Project Note AP03-2018-01

Public DGS Corpus: Annotation Conventions

Authors: Reiner Konrad, Thomas Hanke, Gabriele Langer, Susanne König, Lutz König, Rie Nishio, Anja Regen

DOI: https://doi.org/10.25592/uhhfdm.822

First version: December 2018
Revision 1: February 2019: Clarified section on compounds, hyperlinks into public corpus added
Revision 2: September 2019: Change of type-subtype relation of lexicalized forms of signs based on manual alphabet, initialization, and cued speech
Revision 3: May 2020: Text adapted to changes in layout and content of release 3 (double-token tags, glossing conventions; cf. Hanke et al. 2020)

Abstract

This project note describes the annotation and glossing conventions as they apply to the Public DGS Corpus. In many aspects, they are identical to the annotation guidelines used in the DGS-Korpus project. However, not all aspects dealt with in the annotation show in Public Corpus. E.g. our annotation differentiates between different word forms by using qualifiers whereas the Public Corpus annotation just marks tokens as deviating from the citation form. In addition to explaining the annotation in translation and of mouthings, segmentation and lemmatisation, this texts explains specific approaches such as double glossing, double-token tags and lists special glosses used in the Public DGS Corpus. Glosses as well as some tokens are hyperlinked into the actual data, so the reader is invited to view both the original video as well as the annotation. The first appendix provides an overview of all symbols and gloss categories while the second lists the fingerspelling forms as used in DGS.

Contents

Introduction

This document describes the annotation and glossing conventions used in the public DGS corpus (http://ling.meine-dgs.de). Annotations include translations, glosses – used to identify the types tokens are matched to, comparable to Johnston's ID-glosses (Johnston 2010) –, and mouthings. Mouth gestures are annotated in a simplistic way by just adding "[MG]" in the Mouthing/Mouth gesture tier. These annotations are exported from iLex (Hanke/Storz 2008), our lexical database and annotation tool. Except for mouthings, all annotations are available in English and German (cf. language button DE|EN on the right side of the header). Downloadable annotation files provide both versions in separate tiers.

The update in September 2019 completes the proposed volume of about 50 hours video. It contains almost 356,000 tokens and recent changes of annotations since the first release. Older versions of a transcript or type entry remain available.

Translation

Following experiences from other corpus projects, we started with the translation first, then segmented and lemmatised tokens. In order to make the corpus data accessible to researchers neither knowing DGS nor German, the German translations were also translated into English.

Translation into German

Unlike other projects e.g. the Auslan Corpus Project, we did not aim for a free translation. Instead, the translation was supposed to be as close to the DGS utterance (source language) as possible.

A first and rough translation was made by contracted sign language translators or interpreters. These texts were aligned to the corresponding turns of informant A and B by rough timecodes. Next, student co-workers splitted and aligned these texts into 'sentence'-like utterances. These chunks of signed text should form a coherent and comprehensible meaning or utterance unit (cf. Johnston 2016:13-14). Further cues for chunking (or "boundary markers") are signals inherent to sign languages like palm-up-open-hand gesture (cf. below Gestures), head nod, body shift, gaze or rhythm. The students also proofread the translations and checked back with deaf team members where in doubt about the signing.

The more or less short written German sentences were aligned with the DGS video and serve several functions. They give access to the content of the signed texts even to those not knowing DGS and guide our (hearing) student annotators – with a varying competence of DGS – when lemmatising. They are searchable and define preliminary utterance units when looking for the context in which a sign token is used. As Johnston (2016:14) posits, these "translation sentences are not attempts to segment the [sign language] text into its potential language- specific syntactic or grammatical units".

On the website MEINE DGS (http://meine-dgs.de), the German translations are displayed as subtitles to the video clips.

For anonymisation reasons names of individual persons are represented by variables (#name1, #name2 etc.). Names of persons in public space or well known in the hearing and/or deaf world are not anonymised.

Translation into English

The purpose of the translation into English is to give access to the content of the DGS videos to those knowing neither DGS nor German. It is a free translation being more concise where suitable.

Segmentation

Defining the starting and ending point of a sign's token is a prerequisite for lemmatisation. One can either define the ending point of one sign as the starting point of the next sign – there is no gap between signs in the flux of natural signing, as there is no gap in the flux of speaking – or one can decide to delimitate transitional movements as not being part of the token's form. We opted for the latter in order to minimise visual noise when comparing tokens of one type or subtype. As a consequence, there are gaps between token tags in the annotation grid when there is a transitional movement between signs.
With the implementation of double-token tags, the segmentation is necessarily coarser with respect to the end of the manual activity of two-handed signs than would be possible with separate tiers for each hand. We concentrate on the active hand and ignore the non-active. As a consequence, the end of e.g. a hold is not specified but the fact that there is a hold is annotated in the token tag. For a detailed definition of starting and ending point of a sign cf. AP03-2010-01 (Hanke et al. 2012).

Lemmatisation

Before explaining our glossing conventions in more detail, we will briefly summarise the main aspects in which our approach differs from those of other sign language corpus projects:

- In order to take the iconicity of signs into account and to differentiate between conventional and productive combinations of signs and mouthings, we implemented a type hierarchy (double glossing) in the database model of iLex.
- We use double-token tags in the token tier instead of separate gloss tiers for left and right hand in order to facilitate the token-type matching.
- We use a lexical database where tokens are directly linked to types so that types can be easily accessed starting from a token in the transcript and tokens can be listed and reviewed by customised lists in the type entry.

The role of mouthings

Johnston (2010:115) lists translation and lemmatisation as "absolutely minimally required to begin building a machine readable reference SL corpus". For DGS, we consider it necessary to also include mouthings – and give an indication of apparent mouth gestures – in the first pass of basic annotation as mouthings are an important clue to the meaning of a DGS sign token. Mouthings are used to distinguish between conventional and productive uses of sign types. This approach is modelled in iLex as a type hierarchy. In combination with the sign form, mouthed words can be used to search for the appropriate sign type a token should be matched to.

Type hierarchy (double glossing)

In sign language lexicography (lexical) signs are treated as the words, that is, the units of the respective sign language to be described in a dictionary. One principle applied here is that of

idiomaticity (Johnston/Schembri 1999). However, taking into account that signers, due to the gestural-visual modality of sign languages, have the possibility to visualise meaning by directly linking a sign's form to the visual world alters the picture. The impact of iconicity is also reflected in the lexicon that differs from the lexicon of a vocal language in certain respects:

- Many signs are iconically motivated.
- A typical pattern in signing is lexical signs denoting something (*telling*) followed by so-called productive signs illustrating the intended meaning (*showing*).
- Many lexical signs can be modified gradually or de-lexicalised (re-iconisation for the purpose of *showing*).
- Signs, at least in DGS and many other European sign languages, are often coarticulated with mouthed words (so-called mouthings). This facilitates the phenomenon that signs often cover a wide range of different meanings.

