and words as equivalents. They either have a primary make-up using German as source language showing signed equivalents (e.g. [6]) or they have sign entries and use German equivalents and contexts to list the signs' meaning(s) (e.g. [5c-5f]). The vocabulary lists of learning materials (e.g. [9a, 9b]) and the collection of phrases ([2c]) use glosses as written labels for signs where the gloss word is to be taken as an indication of the meaning of the sign.[11] In the children's visual dictionary ([8a, 8b, 8c]) pictures of objects and sign films are presented without any written words. However, the sign film contains an audio track where the corresponding German word is spoken aloud.[12] All available equivalents and meaning indications assigned to one sign form in different sign collections are taken as a starting point for identifying word senses for signs that are to be included in the Basic Dictionary.

In the Basic Vocabulary we aim at the most central meanings of selected signs. Not all of the equivalents listed in the sign collections are equally basic or central, and it is sensible and necessary to make a selection. This selection is operationalized in the following way: German equivalents are compared

---

10   Additional signs will be included in order to make sure all concepts of a specific learner's basic word list of German (Glaboniat et al. 2005) and some Deaf-specific concepts are covered.

11   In corpus annotation and lexicography glosses are used as a unique identifiers for signs: one gloss always stands for the same form and different forms receive different glosses. In learning materials glosses are often used as a means to represent the sign order of example sentences and at the same time to indicate the signs' meanings in this particular context. Depending on the context a sign might therefore be represented by different glosses and the same gloss might be used for different signs.

12   In the print version of the product German equivalents of signs are printed below the signs' drawings (comp. Kestner 2002).

to a combined list of German words that are considered relevant for a basic (German) vocabulary. This combined list is basically a merge of a foreign language learner's basic vocabulary list of 1071 words (adapted from Glaboniat et al. 2005, levels A1 and A2 (active vocabulary)) and a frequency list of the about 4000 most frequent German words (Jones & Tschirner 2006, 2011). Words that are both listed as equivalents in the sign collections and found in the combined German word list are selected for further processing.

2.3.3 Reversal of bilingual information

After the basic German equivalents of a sign have been determined, the next step is to identify individual senses that the words and the sign potentially share. This step can be understood in terms of the reversal of a mono-directional bilingual dictionary. It goes beyond the reversal on the form level and is aimed at the level of meaning, i.e. the level of lexical units (as outlined by Martin 2013: 1447-1448). However, in our case the equivalents do not stem from one uniform source but from different sources and are of different status, quality and granularity as far as meaning specification is concerned.

It is reasonable to assume that any German word listed as translational equivalent must share at least one sense with the corresponding sign. If the German word is polysemous each of its senses has to be considered separately, whether it could also be a potential sense of the sign. At this stage it is inevitable to do a considerable amount of sense splitting. For this we use German dictionaries. The approach we take is to identify relevant senses of the German word according to two criteria: First, the senses chosen are to be the most basic and central ones with regard to everyday communication. More marginal senses are omitted from further processing. Second, the senses selected for the feedback have to be certain or likely candidates for being senses of the sign as well. Whenever the context information provided in a sign collection clearly identifies a particular sense this information is taken into account and the identified sense goes into the feedback process for verification. The same is true for senses that are attested clearly in the data from previous projects or already transcribed corpus data in our database. In some cases it is very straightforward that a particular word sense is clearly not a sign sense and can be omitted, e.g. because it contradicts with the iconic properties of the sign. In many other cases a decision cannot be reached without further information of the sign's use in context. Uncertain basic senses are prepared for the feedback process to be verified or rejected by the language community (see 4.).

3 Sign Language Specific Issues concerning Word Sense Disambiguation

In the process of identifying and diambiguating word senses of signs two sign language specific properties have to be taken into consideration: iconicity and mouthings.

3.1 Iconicity

Many signs are iconic. The form of an iconic sign resembles something – the underlying image – which is directly or indirectly associated with one or several of the sign's meanings. The underlying image can influence the ways a sign's meaning may be restricted or open for extension. Therefore, when checking possible senses of a sign special attention has to be given to its underlying image.

