Detailed Annotation and Qualifiers

Authors: Cornelia Loos, Reiner Konrad

DOI (last version): 10.25592/uhhfdm.10223

Version 0: 2021-07-09 (draft)

Version 1: 2022-04-04

Abstract

This project note provides an overview of which steps have been taken to make the DGS Corpus searchable for common morphological and syntactic phenomena in DGS: negation, indicating verbs, and compounding. We present the transcription methodology team's process in identifying factors that need to be considered in the transcription of these phenomena and how they should be annotated. Tokens with morphological negation, for instance, are matched to a qualified type, i.e. a child type of a type or subtype. The gloss of a qualified type is a combination of the type's or subtype's gloss and a qualifier code (in this case 'alph) that represents a specific form modification. We also present a list of qualifiers that could potentially be used for further detailed transcription and discuss quality assurance processes for gauging how these have been used by annotators and which qualified types need revision.

Overview of the topics

1 Introduction

1.1 Detailed annotation

From the beginning, the DGS-Korpus Project has aimed to provide detailed annotation[1] beyond token-type matching and roughly sentence-based translations. The following features and categories were envisaged for detailed annotation of lemmatized tokens:

- mouthing/mouth gestures,
- (lexicalized) facial expression,
- annotation of deviations from the citation form and grouping of frequently occurring deviant forms,

[1] We follow the terminology of Johnston (2019) in distinguishing between basic and detailed annotation assuming that annotation covers most of the processing steps of corpus data like translation, transcription, lemmatizing, tagging, or coding. Transcription in its broad sense is synonymous to annotation.

- meaning in context,
- syntactic category,
- use of signing space and referencing to established loci,
- productive use of signs in the mode of 'showing' – in contrast to 'telling'.[2]

Further on, in the lexical database type entries should be differentiated according to phonological and lexical variation, grammatical and iconic modification, and metaphorical uses. In order to determine syntactic and semantic functions of signs, the sentence-based translations should be further segmented into shorter units (clauses or phrases). Productive signs, i.e. classifier constructions, should be distinguished from established (lexical signs) and annotated in a way that supports specific retrieval and categorization. Furthermore, the distinction between conventional and productive uses of established signs (König et al. 2008) should be implemented in the lexical database.

While detailed transcription was first envisioned to follow basic transcription, we soon revised this decision and incorporated some features of detailed transcription into the base annotations conducted by our team of student assistants. They started annotating mouthing and/or mouth gesture during the first pass of basic transcription, i.e. segmentation and lemmatization. The reason was that mouthed words are a strong indication of the meaning of the sign. Besides form and iconic motivation, meaning is crucial to identifying a signed token, in other words to matching the token to the respective type or subtype in the lexical database iLex. Types correspond to signs, while subtypes represent conventionalized form-meaning pairings encoded by a sign. Matching tokens to subtypes and annotating mouthings saves having to specify for each token its meaning in context. This is only required in a few cases:

- When it is a person's name in order to support checking for anonymization.
- When two or more tokens express one meaning unit (e.g. a calque from a German compound) and there is no or insufficient mouthing. In that case the meaning will be annotated on a separate meaning tier with tags covering two or more tokens.
- When it is a productive sign, the meaning in context and the HamNoSys notation of (at least) the handshape(s) are annotated at the token tag.

Furthermore we did not defer the documentation of token form differences to a second annotation pass (see Johnston 2011 for a different strategy). The student annotators have to indicate when the token form differs from the citation form in the first pass of token-type matching, without specifying the exact nature of this difference. With the implementation of qualifiers in iLex in 2012 (Konrad et al. 2012) we aimed to systematically group the signs' potential for variation and modification in context. In the pass of lemma revision (i.e. looking through all the tokens matched to a type entry), all tokens with an unspecified indication of form deviation have to be either 'qualified' by matching them to an appropriate qualified type, or the form deviation has to be specified using HamNoSys. This means that lemma revision and detailed annotation are strongly intertwined.