Ignoring the role of iconicity and following the rule *same form, same meaning => same type* – and in consequence *same form, different meaning => different type –,* lemmatisation taking the observable mouthing into account would end up by mapping the lexicon of the vocal language onto the sign language lexicon, an inadequate and unsatisfying result (König et al. 2008). Ebbinghaus/Heßmann's functional perspective helps to understand the interplay of signs and words as mutual contextualisation. "Some such combinations occur with greater regularity than others and can be regarded as simultaneous collocations" (Ebbinghaus/Heßmann 2001:134). For example, the DGS sign SQUARE1^ (symmetrical two-handed sign with index fingers sketching a vertical square) in the DGS-Korpus frequently covers conventionalised meanings such as 'square', 'page', 'letter', 'recipe', or 'map' – the manual sign is regularly accompanied by the corresponding mouthings. However, SQUARE1^ is also used in combination with mouthings like "newspaper", "visa", "television", or "stole". All these mouthed words denotate concepts compatible with the iconic value of the sign tracing a square on a vertical plane, but these are not conventionalised meanings of this sign.

> Information about regular collocations with nonmanually produced units should be part of the lexicographic description of the manual lexicon of a sign language, though pairs of simultaneously related signals should not be reduced to the status of single signs. (Ebbinghaus/Heßmann 2001:134)

In order to support the lexicographic description in this sense, we modelled the impact of iconicity in iLex by implementing a type hierarchy and using double glossing. Types are linked to each other by a parent-child relation. In the database the parent type (in the following: type) is specified by a citation form. Each child type (in the following: subtype) stands for a conventionalised form-meaning relation. In most cases, these relations correspond to regular combinations of signs and mouthings. A subtype inherits its citation form and iconic value from the parent type. Type glosses should best give a hint to the iconic value of the sign whereas subtype glosses, like keywords, express a core meaning aspect.
At the first run of token-type matching, tokens of conventional sign-mouthing combinations are matched to the appropriate subtype – the token tag in the annotation grid shows the subtype gloss – and tokens of productive sign-mouthing combinations are matched to the type – the token tag shows the type gloss. Type glosses are indicated by a circumflex at the end, e.g. SQUARE1^. All other glosses represent subtypes.
At the end of the lemmatisation process, the range of meanings a sign type can cover is documented in a structured way, separating tokens of different conventionalised uses from each

other and from productive uses. This kind of pre-sorting supports the lexical description of sign types and their meanings. For example, in the type hierarchy the following subtypes are child types of the parent type SQUARE1^ used to express different conventionalised meanings as indicated by mouthings all motivated by the underlying image of the sign's form: CERTIFICATION2, DOCUMENT-OR-CERTIFICATE4, DOCUMENT-OR-RECORD1, FORM1, LETTER-MAIL2, MIRROR2, PAPER4, PICTURE2B, PIECE-OF-PAPER1, PILLOW1, PLAN5, POSTER2, SCREEN1, SHEET-OF-PAPER1, SIGN1, SQUARE1, TOWEL2, WINDOW6.

Double-token tags

Lemmatisation of sign language texts has to deal with the fact that signs are articulated with one or two hands. Two-handed signs can be divided into two-handed symmetrical and asymmetrical signs, in contrast to complex sign constructions where each hand articulates a different sign. In iLex we opted for one token tier that allows for annotating one type for each hand in order to make the annotation less time-consuming. Two-handed signs are either annotated in the right or left hand slot: For one-handed and asymmetric signs the slot of the active hand is used. For symmetric signs either the slot of the hand is filled that starts or moves higher than the other hand (e.g. in the case of point-symmetric signs such as TO-PLAY2 where one hand starts in the higher position than the other) or, when there is no difference in height, in the right hand slot (as default). (iLex automatically blocks filling the other slot if the first slot is filled with a sign whose HamNoSys notation determines it is two-handed.) Whether a signer is left-, right-, or mixed-handed can be evaluated ex post. There is no need to make a prediction of the signer's dominance beforehand.

The online transcript view now shows two columns for the subtype/type-gloss tier ("Lexeme/Sign"). A one-handed sign of the right hand is represented by a left-justified gloss in the left column. For two-handed signs (with an active right hand) both columns of that line are coloured. A one-handed sign of the left hand is represented by a right-justified gloss in the right column. Correspondingly, for two-handed signs (with an active left hand) both columns are coloured (cf. fig. 1).

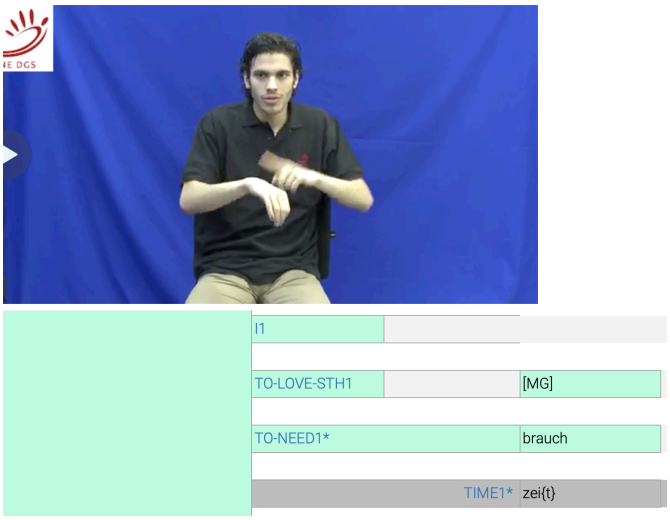

Figure 1: Glosses for right- and left-handed signs in the online transcript

The KWIC concordance view of the tokens of a type/subtype shows right-handed signs with a gloss in the upper line (r), for left-handed signs in the inferior line (l). For two-handed signs, there is no separating line and the gloss fills both slots (cf. fig. 2).

Figure 2: A token of TIME1 with KWIC concordance in the Types list

The downloadable ELAN files are exported from our iLex database. They follow Johnston's (2016:20-22) approach distributing tokens over separate tiers for each hand. Together with the type/subtype distinction this results in four tiers per informant: two tiers for left and right hand with subtype or type glosses (Lexeme_Sign_r_A, Lexeme_Sign_l_A; "Lexeme" corresponds to subtype, "Sign" to type) complemented by two tiers with type glosses (Sign_r_A, Sign_l_A). In addition, the corresponding tiers in German are displayed (Lexem_Gebärde_r_A, Lexem_Gebärde_l_A, Gebärde_r_A, Gebärde_l_A).