3.1.1 Underlying Image and Meaning

The underlying image of a sign may restrict its use to certain specific senses and prohibit its use for related senses. Due to iconic aspects the granularity of meaning differentiation of a sign may be much finer than that of the corresponding word of the surrounding spoken language. When the underlying image of an iconic sign is very specific and the sign's meaning is restricted accordingly, related senses that are not compatible with the iconic properties of the sign in question are often covered by other signs.

Consider for example different DGS signs for *Pfeife*. The German word *Pfeife* can mean either "whistle (object)" or "smoking pipe (object)" while DGS has different signs for these concepts PIPE[13], WHISTLE1 and WHISTLE2 (see figure 1). The sign form of PIPE resembles a person holding a pipe at its bowl some small distance away from the mouth so that the end of its imagined shank touches the lips for smoking. In WHISTLE1 the fingers indicate a whistle (as used by referees) located with its mouthpiece at the mouth. WHISTLE2 shows how a person is holding a whistle to the mouth. The signs' meanings are restricted to the different objects in accordance with their underlying image.

---

13 Since there is no conventional writing system for sign languages it is a convention in sign language linguistics to use capital letter glosses with numberings to identify and represent different signs. The spoken language word used as gloss simply is an easily written and readable label or name for the a sign, but neither is the gloss' part of speech nor are its meanings to be interpreted as indicating the sign's part of speech or meanings.

Figure 1: DGS signs PIPE, WHISTLE1 (Source: [5e]), WHISTLE2 (Source: [6]), AND WHISTLE3 (Source: [5e]).

WHISTLE1 and WHISTLE2 can also mean "producing a whistling sound, whistling". Another DGS sign listed for this meaning is WHISTLE3 (see figure1). On the iconic level the three signs for *whistling* are specific as to how the whistling sound is produced: WHISTLE3 shows how thumb and index finger are rounded to an O and put to the mouth producing the sound without the help of a whistle, while for WHISTLE1 and WHISTLE2 a whistle is involved. The question arises whether these iconic differences are reflected in sign usage or not. In a specific context the sign used will usually be the one that iconically best matches the named or described incident.[14] Thus the "whistling" sense of the signs could be further specified as "producing a whistling sound by blowing into a whistle" for WHISTLE1 and WHISTLE2 and "producing a whistling sound by using your fingers in the indicated way" for WHISTLE3. Sign language users may find this trivial or logical or even unnecessary information since the form of the sign shows these aspects of meaning, but this only proves the point that iconicity plays a role in sign language use. Whenever the selection and use of a particular iconic sign takes its underlying image into account this kind of information should be made explicit in a dictionary and thus be accessible to the language learner. This can be done by listing the specific meaning as determined by the iconic value of the sign as one of the sign's senses.

However, iconic signs do not always exploit their more specific underlying images for subtle meaning differences. Often sign forms are conventionalized for a general meaning without paying attention to the more specific images their forms are derived from. One example is the sign TELEVISION (see figure 2). Its underlying image shows a person turning two knobs (the kind old television sets had for adjusting the frequency). This sign is used for all kind of television sets regardless of their technical type or outward appearance.

---

14   This is a hypothesis based on informal observations and signers' intuition. To what extend this can be confirmed by corpus data remains an open question.

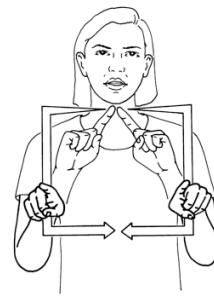**Figure 2: TELEVISION and old television set with turning knobs.**

Figure 3: RECTANGLE.

Signs – just like words of spoken languages – can extend their meaning from the specific to the more general. With this we come back to the whistling examples: When the particular circumstances of sound production are not known to the signer or are not relevant or if one talks about whistling in general, is it then possible that any of the three signs may be used for the more general concept of "whistling" independent of how the sound is produced? And what about the whistling of a boiling water kettle or of an old steam engine train? Can the three signs for "whistling" also be used in a transferred sense for these kinds of whistling or are there other signs to be used instead? Do WHISTLE1, WHISTLE2 and WHISTLE3 additionally to their very specific sense matching their underlying image have a more general meaning "producing a whistling sound" (no matter how it is produced)? These are questions to be asked and hopefully answered in the feedback and – at a later stage of the project – by analysis of corpus data.