We have deliberately refrained from assigning (sequences of) signs a syntactic category since there is as yet no standard annotation procedure in sign language corpus linguistics for defining sentences or clauses, nor for how to tag parts of speech or argument structure. Proposals like Johnston (2019) as well as our own preliminary analyses (see Section 5 on reference tracking) show that it is crucial to define sentence or clause-like units on the basis of

[2] 'Showing' concerns classifier constructions (in our terminology 'productive signs') built on the spot to visualize meaning, but it also encompasses established (iconic) signs that may be re-iconized in order to express contextually supported meaning (Konrad et al. in press). In contrast 'telling' refers to a mode of signing that relies mostly on conventionalized signs.

not only manual signs but also nonmanual components (e.g. in constructed action), and mouthing/mouth gestures.

1.2 Qualifiers

Qualified types are combinations of types with qualifiers. Qualifiers are built-in form features in iLex with feature values as controlled vocabularies. They represent such common form deviations as single vs. repeated path movement, variation in the number of hands with which a sign is produced, or changes in the place of articulation indicated in the citation form. Qualifiers are designed to signal a deviation in form, but in some cases, this regularly occurring form is associated with a particular meaning. The qualifier alpha_negation ('alph)[3], for example, is used to annotate a particular movement pattern superimposed on a citation form in order to negate the sign's meaning.

When specifying a token's form deviation, the annotator can either choose from a list of already existing qualified types[4] or they can build a new type-qualifier combination resulting in a new qualified type. Using the type hierarchy (Konrad et al. 2020), these qualified types are child types of a parent type. With a few exceptions, the form of these qualified types is notated in HamNoSys, just like it is for the parent type. The following illustrations show how a token can be matched to a qualified type:

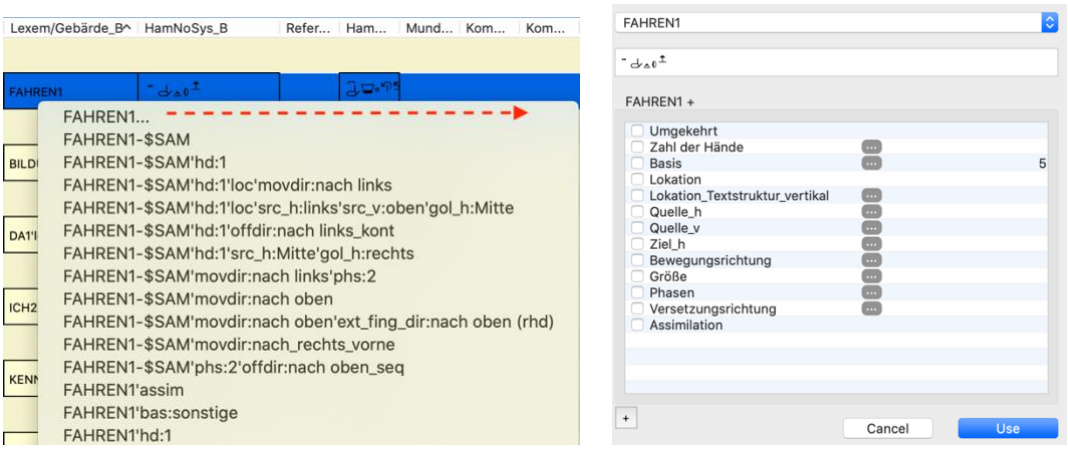

Figure 1: Use the context menu in an iLex transcript to either choose an existing qualified type or to create a new one.

Each type and subtype entry has a 'child' tab, which lists all dependent types in the type hierarchy. For the subtype TO-DRIVE1[5] (FAHREN1;) this looks as follows:

[3] When creating a qualified type, the code of a qualifier preceded by an apostrophe and the selected feature value are automatically added to the type or subtype gloss. For examples see Figure 1.

[4] In the following, all that is said about qualified types is also valid for qualified subtypes.

[5] For further information on the use of glosses in this text see footnote 9 below.