Lemmatising with iLex

Token-type matching or identifying (lexical) signs is ideally a strict top-down procedure. However, without a more or less complete lexical resource of DGS at hand, lemmatisation and building a lexical resource go hand in hand with constantly switching back and forth between top-down and bottom-up (König et al. 2010). Although basic annotation of sign language texts should be as theory-neutral as possible, it cannot be done without any theoretical assumptions. One assumption is the distinction of three sign categories: lexical signs (cf. Johnston/Schembri 2010: „fully-lexical signs"), productive signs (cf. Johnston/Schembri 2010: „partly-lexical signs"), and others (cf. Johnston 2016: "non-lexical signs"). The following glossing conventions allow to group and sort tokens of each sign category and to distinguish further subgroups.

With iLex lemmatisation is supported by a lexical database. Types and subtypes are separate database entities with unique IDs token are linked to. A type/subtype entry is defined at least by a gloss and a citation form in HamNoSys. For each type/subtype the matched tokens are listed with further information.

Constraints implemented in iLex guarantee that gloss names are unique. As glosses are labels for sign types/subtypes representing unique type entities in the lexical database there is no need to use the term "ID-glosses" (Johnston 2010) as every gloss in our system fulfils the criteria for ID-glosses, i.e. the glosses in the online transcript or the ELAN export files can be taken as ID-glosses. This is also true for the English glosses. German glosses on the type and subtype level also have different and unique English glosses. The English glosses also group lexical and phonological variants (cf. below: Glossing conventions) exactly as the German ones do.

Except for productive signs, glosses in the public DGS corpus correspond to unique type entries in our lexical database. As the distinction of productive signs in iLex is tentative and awaits further detailed transcription, tokens of productive signs are all labelled $PROD.

Glossing conventions

A gloss is a German word that corresponds – on the subtype level – to a core meaning (keyword) of the sign. When several words are needed, they are separated by hyphens. No spaces are used within gloss names. Glosses are written in upper case to indicate that they represent signs and not German words.

When adding a new type or subtype in the lexical database, the gloss name usually is followed by a number. Different numbers are used to differentiate between lexical variants, e.g.

WOMAN5 and WOMAN8. As we use German words for gloss names, we cannot get around the problem of German synonyms. So numbers are also used to differentiate between different signs, e.g. ZU3^ (to squeeze, squeezed) – ZU7 (closed) – ZU9 (towards; a sign used in cued speech) that share the same gloss name because their different meanings are covered by the same German synonym. When translating the German gloss names into English, we tried to take differences in meaning into account so that the corresponding English glosses were SQUEEZED3^ – CLOSED7 – TOWARDS9. In order to bridge the gap of English synonymy, we used combinations of English words like GAP-OR-DISTANCE1 (German gloss AB-STAND1) and DISTANCE-OR-RANGE1 (German gloss ENTFERNUNG1).
Phonological variants are grouped together by using the same gloss name and number followed by different letters, e.g. WOMAN2A, WOMAN2B, WOMAN2C, and WOMAN2D. English glosses get the same numbers and letters to distinguish lexical and phonological variants as the original German glosses do. These numbers and letters are not re-ordered to eliminate gaps, neither in the lexical database when types are deleted or gloss names changed, nor when glosses are exported for the public corpus. In a few cases, phonological variants on the subtype level do not correspond to phonological variations on the type level so that e.g. the subtypes TO-ACCOMPANY1C and TO-ACCOMPANY1D are both grouped under the type TO-ACCOMPANY1A^.
Clicking on a gloss in the transcript leads to the corresponding Types list entry where all occurrences of the corresponding type and subtypes are listed.
In release 3 these occurrences were given in a KWIC concordance view with three left and right neighbour glosses. Additionally, the translation of the accompanying tag is shown right after the link leading to the line in the online transcript (cf. fig. 2). And vice versa, clicking on one of these tokens leads to the corresponding transcript line where the video sequence can be viewed.

Figure 3: Heading of the entry TIME1^in the Types list

In case of existing pre-released entries of the dictionary DGS–German[1], a movie and the HamNoSys of the citation form as well a link to the dictionary entry (here: 354) and further lexical resources are given (cf. fig. 3; Müller et al. 2020).

When tokens differ from the type/subtype citation form, they are marked with an asterisk in the subtype/type-gloss tier of the transcript as well as in the Types list, e.g. PLANE1*. Tokens glossed without an asterisk are supposed to represent the citation form of the type or subtype.

[1] See http://dw-dgs.meine-dgs.de

Up to now, only a part of nearly 356,000 tokens have already passed lemma revision (Konrad/ Langer 2009).

Lexical signs

In order to build a corpus-based DGS dictionary, the focus of the DGS-Korpus project is on lexical signs, i.e. lexemes as entries of a sign language dictionary or lexical resource. Through language use, form and meaning of lexical signs became highly conventionalised units. That is why they are also called *established* or *frozen signs* (Brennan 1992). They are "relatively stable and consistent across contexts" (Johnston 2016:15). Out of context, either the meaning – asking a native signer for the meaning of this manual form – or the form – asking for a manual sign for e.g. 'alcohol' – will be spontaneously associated with each other. This means that lexical signs have a citation form and at least one core meaning.

Johnston/Schembri (1999) focus on the idiomaticity of lexical signs. Although the form of many lexical signs is iconically motivated – their formational components handshape, orientation, location, and movement express a general meaning (first level of conventionalisation), e.g. a flat hand represents 'something flat' –, the meaning of lexical signs is idiomatic i.e. it cannot be composed by summing up the meanings of its components: They have passed a second level of conventionalisation. For example taking the sign ALCOHOL1^, all its parameters are meaningful and can described as "holding something with a cylindric shape and pouring its content into the mouth". The conventionalised meaning 'alcohol' of the lexical item fits with this general meaning, but there is no rule that leads from the general meaning of the sign to 'alcohol'. The sign must be learned as a lexicalised and stable form-meaning pair. In contrast to productive signs (see below), Johnston/Schembri (2010) call them "fully-lexical signs".

There is a small number of signs where the general meaning coincides with the conventionalised meaning like e.g. PLANE1^ (subtype PLANE1: 'plane') or TO-BEAT1^ (subtype TO-BEAT1: 'to beat'). Johnston/Schembri (1999:133-134) call these signs "general signs" because the criterion of idiomaticity is not applicable (cf. also Fenlon et al. 2015:191). Nevertheless, as we do not emphasise idiomaticity but conventionality and take the iconicity of signs into account (see above) these signs are highly conventionalised form-meaning units and thus also treated as lexical signs.

When using iconically motivated lexical signs, the iconic value of its components is deactivated. Nevertheless, it can be re-activated (de-lexicalised) in order to express further meaning aspects, For example, the movement component of the sign ALCOHOL1^ can be modified accordingly to express '(to) abusively drink alcohol', maybe accompanied by appropriate (expressive) facial expression and body movement or posture.

Among lexical signs the following subgroups are distinguished by using prefixes or suffixes to gloss names. The prefixes are preceded by the dollar sign ($) in order to group these types together when sorting glosses alphabetically.