Whenever a sign's form resembles one particular example of a more general category its meaning can either be restricted to that specific kind resembled, or the sign can be used for the more general concept. There is no general rule as to what applies to a particular sign. For each sign this has to be determined separately. Thus, for iconic signs special attention has to be given underlying image when analysing their meanings.

3.1.2 Underlying Image as a Source for Polysemy

The iconicity of signs is an important source for polysemy and meaning extensions in DGS. This is especially true when the sign's form represents a very unspecified image that may stand for a number of objects, situations or actions. The underlying image can serve as a common core for rather diverse meanings which either share a common visual trait in their real-world manifestations or which utilize the visual metaphor provided. For example, among the translation equivalents listed in various sign collections as meanings for the sign RECTANGLE (see figure 3) quite a number refer to objects or convey meanings that either consist of or are associated with a piece of paper, such as: "paper", "piece of paper", "slip (paper)", "page", "form (to be filled in)", "map", "recipe (cooking)", "prescription", "sick note", "official notification", "certificate", "report card". Other meanings refer to rectangular and predominantly two-dimensional objects that are solid such as "window", "signboard", "mirror", "screen (computer, TV)", or that are soft such as "pillow", "towel", and "(cleaning) rag". On the semantic level a

"pillow" does not have much in common with a "computer screen" neither does an "official notification" have anything to with a "cleaning rag" – the characteristic all these share is their rectangular shape or their association with something of a rectangular shape. In cases like this the iconic potential depicted in the sign's form serves as the basis for meaning extension of the sign.[15]

This kind of extensive polysemy in DGS is facilitated by an interaction of iconicity and mouthings. The mouthing helps to contextualize and disambiguate the polysemous sign (see below 3.2.).[16]

In DGS there are many ways of how the meaning of a sign can be extended. Often the meaning is extended along the lines of regular polysemy. Bentele, Konrad & Langer (2000: 621) list some conventional and productive uses of the sign STIRRING that exemplify some of these patterns (see Table 1). The form of STIRRING (see figure 4) depicts someone holding a tool, such as a cooking spoon, and stirring something with it.

It is very common in DGS that in addition to a sign's core meaning (an action represented by the sign form as in STIRRING) it is also used in many other related senses – for example – for a profession typically associated with the activity, for a place or institution where the activity takes place, and for an object produced or manipulated by the activity.
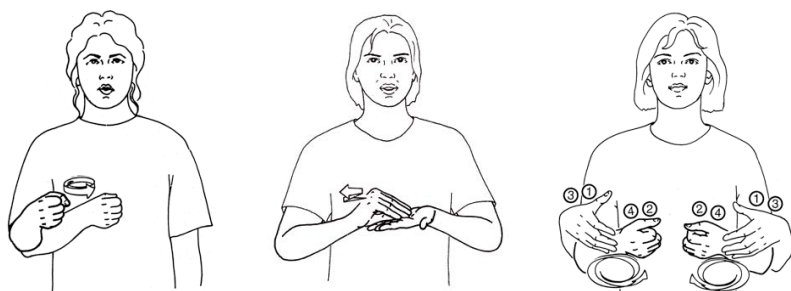

Figure 4: STIRRING, WRITING, and KNEADING.   **Figure 5: BEAK.**

---

15   The meanings listed here are conventional uses of the sign RECTANGLE as they appear in sign collections. The list here is not complete. These conventional meanings are strongly and regularly associated with this sign. However, in addition the sign may also spontaneously be used to denote other rectangular two-dimensional objects in a specific context, often in combination with a corresponding mouthing. This is a common strategy to be used, especially when there is no other conventional sign or when the signer doesn't know such a sign. Such more spontaneous uses are backed by the sign's iconic potential and will be easily understood in context. In contrast to the more *conventional uses* these have been called *productive uses* of a sign (cf. König, Konrad & Langer 2008).

16   Specific mouth gestures (mouth movements not derived from spoken language words) and facial expressions can also contribute to the disambiguation of meanings. These are not discussed here.