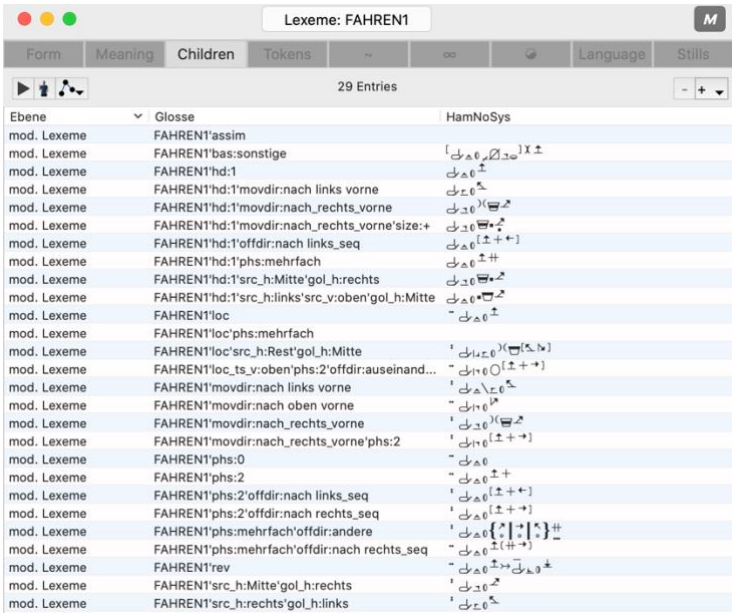

Figure 2: List of qualified types of the subtype TO-DRIVE1 (FAHREN1)

Up to now, 32 qualifiers are used to group form features of signs. In addition, 8 qualifiers refer to signs of the manual alphabet, 7 qualifiers apply to signs for numbers, one signals cued speech and another is used if a number is signed as a super- or subscript. Note that qualifiers and their values are a practical way of annotating the forms and meanings of manual alphabet and number signs. In these cases, however, the feature value always represents the (form and) meaning of a token. Appendix 1 lists all 49 qualifiers grouped by parameter and function. Of the 32 qualifiers that group form features of signs, 22 have predefined feature values (i.e. controlled vocabularies) to specify the form feature. Most of the qualifiers grouping form features are used to specify the movement parameter (22). Some qualifiers have to be combined, e.g. a repeated sideward movement is annotated by a combination of the qualifier phases ('phs; indicating the number of movement phases, i.e. repetitions) and offset_direction ('offdir). Appendix 2 lists the same qualifiers sorted by the number of tokens that are matched to qualified types with the respective qualifier. In other words, Appendix 2 gives an indication how often a particular qualifier has been used in the annotation process. Due to the fact that many qualified types are built by combinations of two or more qualifiers, the number of tokens (column 'Uses_of_qualifiers') is much higher than the absolute number of tokens matched to qualified types.

Table 1 illustrates the proportion of tokens matched to qualified types compared to tokens that have been flagged by student assistants as having a form deviation. The total number of tokens in the DGS Corpus is ~ 658,000 (end of July, 2021), the Public DGS Corpus has 359,450 tokens. In the following, we exclude tokens for the manual alphabet and number signs as they do not constitute form deviations that we annotate in order to further analyze their grammatical or lexical function.

	Tokens	**Tokens of qualified types**	**%**	**Tokens with deviant form**	**%**
DGS Corpus	633,871	111,106	17.5	178,609	28.2
Public DGS Corpus	347,801	68069	19.6	97,919	28.2

Table 1. Overview of annotated tokens in the Public DGS Corpus that have a form deviation vs. that been matched to a qualified type.

This table is instructive, since the Public DGS Corpus marks both tokens with qualified types and tokens with a form deviation with an asterisk (*) (except for tokens from the manual alphabet or numbers). The table shows that taken together, both account for nearly 50 per cent of all tokens in the Public DGS Corpus. We intend to review and further analyze some of these qualifiers in order to publish this information in future releases and make the Public DGS Corpus searchable for phonological, morphological, syntactic, and semantic phenomena. The semantic and grammatical aspects in question are (sentential) negation, directionality in verbs, reference tracking, movement repetition (e.g. for reduplication or aspect marking), and number of hands used in a sign (e.g. for marking of lexically inherent plurality).

In the following we discuss our annotation methodology with respect to these phenomena in light of the sign language research literature. We start with how complex signs are annotated in general (Section 2) to provide enough background for the discussion of alpha and other types of morphological negation (Section 3). We then discuss argument marking via directional verbs and an annotation schema for referent tracking in the corpus (Section 4 and 5). Lastly, we lay out our annotation scheme for two form deviations associated with different semantic and grammatical notions: movement repetition (Section 6) and variation in the number of hands (Section 7).