Name signs ($NAME)

Name signs for persons are labelled by $NAME, a collective gloss (subtype) for all name signs of private individuals. For anonymisation reasons, in the public DGS corpus no further differentiation is made. When a person in public space or well-known in the hearing and/or deaf world, the prefix is followed by the person's name, e.g. $NAME-ANGELA-MERKEL1.

Unknown (regional) signs ($CANDIDATE)

When a new sign is found that no one in our team knows and its data is not sufficient to verify its conventionality the sign is marked as a lexical sign candidate. Regular use by one or several informants of one region may suggest that it is a (regional) variant. Candidate signs are labelled by a keyword for the assumed meaning, and the suffix $CANDIDATE, followed by the code for the region where the data were collected and a running number as e.g. in AUGUST-$CANDIDATE-MST05. By this, these signs are sorted alphabetically in the Types list and extra labelled as candidates for lexical sign entries. Additional corpus data, verification outside the corpus or a more detailed analysis of the data may lead to the re-evaluation of the sign's status as a lexical sign and the removal of the $CANDIDATE suffix in the future.

Signs for (bound) morphemes in German ($MORPH)

A few signs are exclusively used to express bound morphemes of German words like e.g. the suffix "-in", a morphological marker to express female gender like in "Lehrerin" (female teacher). These signs are grouped by the prefix $MORPH-, e.g. $MORPH-FEMALE1.
In DGS like e.g. in Auslan (Johnston 2001) many lexical signs can function as noun, verb or other parts of speech without any formal morphological marker (Schwager/Zeshan 2008). To mark the difference e.g. between 'to teach' and 'teacher' the DGS sign TO-TEACH1 can be followed by the sign PERSON1 to make clear that the person who teaches is meant and not the act of teaching (token: TO-TEACH1). In the lemmatisation process these cases were segmented as two tokens and matched to the corresponding types/subtypes. Whether occurrences of PERSON1 in these contexts should be treated as morphemes or grammaticalised signs (like function words) has to be checked later on.

Foreign signs

When informants use foreign signs, e.g. in a constructed dialogue sequence or when discussing differences between sign language vocabularies, these tokens are labelled with gloss names of the surrounding vocal language and an indication of the respective sign language, as in NO-ASL1. Glosses for International Signs are marked by INTS, e.g. GERMANY-INTS1.

Among lexical sign, some signs are regularly accompanied by mouth gestures, so called multi-channel signs. These signs as well as compounds are not marked by prefixes or suffixes. The former do not require special segmentation conventions whereas the latter do.

Multi-channel signs (so-called idiomatic signs)

Some signs are regularly combined with one or more mouth gestures and do not have strongly associated mouthings. Some of these so-called *multi-channel signs* (Brennan 1992:128, Johnston/Schembri 1999:154-155) are used to express context-sensitive meanings and are difficult to translate into the surrounding vocal language by a single equivalent. The translation of such signs may result in a paraphrasing of the contextual meaning. The existence of these signs was felt to be an outstanding feature that makes sign languages distinguished from its surrounding vocal language. In the DGS, Swiss-German Sign Language (DSGS), and Austrian Sign Language (ÖGS) communities, there was a need to give these signs an extra name, *Spezialgebärden* (special signs, in use since the early 1990ies). Later on, these signs were also called 'idiomatic' signs (cf. Konrad 2014). However, from a language-internal perspective these signs are regular lexical signs and do not constitute their own distinct class of signs (König et al. 2012:142). In the public DGS corpus multi-channel signs are therefore not treated as a special category but as lexical signs among others, e.g. TO-LOSE-ONES-TRAIN-OF-THOUGHT1^.

Compounds and multi-word expressions

Becker (2003) claimed that composition is a minor and not active process in expanding the lexicon in DGS. Sequences of signs mirroring German compounds are not treated as fixed units but as sequences of individual signs in the lemmatisation. Whether these sequences should be classified as loan translations, collocations, (sequential) DGS compounds or something else is left to further analysis. The same applies to other types of multi-word expressions.

In contrast, simultaneous constructions such as blends and simultaneous compounds are treated as a single type. In simultaneous compounds, the members of the compound are produced at the same time, e.g. in initialised signs or AT-HOME1A^, which combines the handshape of TO-SIT1A^ with orientation, location, and movement of HOUSE1A^. In blends, the members of the compound are signed one after the other but at least one has been phonologically reduced. For example, in TO-KNOCK-ON-WOOD1^, the repeated movement of the initial member of the compound, TO-HOPE1B^, is deleted.

Productive signs ($PROD)

Unlike in vocal languages, signing discourse is not only a succession of lexical units, but an interplay between lexical signs (predominantly denoting something: *telling*) and productive signs (predominantly illustrating the intended meaning: *showing*). These productive signs – also known as classifier or polymorphemic signs – are fully iconically motivated. They have a general meaning to which each iconic value of its components contributes.

Following Johnston/Schembri (1999) they have passed a first level of conventionalisation and are built of conventional and non-conventional elements and are therefore called "partly-lexical signs" (Johnston/Schembri 2010). Their meaning can only be interpreted in context. A change of the sign's form always expresses a change of meaning. Productive signs did not pass a second level of conventionalisation. As a consequence, productive signs do not have a fully specified citation form. All instantiations of productive signs have the collective gloss $PROD (abbreviated for productive sign).

Pointing signs ($INDEX)

In Johnston's (2016) classification, pointing signs are *partly-lexical signs*. They have a conventional element, the handshape, and highly contextual ones: orientation and movement. However, unlike productive signs, pointing signs very likely stem from pointing gestures and cover a limited range of meanings: They point to a referent or a location (or both) and serve several functions.

Pointing signs are glossed by $INDEX. Variation in handshape (and orientation) is distinguished by adding numbers:

- $INDEX1: index handshape,
- $INDEX2: flat hand with palm up and finger tips pointing,
- $INDEX4: thumb.

In our studio we used screens placed low in front of the informants to show stimuli and videos explaining the elicitation task. As a consequence informants often point to the monitor when signing. In order to count these tokens separately we labelled them with $INDEX-TO-SCREEN1.

Tokens of pointing with the index handshape to the mouth in a pragmatic function – attracting the addressee's attention to lip-reading – were matched to the subtype $INDEX-ORAL1.

Tokens of the flat hand (palm down) making a circular movement in order to indicate or locate a referent to a specific area in the signing space, were glossed by $INDEX-AREA1. These signs can be analysed as a combination of an indexical sign with a sketching movement (Langer 2005:265).
Pointing signs can also be lexicalised. We opted for the lexical items I1 (pointing to the signer with contact to the chest meaning 'I') and YOU1 (pointing straight away from the body towards the addressee meaning 'you') because these are stable form-meaning units (but cf. Johnston 2016:25-29 who does not single out these signs from the class of pointing signs). Pointing to a part of the body can be a productive way to express the intended meaning, but there are also lexicalised signs for certain body parts like NOSE1A or HEART1A.