Other examples of this type are WRITING ("writing", "author, writer", "school" and "literature"), and KNEADING ("kneading", "baker", "bakery", "dough"). Our impression is that in addition to well-known patterns of regular polysemy found in spoken languages (e.g. Atkins & Rundell 2008: 140) some DGS patterns may be specific to a visual language.[17]

German equivalents	English equivalents	meaning extension pattern
rühren, umrühren	to stir	core meaning: action directly represented by sign's form
kochen	cook (prepare food)	more comprehensive activity typically involving the represented action
Küche	kitchen	place or area where activity usually takes place
Koch	cook, chef	person who usually or professionally carries out the activity
(Koch-)Löffel	(cooking) spoon	tool used in the represented action (productive use)
Eintopf	stew	object or substance manipulated by action (productive use)
Hauswirtschaft	home economics	superordinate domain the activity belongs to (productive use)

**Table 1: Some meaning extensions for STIRRING
(adapted from Bentele, Konrad & Langer 2000: 621).**

The conclusion that we draw from these observations is that the underlying image and image producing technique of an iconic sign can be and should be considered for decisions on lemmatisation and in the analysis of word senses. For us a common underlying image connecting diverse meanings is a valid reason to treat these meanings as different senses of a polysemous sign in one entry rather than treating them as homonyms in different entries (cf. also König, Konrad & Langer 2008).

3.2 Mouthings

Many manual signs are accompanied more or less regularly by mouthings. Mouthings are mouth movements that remind of the articulation of spoken language words. Some of these mouthings are quite conventionally attached to certain signs, while others appear to be more occasional, dynamic or even spontaneously combined (cf. Ebbinghaus & Heßmann 1996, 2001). While the linguistic status of these mouthings is not yet finally agreed upon in sign language linguistics, we consider mouthings to be some kind of lexical material from German integrated into DGS signing.

Most signers are to some degree bilingual and have at least some basic knowledge of German words and their core meanings. Mouthings in DGS originate from German words and bear the potential of

---

17   Further research might reveal that certain kinds of regular polysemy may be typical for or restricted to certain iconic sign types – whether the underlying image is a representation of an action (manipulative technique as in STIRRING), the outline of the form of an object (sketching technique as in VIERECK) or representing an object and its movement (substitutive technique as in BEAK see figure 5 above). For a description of image producing techniques in DGS cf. Langer (2005) and König, Konrad & Langer (2008: 388-389).

carrying and utilizing the semantic load of the mouthed word. This both supports and at the same time complicates the process of determining the various senses of a sign.

Mouthings can help to disambiguate polysemous signs. Often the mouthing is an important factor to determine the meaning of a polysemous sign in a specific context. It is not unusual to have one sign (for example BEAK see figure 5) to cover a basic category concept (such as "bird") and a set of subordinate concepts (such as "sparrow", "blackbird", "pigeon", "chicken" and a number of other birds with small narrow beaks) in everyday non-expert signing. One sign form can also cover complementary concepts such as "brother" and "sister" or antonymic concepts such as "dry" and "moist". The occurrences of such polysemous signs are usually accompanied by mouthings that have a disambiguating function and correspond to the meaning.

When a sign is conventionally associated with a particular mouthed word not all of the word's meanings automatically are also conventional meanings of that particular sign. Some of a word's meanings might not match the iconic value of the sign. However, in a specific context the sign form might just be used to contextualize that word, occasionally even for a meaning that is not lexicalized for the sign. Also, a polysemous German word might contribute some of its other or more marginal senses into sign utterances by way of mouthing and finally extend the sign's meaning in a way that might not be expected or not be in accordance with the sign's iconic core but parallels the meanings of the German word. One example for this is the sign CRAB. The underlying image of the sign shows the opening and closing of a pair of crab pincers. The German name for this animal is *Krebs*. But *Krebs* has also the meanings "cancer (disease)" and "Cancer (zodiac sign)". This is paralleled in DGS in that the DGS sign CRAB is also used in all three senses of the German word.

The interplay of a sign's iconic potential and its productive exploitation, the sign's lexicalized (conventional) meanings and the meanings of associated mouthed words make it difficult to determine the degree of conventionalization for a particular meaning. One major issue here is that a frequently used mouthing has the potential to bring additional meanings from another language (German) into the equation. This meaning potential of the word has to be considered especially when dealing with rather isolated signs as listed in sign collections and dictionaries that use spoken language equivalents to indicate the sign's meaning(s). Analyses of usage data from a large corpus can provide evidence for which senses are common usage and which are to be considered marginal or even occasional or creative occurrences. Until corpus data analyses of this kind will be possible we resort to the data as presented in the sign collections, documented in our internal database from previous projects and signer's intuition to identify and disambiguate possible senses to be validated in the feedback process.

4   Feedback from the Signing Community

In order to verify and complement our data, e.g. for regional variation and passive vocabulary, we rely on feedback from the signing community. This holds especially for the signs to be included in the Basic Vocabulary.

4.1   Preparation of Content

Each potential combination of sign form and meaning identified in the process described above has to be verified by the community. The crucial point is how to convey the intended meanings. Ideally context information in DGS would be given for each instance, however such an approach is cumbersome and not feasible within the project. Furthermore, in several cases senses seem clear enough from a short disambiguating written context such as "mouse (computer)" vs. "mouse (animal)". In cases like this we can spare the participants the effort of having to watch an additional film. For the feedback we chose a rather pragmatic approach: In order to make it as easy for the participants as possible, we present German equivalents of the signs, often along with German disambiguating contexts in brackets, and add signed contexts wherever we feel it is needed. German contexts can be a superordinate term, a domain, a relevant dimension (e.g. space, time), further disambiguating synonyms, a short explanation, collocations or other suitable context information. For DGS contexts we mainly use examples, but we also try short explanations, collocations and sign synonyms. For this we rely largely on the linguistic knowledge and intuition of a Deaf colleague. To a smaller extent, as far as it has been already made accessible through transcription, we also use data from our corpus as basis for the construction of examples.

4.2   The Feedback System

For the realisation of the feedback we decided for a crowd sourcing approach using an online feedback platform. Our aim is to get as many members of the signing community involved as possible, i.e. not only Deaf people but everybody using DGS.[18] Dealing with a relatively small language community spread across the country, the crucial point is not only how to attract enough participants, but also how to make them check back regularly. The newly developed online platform, running on desktop computers as well as mobile devices, allows for as much freedom for the participants as possible.[19] While the target community's pride of their own language and interest in supporting the project can be expected to be the main motivation for taking part in the feedback, computer game elements such

---

18   Metadata information will be used to weigh the participants' answers depending on their proficiency in DGS.

19   A more detailed description of the Feedback System in DGS and German can be found at http://feedback. dgs-korpus.de.

as levels and high score are used as extrinsic motivation. An important point for developing the system was to make it accessible by presenting all information in DGS and also to allow for answers and comments to be given in DGS (video upload via webcam). The system is optimised for our use, but the software is open source and can be adjusted for other purposes.

Participants create individual accounts where their answers and metadata are stored. Within the feedback system tasks are organised in categories and work packages. Due to the complexity of the tasks, users are asked to deal with only one type of task at a time. Categories therefore contain work packages of generally one type of task, and a new category is only released after a certain amount of work packages of the previous category have been completed. A so-called "golden trail" leads through the work packages. A help page contains explanations dealing with different aspects of the feedback system and with problems and questions that might arise. A "comment"-button on each page allows the participants to leave comments (in DGS or written German) that provide us with extra information or help us to spot problems regarding the tasks and the system.

4.3 Presentation of Content in the Feedback System

Work packages are organised in pages and rows. Task explanations and questions are all presented in DGS and written German (a button allows to switch between both languages), other content is provided as film, text or pictures depending on the purpose.

Figure 6: Example page of a work package "Form and Meaning"[20].

---

20   In this example the sign is made by two hands with "V" handshape moving up and apart. The meanings listed for verification are: "famous", "celebrity", and "public" (rows are moving upwards when answer button has been clicked).

Our first question type deals with sign forms and meanings as described above. Each sign form is first shown without mouth pattern in order to have the participants concentrate on the hands.[21] Participants are asked whether they know the sign form presented. Clicking "no" leads directly to the next page with a different sign form. When pressing "yes" further questions are revealed concerning the meanings of the respective sign form. In each row one meaning of the sign is presented by a film clip of the sign with a common mouth pattern and one or more German equivalents (often including sense indicating contexts in brackets). If needed an additional film clip provides context information in DGS also. Participants are asked whether they use the sign with the respective meaning themselves, don't use it but have others seen using it, or don't know this sign with this meaning at all. At the end of each page the system allows to add further meanings in written German or DGS (via webcam).