Others

The potential of corpus linguistics lies in the fact that linguistic units can be further analysed with regard to their context in natural speaking/signing, i.e. the signs preceding or following the sign in question are important clues to address various research questions. That is why lemmatisation should be done continuously without gaps in order to create running text words, in case of sign language annotation, running glosses. Besides lexical, productive, and pointing signs there are several linguistic phenomena in sign languages that must be covered by glossing conventions as well:

Fingerspelling ($ALPHA)

In DGS, fingerspelling is conventionalised using the one-handed manual alphabet (cf. Appendix 2) with some variation of handshape for the letters D, F, G, J, K, T, ß. Tokens are labelled by $ALPHA. According to the expressed meaning sequences of fingerspelling are segmented as one token tag. In iLex we use the type hierarchy with qualifiers (Konrad et al. 2012) and an open vocabulary to lemmatise fingerspelling. In the Types list of the public DGS corpus, all occurrences of fingerspelling are grouped by the type gloss $ALPHA^. By using different subtype glosses we distinguish different kinds of fingerspelling:

- $ALPHA1: one-handed manual alphabet,
- $ALPHA2: two-handed manual alphabet where letters are represented by using two hands to depict the form of the letter, e.g. using the small C-handshape with contact of the tips of index and thumb to form the letter S. This way of depicting the letter S is lexicalised for 'September' (SEPTEMBER2A).
  This kind of two-handed fingerspelling is different from the two-handed manual alphabet of BSL, glossed by $ALPHA-BSL^. Fingerspelling from manual alphabets of other sign languages are labelled correspondingly, e.g. $ALPHA-NZSL^.
- $ALPHA-SK: sketching the form of the letter in the air with the index finger (with the exception of Z as sketching the Z is the standard form in one-handed fingerspelling and thus glossed with $ALPHA1).

In addition to the gloss name we add the fingerspelled letters to give a hint to the expressed meaning, e.g. $ALPHA1:B-U-S. The letters added to the gloss name document what is fingerspelled and may differ from the intended word or meaning, e.g. $ALPHA1:A-L-F-E-D-O (proper name: Alfredo). If there is variation in handshape, the letters are extended by a number, e.g. $ALPHA1:T_2. When the signer intentionally does not fingerspell letters but moves the fingers in an indifferent way just to suggest that there should be some fingerspelling, these tokens are labelled by $ALPHA1:#.
Lexicalised forms of the one- or two-handed manual alphabet showing only one letter of a word like e.g. EX-PARTNER2 or WUPPERTAL2 are listed at the end of the type gloss $AL-

PHA^ by corresponding subtype glosses. Lexicalised forms showing several letters by manual alphabet handshapes are listed as separate types like e.g. BIO^.

Initialisation ($INIT)

In sign language linguistics initialised signs are lexical signs combined with a handshape from the manual alphabet usually representing the first letter of the corresponding (written) word, e.g. CHANCE1^. In the DGS Corpus, these signs are glossed separately like any other lexical signs when used conventionally.
Sometimes a phonological variant of a sign is an initialised form as e.g. RULE1C. In the Types list these forms are listed as subtypes under the respective type entry, in this example RULE1D^. The tilde symbol (~) ahead of the gloss indicates that the form of the sign is derived by changing one (or more) parameter of the type sign. In the case of initialisation the handshape is replaced by a handshape of the one-handed manual alphabet.
What we call initialised signs and label with the prefix $INIT are spontaneously produced signs with a handshape from the one-handed manual alphabet and a simple straight or circular movement or a wrist shaking. Often these signs are used for proper names or technical terms where the signer does not have a conventional sign at hand. Usually the handshape represents the first letter of the name or written technical term. Depending on the movement tokens are matched to the following types:

- $INIT-STRAIGHT1^,
- $INIT-HAND-WRIST1^,
- $INIT-CIRCLE1^,
- $INIT-CIRCLE2^.

Number signs ($NUM)

As there is a great variation in number signs of DGS, it was our aim to document each occurrence of number signs. We therefore segmented each token separately, even in compound numbers. For example the signed sequence to express the number 1989 is segmented into three tokens: $NUM-TEEN1:9 (nineteen), $NUM-ONE-TO-TEN1A:9 (nine), $NUM-TENS1:8d (eighty). (DGS shares an idiosyncracy with German that the last digit is articulated before the last-but-one if that is 2 or larger. In German, the example is articulated as “neunzehn|(hundert)|neun|und|achtzig” (nineteen|(hundred)|nine|and|eighty)).
As with fingerspelling, in iLex qualifiers are used to lemmatise number signs. In the public DGS corpus glosses for number have the prefix $NUM. Number signs are grouped by their form and according to the range of numbers they cover:

- 1-10 ($NUM-ONE-TO-TEN1A etc.),
- 11-19 ($NUM-TEEN1 etc.),
- 10, 20…90 ($NUM-TENS1 etc.),
- 100, 200…900 ($NUM-HUNDREDS1),
- 1000, 2000…10,000 ($NUM-THOUSANDS1).

The root of these signs is formed by a conventionalised movement and hand orientation. In the same way as in number incorporating signs the handshape(s) indicate(s) the numerical digit needed to express the corresponding number. Variants differing in movement and/or orientation are glossed with additional numbers and letters to the gloss name, e.g. $NUM-TEEN1, $NUM-TEEN2A, $NUM-TEEN2B, and $NUM-TEEN2C in the same way as used for lexical signs.

We add a code for the numerical digit handshape to the gloss name which in combination with the root is used to indicate the intended meaning of the sign represent a certain number as e.g. $NUM-TEEN1:3d (“thirteen”). Variation in handshape is coded with an additional letter in lower case (“d” for thumb (German: Daumen) when the thumb is used to represent the number, “f” for the F-handshape to represent the number 3, “w” for the W-handshape to represent the number 3).
For repdigits there are also conventionalised signs: $NUM-DOUBLE1A etc.
A tapping movement of the index finger or index plus middle finger on the thumb is used to express the numbers 11 and 12. The glossing reflects this typical movement: $NUM-TAPPING1.
A sniping movement is used express the numbers 11 to 19. These signs are glossed $NUM-SNIP1.
The following number signs do not allow number incorporation:

- $NUM-HUNDRED1 etc.,
- $NUM-THOUSAND1,
- $NUM-MILLION1.