The completion time for each work package is planned to be approximately 20 minutes. On average a work package contains about eight different sign forms depending on the number of meanings to be verified (so far up to a maximum of 12 meanings with one sign form). When a work package is submitted the participant receives the number of points as defined for this package (i.e. independent of the number of "positive" or "negative" answers given) and is listed in the high score table. While this game-like approach primarily aims at motivating participants and boosting competition between them it also serves as a tool to stir which tasks are assigned to a participant (i.e. a certain number of points is needed to access the next category with different types of questions).

5   Conclusion

While the corpus collected within the project is still in the process of transcription and annotation we rely on other sources for our lexicographic work, especially for the production of our Basic Vocabulary. We have compared the signs of nine other previously produced sources (sign collections, learning materials) and based our lemma selection on the highest overlap in these products. The equivalents listed are taken as a starting point for a process of identifying different senses of the signs in a process of reversal of mono-directional bilingual information. In this process the iconicity of signs is also considered. The given meanings of signs are split into a finer and more sufficient granularity of senses, disambiguated through suitable contexts and presented via an online feedback system to the language community for verification or rejection.

The described procedure is not ideal to identify word senses of signs. It does things a little backwards, starting with the meanings of the German words given as equivalents instead of looking at signs and their contexts of actual use. However, this procedure takes the already published material containing

---

21   Considering mouthings as an essential part of signing raises the issue of how to present sign forms in the feedback as signers should not rate sign-mouthing but sign-meaning combinations. Our solution so far is to omit any mouth pattern for the more general question on sign form and show a separate film of the sign with the presumable most commonly used mouth pattern for each meaning. An accompanying explanation is included in the tasks explanation and pre-test have revealed no major problems.

information on signs' meanings seriously. Also, involving the language community in the feedback process has some advantages: While the corpus contains signing from only 330 informants we expect that through the feedback we can obtain data from even considerably more signers. This way we can obtain a more detailed account on regional variation concerning signs and sign senses. Direct questions also allow us to elicit data on passive as well as active vocabulary.

Nevertheless the result of such a feedback process can only be a first incomplete and approximated account of a sign's senses. This is the reason why we regard the Basic Vocabulary as preliminary. Once the corpus is available it will be the basis for further lexicographic work. Results from the feedback can be confirmed and complemented by corpus data, while the feedback process (crowd sourcing) may prove a valuable complementary method to obtain data for signs and questions the corpus does not cover.

6   References

Atkins, B.T.S, Rundell, M. (2008). *The Oxford Guide to Practical Lexicography.* New York: Oxford Univ. Pr.

Bentele, S., Konrad, R. & Langer, G. (2000). Transkription und Analyse. In R. Konrad, T. Hanke, A. Schwarz, S. Prillwitz & S. Bentele. Fachgebärdenlexikon Hauswirtschaft. Vol. 2. Hamburg: Signum, pp. 617-632.

Brien, D. (ed.) (1992). *Dictionary of British Sign Language/English.* London: Faber and Faber.

Ebbinghaus, H., Heßmann, J. (1996). Signs and words. Accounting for spoken language elements in German Sign Language. In W. H. Edmondson, R.B. Wilbur (eds.) International Review of Sign Linguistics 1. Mahwah, N.J.: Erlbaum, pp. 23-56.

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 language.* Hamburg: Signum, pp. 133-151.

Glaboniat, M., Müller, M., Rusch, P., Schmitz, H., & Wertenschlag, L. (2005). *Profile deutsch. Gemeinsamer europäischer Referenzrahmen.* Berlin, München: Langenscheidt. (Includes CD-ROM).

Hanke, T. (2004). HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts. In O. Streiter, C. Vettori (eds) *LREC 2004, Workshop proceedings: Representation and processing of sign languages.* Paris: ELRA, 2004, pp. 1-6.

Hanke, T., Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood & E. Thoutenhoofd (eds.) *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages.* Paris: ELRA, 2008, pp. 64-67.

Johnston, T. (1989). AUSLAN Dictionary. A dictionary of the sign language of the Australian deaf community. Victoria: Aust. Print Group.

Jones, R.L., Tschirner, E. (2006). Frequency Dictionary of German. Core Vocabulary for Learners. London: Routledge.