Other roots requiring number incorporation are also labelled with the $NUM prefix:

- $NUM-CLOCK1A etc.,
- $NUM-GERMAN-MARK1,
- $NUM-GRADE1 (e.g. school),
- $NUM-MARK1A (e.g. school),
- $NUM-NUMBER-OF-PEOPLE1 etc.,
- $NUM-TIME-AFTER-NOW1^,
- $NUM-TIME-BEFORE-NOW1^,
- $NUM-WEEK-AFTER-NOW1,
- $NUM-WEEK-BEFORE-NOW1,
- $NUM-YEAR-AFTER-NOW1,
- $NUM-YEAR-BEFORE-NOW1,
- $NUM-YEAR-OF-APPRENTICESHIP1 etc.,
- $NUM-OR-BETWEEN1,
- $NUM-FROM-TO1,
- $NUM-RELATION1 etc. (match result),
- $NUM-DENOMINATOR1 (math.),
- $NUM-NUMERATOR1 (math.).

When lexical signs incorporate numbers, the handshape of the citation form will be replaced by the respective numerical digit handshape. In iLex this kind of modification is handled by adding the qualifier “quantity” in combination with a handshape code, in the public corpus an asterisk is added to the gloss because the token form deviate from the citation form of the types/subtype, e.g. YEAR1A*. Lexical signs incorporating numbers are the following: BEFORE1E, COMPARISON2, DAY-AFTER1B^, DAY-BEFORE1C^, DAY2, HALF6, HOUR1B, HOUR2A, HOUR2B, HOUR2C, MONTH1, OLD8B, PFENNIG1, SEVERAL1, STOREY1, TIMES3, TO-LIST1B, TO-LIST1C, TOGETHER3B, WEEK1A, WEEK1B, WEEK1C, YEAR1A, YEAR1B, YEAR3A, YEAR3B.
When numbers are signed depicting the Roman script we use the gloss $NUM-ROMAN1.[2]

[2] The Roman number IV is signed by tracing IIII with a four-finger handshape.

As for cardinal numbers, there are also signs with a root indicating ordinal numbers numerical digit handshapes, labelled $NUM-ORDINAL1 and $NUM-ORDINAL2.

List buoys ($LIST)

Another way of enumeration or counting is by pointing with the index finger (or flat hand) to the finger of the passive hand depicting an imaginative list. These "list buoys" (Liddell 2003:223-242) are used as ordinal or cardinal number signs or have a discourse structuring function. They are labelled by $LIST. In iLex list buoys are lemmatised as number signs using qualifiers and closed vocabularies. In the public DGS corpus we use $LIST1 for signs where the index finger pointing out one number of the list, and $LIST2 when the index finger is replaced by the flat hand.
Besides indicating one item of a given list, the list can be extended while indicating one item after the other ($LIST-TO-LIST1 etc.), items can be removed ($LIST-TO-REMOVE1A etc.) or grouped together ($LIST-TOGETHER1C). Grouping can be realised by a manipulative technique (Langer 2005, König et al. 2008, Ebling et al. 2015; $LIST-TOGETHER1C), by pointing consecutively with the index finger to several items ($LIST-TOGETHER3) or by incorporating the number of grouped items in the handshape of the indicating hand (cf. Liddell's TWO-LIST buoy; $LIST-TOGETHER2).
Tokens of list buoys always show two numbers, the first one representing the list item (the intended meaning), the second giving the total number of the shown list items (handshape of the passive hand) like in $LIST1:2of4. When items are grouped by number incorporation, there is a first number for the number of grouped items (handshape of the dominant hand), followed by the other numbers, e.g. $LIST-TOGETHER2:2:1-2of4.

Gestures ($GEST)

Signers just like hearing speakers use manual and nonmanual gestures. Manual gestures are holistic and expressive signs that are produced spontaneously. Manual activity that is neither a conventional nor a productive sign is likely to be a gesture when this activity is also used in the hearing majority with same or similar functions. Many gestures are culturally shared, some of them highly conventionalised (emblems), but there is still a great individual variation e.g. when expressing feelings.
When our annotators identify a gesture, the token is either matched to $GEST^ (unspecified type for all manual gestures) or $GEST-NM^ (unspecified type for all nonmanual gestures). The palm-up-open-hand gesture (aka PUOH) is glossed by $GEST-OFF^ (abbreviated for "offene Hand" (open hand)).
Beside these collective types, there are several gesture type entries specified by form and meaning much like lexical signs. Annotators can use these types when appropriate. Consequently, most of the gesture tokens in the public DGS corpus are grouped into manual, nonmanual, and PUOH gestures, others are already differentiated and matched to specified gesture types like e.g. $GEST-TO-PONDER1^. This differentiation is tentative and has not yet undergone review.

Mouthing (without manual activity; $ORAL)

When there is no significant manual movement and the meaning is only expressed by mouthing – which can be observed particularly in elder informants –, these tokens are labelled with $ORAL^, a placeholder for (voiceless) oral articulation (of German words) only. As a consequence, in the accompanying Mouthing/Mouth gesture tag a word or sequence of words has to be annotated, not a mouth gesture ([MG]).

Cued Speech ($CUED-SPEECH)

In the 1970ies a cued speech system was developed in Germany for better teaching the articulation of phonemes to deaf children was developed. Some of these cued speech hand signs are used in DGS, much like initialised signs, e.g. to express names where no conventional sign is at hand. The tokens are glossed with $CUED-SPEECH^. Over the time, several cued speech signs lexicalised. If the form of such a lexicalised sign simply corresponds to a cued speech sign, the subtype is listed at the end of the $CUED-SPEECH^ type entry as e.g. WHITE11. Sometimes a cued speech sign has a variant form as e.g. IF-OR-WHEN1A (cued speech sign for the articulation of the nasal 'N') and IF-OR-WHEN1B (different movement). These signs have separate type entries as well as lexicalized combinations of cued speech signs (see manual alphabet ($ALPHA)).

e.g. IF-OR-WHEN1 (cued speech sign to support the articulation of the letter N). Tokens of cued speech as well as tokens of lexicalised signs (subtypes) are grouped in the Types list under $CUED-SPEECH^.

Unsolved cases ($UNCLEAR)

Undecipherable manual activity which is assumed to be part of linguistic activity is labelled by the gloss $UNCLEAR^.

Extra-linguistic manual activity ($$EXTRA-LING-ACT)

Where particularly noticeable, extra-linguistic manual activity like e.g. rubbing ones nose or brushing off ones clothes, is tagged with $$EXTRA-LING-ACT^. This is meant to support machine learning of hand activity, i.e. these "tokens" identify sequences of strong visual noise (and not a kind of non-tokens, cf. Langer et al. 2016).

Annotation of mouthings and mouth gestures

Mouthings are very frequent in DGS. They are an important clue to the meaning of a DGS sign token which, in combination with the sign form, can be used to search for the appropriate type the token should be matched to. That is why we decided to also annotate mouthings in the phase of basic annotation.

In order to make the annotation less time-consuming we do not segment mouthings or mouth gestures independently from sign tokens. The Mouthing/Mouth gesture tier depends from the subtype/type-gloss tier, tags can spread over one or several token tags, always starting and ending when a token tag starts or ends.