Jones, R.L., Tschirner, E. (2011). *Frequency Dictionary of German. Core Vocabulary for Learners.* (CD-ROM). London: Routledge.

Kennedy, G. (ed.) (2008). *A Concise Dictionary of New Zealand Sign Language.* Reprint. Wellington: Bridget Williams Books.

Kestner, K. (2002): Tommys Gebärdenwelt. Das Gebärdensprachbuch zur CD-ROM. Guxhagen: Manual Audio Devices.

König, S., Konrad, R., Langer, G. (2008). What's in a sign? Theoretical lessons from practical sign language lexicography. In J. Quer (ed.) *Signs of the Time. Selected Papers from TISLR 2004.* Hamburg: Signum, pp. 379-404.

König, S., Schmaling, C. (2012). Gebärdenschriften: Flüchtiges fixieren. In H. Eichmann, M. Hansen, J. Heßmann (eds.) *Handbuch Deutsche Gebärdensprache. Sprachwissenschaftliche und anwendungsbezogene Perspektiven.* Hamburg: Signum, pp. 341-356.

Kristoffersen, J.H., Troelsgård, T. (2010). Compiling a sign language dictionary. Some of the problems faced by the sign language lexicographer. In M. Mertzani (ed.) *Sign Language. Teaching and Learning. Papers from the 1ˢᵗ Symposium in Applied Sign Linguistics, Centre for Deaf Studies, University of Bristol, 24-26 September 2009.* Bristol: Centre for Deaf Studies, pp. 1-10.

Kristoffersen, J.H., Troelsgård, T. (2012). The electronic lexicographical treatment of sign languages: The Danish Sign Language Dictionary. In S. Granger, M. Paquot (eds.) *Electronic Lexicography.* Oxford: Oxford Univ. Pr., pp. 293-315.

Langer, G. (2005). Bilderzeugungstechniken in der Deutschen Gebärdensprache. In *Das Zeichen* 70, pp. 254-270.

Martin, W. (2013). Reversal of Bilingual Dictionaries. In R.H. Gouws, U. Heid, W. Schweickard & H.E. Wiegand (eds.) Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. (Vol. 5,4 of Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK)). Berlin: De Gruyter, pp. 1445-1455.

Nishio, R., Hong, S., 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. 4th Workshop on the Representation and Processing of Sign Languages.* Paris: ELRA, 2010, pp. 178-185.

Ordbog over Dansk Tegnsprog. Accessed at: http://www.tegnsprog.dk/ [20/03/2014].

Stokoe, W.C. (1993). Dictionary Making, Then and Now. In *Sign Language Studies*, 79 (22), pp. 127-146.

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

Sutton, V.J., Paul, F.A., Candelaria, I., Gunderson, J. (2009). *SignWriting Basics Instruction Manual.* SignWriting Pr.

Zwitserlood, I. (2010). Sign Language Lexicography in the Early 21ˢᵗ Century and a Recently Published Dictionary of Sign Language of the Netherlands. In *International Journal of Lexicography*, 23 (4), pp. 443-276.

Sign Collections for DGS

Sign collections for the Basic Vocabulary. Related or complementary works are grouped together.

[1] *777 Gebärden 1-3. Alle 3 Folgen auf einer DVD.* (2002). Version 2.0. Guxhagen: Manual Audio Devices. (DVD-ROM).

[2a] *DGS-Aufbau-Lexikon.* (1998). Aachen: Microbooks/Desire. (CD-ROM).

[2b] *DGS-Basis-Lexikon.* (1998). Aachen: Microbooks/Desire. (CD-ROM).

[2c] *DGS-Phrasensammlung.* (1998). Aachen: Microbooks/Desire. (CD-ROM).

[3a] Maisch, G., Wisch, F.-H. (1987). *Gebärdenlexikon. Band 1. Grundgebärden.* Hamburg: Verlag hörgeschädigte kinder.

[3b] Maisch, G., Wisch, F.-H. (1988). *Gebärdenlexikon. Band 2. Mensch.* Hamburg: Verlag hörgeschädigte kinder.

[3c] *Grundgebärden 1. Für Einsteiger.* (1999). Hamburg: Verlag hörgeschädigte kinder. (CD-ROM).