Mouthings are annotated in lower case to make them distinct from German words. In contrast to fingerspelling (see above) where we want to document the fingerspelled letters regardless of the intended word, in annotating mouthing we focus on the identification of the mouthed word and not its actual articulation (form). This means that at least the intended word (word stem) to be lip-read should be annotated. Incomplete mouthings are supplemented in curly brackets. No supplement is made when the signer articulates part of a word in the process of searching for the right word.

Uncertainties, i.e. mouthings we could not identify, are marked by "??".

As mouthings in DGS refer to German words, the articulation features are different from e.g. mouthed English words. We therefore do not provide a translation of mouthings.

Mouth gestures are movements of the mouth region with no connection to words of the vocal language. With a focus on lexical signs, we did not aim for classifying mouth gestures by

form features. They are annotated in a simplistic way by just adding “[MG]” in the Mouthing/ Mouth gesture tier. Two exceptions apply:

- Mouthings are preceded by a hash tag (“#”, e.g. “#lalala”) when the signer does not want to articulate a certain word but to indicate that a person was speaking (orally articulating, e.g. instead of signing).
- When the signer imitates sounds of persons (e.g. interjections), animals, objects, or events like onomatopoetic expressions, these mouth actions are annotated by “[LM]” (German: Lautmalerei) followed by an appropriate orthographic transcription of the sound like “[LM:ähm]”, “[LM:bam bam]”, or “[LM:miau]”.

References

Becker, Claudia. 2003. *Verfahren der Lexikonerweiterung in der Deutschen Gebärdensprache*. Hamburg: Signum.

Brennan, Mary. 1992. The visual world of BSL. An introduction. In: Brien, David (ed.): *Dictionary of British Sign Language/English*. London: Faber and Faber, pp. 2-133.

Ebbinghaus, Horst / Heßmann, Jens. 2001. Sign language as multidimensional communication: Why manual signs, mouthings, and mouth gestures are three different things. In: Boyes Braem, Penny / Sutton-Spence, Rachel (eds.): *The hands are the head of the mouth: The mouth as articulator in sign language*. Hamburg: Signum, pp. 133-151.

Ebling, Sarah / Konrad, Reiner / Boyes Braem, Penny / Langer, Gabriele. 2015. Factors to Consider When Making Lexical Comparisons of Sign Languages: Notes from an Ongoing Comparison of German Sign Language and Swiss German Sign Language. In: *Sign Language Studies*, 16 (1), pp. 30-56.

Fenlon, Jordan / Cormier, Kearsy / Schembri, Adam. 2015: Building BSL SignBank: The lemma dilemma revisited. In: *International Journal of Lexicography*, Vol. 28, No. 2, pp. 169-206.

Hanke,Thomas / Storz, Jakob. 2008. iLex - A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In: Crasborn, Onno / Efthimiou, Eleni / Hanke, Thomas / Thoutenhoofd, Ernst D. / Zwitserlood, Inge (eds.): *LREC 2008 Workshop Proceedings. W 25: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*. Paris: ELRA, pp. 64-67. [Online resource; URL: http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf; last access: 2018-12-21].

Hanke, Thomas / Hong, Sung-Eun / König, Susanne / Konrad, Reiner / Langer, Gabriele / Matthes, Silke /Nishio, Rie / Regen, Anja. 2012: Segmentation. [Project Note AP03-2010-01, 1. Revision: Nov. 2012; online resource: URL: https://www.sign-lang.uni-hamburg.de/dgs-korpus/arbeitspapiere/AP03-2018-01.html].

Hanke, Thomas / Konrad, Reiner / Jahn, Elena / Schulder, Marc. 2020. Extending the Public DGS Corpus in Size and Depth. In: Efthimiou, Eleni et al. (eds.): *Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages. 12th International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, France, 2020. Paris, France: ELRA, pp. 75-82. [Online resource: URL: http://www.lrec-conf.org/proceedings/lrec2020/workshops/SIGN2020/pdf/2020.signlanglrec-1.12.pdf; last access: 2020-09-16].

Johnston, Trevor. 2001. Nouns and Verbs in Australian Sign Language: An Open and Shut Case? In: *Journal of Deaf Studies and Deaf Education* 6 (4), pp. 235-257. DOI: 10.1093/deafed/6.4.235

Johnston, Trevor. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. In: *International Journal of Corpus Linguistics*, 15, 1, pp. 106-131. DOI: 10.1075/ijcl.15.1.05joh

Johnston, Trevor. 2016. Auslan Corpus Annotation Guidelines. November 2016 version. [Online resource; URL: http://www.academia.edu/29690332/Auslan_Corpus_Annotation_Guidelines_November_2016_revision; last access: 2018-12-21].

Johnston, Trevor / Schembri, Adam. 1999: On Defining Lexeme in a Signed Language. In: *Sign Language & Linguistics* 2, 2, pp.115-185.

Johnston, Trevor / Schembri, Adam. 2010. Variation, lexicalization and grammaticalization in signed languages. In: *Langage et société* n° 131, mars 2010, pp. 19-35.

König, Susanne / Konrad, Reiner / Langer, Gabriele. 2008. What's in a sign? Theoretical lessons from practical sign language lexicography. In: Quer, Josep (ed.): *Signs of the time. Selected papers from TISLR 2004*. Hamburg: Signum, pp. 379-404.

König, Susanne / Konrad, Reiner / Langer, Gabriele / Nishio, Rie. 2010. How Much Top-Down and Bottom-Up do We Need to Build a Lemmatized Corpus? Poster presented at the Theoretical Issues in Sign Language Research Conference (TISLR 10), Sept 30 - Oct 2, 2010 at Purdue University, Indiana, USA. [Online resource; URL: https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/PosterTISLRTranskription1_R_12.pdf; last access: 2018-12-21].

König, Susanne / Konrad, Reiner / Langer, Gabriele. 2012: Lexikon – Der Wortschatz der DGS. In: Eichmann, Hanna / Hansen, Martje / Heßmann, Jens (eds.): *Handbuch Deutsche Gebärdensprache: Sprachwissenschaftliche und anwendungsbezogene Perspektiven*. Hamburg: Signum, pp. 111-164.

Konrad, Reiner. 2014. "Where have all the Idioms Gone". Ergänzungen zu Ines Schütte: "'Idiome und Redewendungen' in der DGS – Begriffsdefinition und Versuch einer Kategorienbildung". In: *Das Zeichen* 97, pp. 64-67.

Konrad, Reiner / Langer, Gabriele. 2009: Synergies between transcription and lexical database building: The case of German Sign Language (DGS). In: Mahlberg, Michaela / González-Díaz, Victorina / Smith, Catherine (eds.): Proceedings of the Corpus Linguistics Conference (CL2009). University of Liverpool, UK, 20-23 July 2009. [Online resource; URL: http://ucrel.lancs.ac.uk/publications/cl2009/346_FullPaper.doc; last access: 2018-12-21].