[3d] *Grundgebärden 2.* (2000). Hamburg: Verlag hörgeschädigte kinder. (CD-ROM).

[4a] Metzger, C., Schulmeister, R. & Zienert, H. (2000). *Die Firma. Deutsche Gebärdensprache do it yourself.* Hamburg: Signum. (CD-ROM).

[4b] Metzger, C., Schulmeister, R. & Zienert, H. (2003). Die Firma 2. Deutsche Gebärdensprache interaktiv. Aufbaukurs in Deutscher Gebärdensprache - Schwerpunkt Raumnutzung. Hamburg: Signum. (CD-ROM).

[5a] Arbeitsgruppe Fachgebärdenlexika (ed.). (1994). *Fachgebärdenlexikon Computer.* 2 Vol. Hamburg: Signum.

[5b] Arbeitsgruppe Fachgebärdenlexika (ed.). (1996). *Fachgebärdenlexikon Psychologie.* 2 Vol. Hamburg: Signum. Online-Version can be accessed at: http://www.sign-lang.uni-hamburg.de/plex/ [20/03/2014].

[5c] Konrad, R., Hanke, T., Schwarz, A., Prillwitz, S. & Bentele, S. (2000). *Fachgebärdenlexikon Hauswirtschaft.* 2 Vol. Hamburg: Signum. Online-Version can be accessed at: http://www.sign-lang.uni-hamburg.de/hlex/ [20/03/2014].

[5d] Konrad, R., Schwarz, A., König, S., Langer, G., Hanke, T. & Prillwitz, S. (2003). *Fachgebärdenlexikon Sozialarbeit/Sozialpädagogik.* Hamburg: Signum. Online-Version can be accessed at: http://www.sign-lang.uni-hamburg.de/slex/ [20/03/2014].

[5e] Konrad, R., Langer, G., König, S., Schwarz, A., Hanke, T. & Prillwitz, S. (eds.) (2007). *Fachgebärdenlexikon Gesundheit und Pflege.* 2 Vol. Seedorf: Signum. Online-Version can be accessed at: http://www.sign-lang.uni-hamburg.de/glex/ [20/03/2014].

[5f] Konrad, R., Langer, G., König, S., Hanke, T. & Rathmann, C. (eds.). (2010). *Fachgebärdenlexikon Gärtnerei und Landschaftsbau.* 2 Vol. Seedorf: Signum. Online-Version can be accessed at: http://www.sign-lang.uni-hamburg.de/galex/ [20/03/2014].

[6] Kestner, K., Hollmann, T. (2009). Das große Wörterbuch der deutschen Gebärdensprache. Der erste umfassende Gebrauchswortschatz in DGS als elektronisches Wörterbuch. 18000 Begriffe in Schrift und Video von A wie Aachen bis Z wie Zypresse. Deutsch – DGS, DGS – Deutsch. Bundeselternverband Gehörloser Kinder (ed.). Guxhagen: Kestner. [DVD-ROM with booklet.]

[7] Arbeitsgruppe ProViL. (2006). *DGS 3 und DGS 4. Aufbaukurse in Deutscher Gebärdensprache.* [Located at the eLearning platform WebCT/Blackboard; for members of the University of Hamburg accessible via OLAT - Online Learning And Training. Accessed at: *www.olat.uni-hamburg.de* [20/03/2014].

[8a] *Tommys Gebärdenwelt.* Version 3.0. (2007). Guxhagen: Kestner. CD-ROM.

[8b] *Tommys Gebärdenwelt.* 2. Tommy und Tina. Version 3.0. (2008). Guxhagen: Kestner. CD-ROM.

[8c] *Tommys Gebärdenwelt.* 3. Mit Tommys erstem Lexikon. Version 3.0. (2009). Guxhagen: Kestner. DVD-ROM.

[9a] Keller, J., Zienert, H. (2000). *Grundkurs Deutsche Gebärdensprache. Stufe I. Vokabel CD-ROM.* Hamburg: Signum. (CD-ROM).

[9b] Keller, J., Zienert, H. (2002). *Grundkurs Deutsche Gebärdensprache. Stufe II. Vokabel-Video.* Hamburg: Signum. (VHS-Video).

Acknowledgements