Konrad, Reiner / Hanke, Thomas/ König, Susanne / Langer, Gabriele / Matthes, Silke/Nishio, Rie / Regen, Anja. 2012. From form to function. A database approach to handle lexicon building and spotting token forms in sign languages. In: Crasborn, Onno / Efthimiou, Eleni / Fotinea, Stavroula-Evita / Hanke, Thomas / Kristoffersen, Jette / Mesch, Johanna (eds.): *Interaction between Corpus and Lexicon. Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages. 8th International Conference on Language Resources and Evaluation, LREC 2012*, Istanbul, Turkey. Paris: ELRA, pp. 87-94. [Online resource; URL: http://www.lrec-conf.org/proceedings/lrec2012/workshops/24.Proceedings_SignLanguage.pdf; last access: 2018-12-21].

Hanke, Thomas / Hong, Sung-Eun / König, Susanne / Konrad, Reiner / Langer, Gabriele / Matthes, Silke /Nishio, Rie / Regen, Anja. 2012: Segmentation.
[Project Note AP03-2010-01, 1. Revision: Nov. 2012. Online resource: URL: https://www.sign-lang.uni-hamburg.de/dgs-korpus/arbeitspapiere/AP03-2018-01.html; letzter Zugriff: 2018-12-21].

Langer, Gabriele. 2005. Bilderzeugungstechniken in der Deutschen Gebärdensprache. In: *Das Zeichen* 70, pp. 254-270.

Langer, Gabriele / Hanke, Thomas / Konrad, Reiner / König, Susanne. 2016. "Non-Tokens": When Tokens should not Count as Evidence of Sign Use. In: Efthimiou, Eleni / Stavroula-Fotinea, Evita / Hanke, Thomas / Hochgesang, Julie / Kristofffersen, Jette / Mesch, Johanna (eds.): *Workshop Proceedings. 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining. Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia, 28 May 2016. Paris: ELRA, pp. 137-142. [Online

resource; URL: http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-SignLanguage_Proceedings.pdf; last access: 2018-12-21].

Liddell, Scott K. 2003. *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Müller, Anke / Hanke, Thomas / Konrad, Reiner / Langer, Gabriele / Wähl, Sabrina, 2020. From Dictionary to Corpus and Back Again – Linking Heterogeneous Language Resources for DGS. In: Efthimiou, Eleni et al. (eds.): *Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages. 12th International Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, France, 2020. Paris, France: ELRA, pp. 157-164. [Online resource: URL: http://www.lrec-conf.org/proceedings/lrec2020/workshops/SIGN2020/pdf/2020.signlanglrec-1.26.pdf; last access: 2020-09-16].

Schwager, Waldemar / Zeshan, Ulrike. 2008. Word classes in sign languages. Criteria and classifications. In: *Studies in Language* 32, 3, pp. 509-545. DOI: 10.1075/sl.32.3.03sch

Appendix 1: Symbols and glossing conventions (summary)

Types list

Symbol	Explanation	Example
≙	Type gloss (not subtype) as heading for all tokens that are matched to this type	≙ PLANE1^
=	Subtype gloss (inheriting the same citation form as the parent type)	= BASE-GROUND3
~	Lexicalised signs (subtypes) that are derived from the type form by a change in one or more sign parameters	~ WALL-OR-SIDE3

Glossing

^	Type gloss (in contrast to subtype glosses which do not end with ^)	PLANE1^ vs. PLANE1
GLOSS1, GLOSS2 …	Lexical variants (or distinct signs with different meanings because of synonymic gloss names)	WOMAN5, WOMAN8 (ZU3^, ZU7, ZU9)
GLOSS1A, GLOSS1B …	Phonological variants	WOMAN2A, WOMAN2B
GLOSS-ASL, -BSL …, -INTS	Foreign signs from other national sign languages or international signs	GERMANY-INTS1
-$CANDIDATE-	Unknown (possibly regional) sign which is assumed to be a candidate for a lexical sign entry	AUGUST-$CANDIDATE-MST05
$	Precedes gloss name in order to group signs (subgroups of lexical signs, productive signs and others)	$PROD, $ALPHA1 … (see below)
$$EXTRA-LING-ACT	Extra-linguistic (manual) activity	$$EXTRA-LING-ACT^
$ALPHA	Fingerspelling	$ALPHA1, $ALPHA2, $ALPHA-SK, $ALPHA-BSL^, $ALPHA-NZSL^
$CUED-SPEECH	Signs of the cued speech system	$CUED-SPEECH^

$GEST	Gestures	$GEST^, $GEST-NM^, $GEST-OFF^, $GEST-TO-PONDER1^ …
$INDEX	Indexical (pointing) signs (with index finger, thumb, or flat hand/5-hand)	$INDEX1, $INDEX2, $INDEX4, $INDEX-AREA1, $INDEX-TO-SCREEN1, $INDEX-ORAL1
$INIT	Initialised signs (combination of manual alphabet and simple movement)	$INIT-STRAIGHT1^, $INIT-CIRCLE1^, $INIT-HAND-WRIST1^
$LIST	List buoys	$LIST1, $LIST2, $LIST-TO-LIST1, $LIST-TO-REMOVE1A, $LIST-TOGETHER1C
$MORPH	Lexical sign to visualise bound morphemes in German	$MORPH-FEMALE1 …
$NAME	Name sign	$NAME
$NUM	Number signs	$NUM-ONE-TO-TEN1A, $NUM-TEEN1, $NUM-TENS1, $NUM-HUNDREDS1, $NUM-THOUSANDS1 …
$ORAL	Oral articulation (of words) without significant manual activity	$ORAL^
$PROD	Productive signs (in contrast to lexical signs and others)	$PROD
$UNCLEAR	Undecipherable manual activity	$UNCLEAR^

Tokens

*	Token form differs from type/subtype citation form	PLANE1*
$ALPHA1:#	Intentionally suggested fingerspelling (and therefore not properly articulated)	$ALPHA1:#
_2, _3	Handshape variants	$ALPHA1:T_2

1d, 2d, 3d, 3f, 3w, 4d	Variation of handshape in number signs and list buoys: d = thumb (German: Daumen) f = F-like handshape (one-handed manual alphabet) w = W-like handshape (one-handed manual alphabet)	$NUM-TEEN1:3d
$LIST1:[number]of[number]	Indicated item of a given list (handshape)	$LIST1:2of4
$LIST-TOGETHER2:[number]:[number]-[number]of[number]	Number of items indicated simultaneously of a given list (handshape)	$LIST-TOGETHER2:2:1-2of4

Mouthings/Mouth gestures

#	Intentionally suggested oral articulation (and therefore not properly articulated mouthed words)	#lalala
??	Undecipherable mouthing	??
[MG]	Mouth gesture	[MG]
[LM:…]	Mouth action imitating sounds	[LM:bam bam]

Appendix 2: Manual alphabet (DGS)