2 Complex signs

Adjacent signs in the signing stream may simply be integrated syntactically, or they may form a more tight-knit lexical unit such as a compound or a root-affix structure. The decision whether two adjacent signs should be treated as one lexical unit or not depends on several factors including semantic transparency, frequency of co-occurrence, and phonological reduction over time. Since the goal of (detailed) transcription is to document the data and thus provide a foundation for future decisions on, e.g. the compound status of frequently co-occurring signs, we refrain from interpretation and annotate each sign separately. In line with this general rule, entries in the DW-DGS only list 'frequent combinations' and 'phrases' without making claims about their morphological or lexical status. The one exception are *Zusammensetzungen* (combinations) such as EHREN1A^AMT1A (TO-HONOUR1A^AGENCY1A),[6] which point to calques of German compounds that are assumed to have compound status in DGS as well. In the DGS corpus, only complex signs whose components have undergone some level of phonological fusion are annotated as compounds or affixoids.

In order to adequately represent morphological processes such as negation in the process of detailed transcription, we looked at different types of complex signs. The goal was to annotate the spectrum of negated forms attested in the corpus consistently.

2.1 Compounding

Compounds are complex signs in which two or more independent roots combine into a new word whose meaning can be derived from the meaning of its parts (for endocentric compounds) (Lieber & Štekauer 2011; Bauer, Lieber & Plag 2013). Hence *homework* is typically work done at home (as opposed to at school) and when you *waterproof* your home, you protect it from water. Compounds can be classified according to semantic, morphological, and phonological criteria. Semantically, we distinguish determinative and coordinative compounds. Only the former have a clearly identifiable head that determines the semantic class of the compound (e.g. *homework* is a type of *work*) as well as its morphological behavior (e.g. word class and plural formation are determined by the head *work*, not by *home*). An example of a coordinative compound is the ASL sign APPLE^ORANGE^BANANA 'fruit' (Bellugi & Newkirk

[6] See the DW-DGS entry 628 (STEMPEL1^(STAMP1^)), Bedeutung #2 (Reading #2).

1981), where the compound denotes a concept that serves as hyperonym to each compound member. Similarly, in DGS we find FORK-FOR-EATING1^KNIFE1A (GABEL-ESSEN1^MESSER1A) used for 'cutlery'.[7] Morphological considerations have led to further classification of compounds in sign languages. Classifier compounds, for instance, consist of a lexical sign and a classifier such as RED^CL:rectangular 'brick' (Klima & Bellugi 1979). Chain compounds, a term introduced by Lepic (2015), are complex signs that consist of a fingerspelled borrowing from the ambient spoken language and an established lexeme or a classifier. These are commonly used in ASL when there is no conventionalized sign for a particular concept yet, or in order to establish a sign. The ASL compound A-B-S-T-R-A-C-T^SUMMARIZE 'abstract', for instance, establishes the sense 'abstract' for the sign SUMMARIZE. This is comparable to spoken language compounds such as the German *Chai-Tee* (chai), where the first member is a recent borrowing from another language and contributes a slightly more specific meaning than the second member ('spiced milk tea' vs. 'tea').[8] Lastly, we classify compounds along phonological criteria into sequential, blend, and simultaneous compounds. Since this latter distinction bears the most weight in our annotation process, we discuss it in the following section.

2.2 Sequential, blend, and simultaneous compounds

During the process of conventionalization, members of a compound typically undergo phonological reduction and assimilation (Klima & Bellugi 1979). The duration of a compound is shorter than that of the corresponding phrase, the components assimilate in terms of handedness, place of articulation, and handshape, and the transitional movement between the signs becomes more fluid. Phonotactic constraints exert pressure on compounds to conform to the monosyllabic structure of single lexemes.

Some sequences of signs should be classified as compounds based on their not fully predictable meaning relation (a 'wooden house' SHELF2C^^HOUSE1A[9] (FACH2C^^HAUS1A) is made of wood but a 'boat house' TO-ROW1^^HOUSE1B (RUDERN1^^HAUS1B) is not built of boats but contains them) – but they exhibit no degree of phonological fusion. We have termed these *sequential compounds*. In order to have access to the tokens of each component of such constructions, the tokens are segmented and lemmatized individually in the DGS Corpus. This allows us to analyse frequent collocations and decide afterwards whether to interpret them as calques or as compounds in their own right. The latter would result in further tagging of multi-sign lexemes as proposed by Hanke et al. (2012).

By now, these (potential) compounds can be found in the iLex database by searching for tags in the tier *Gesamtbedeutung* (overall meaning). Tags in this tier span two or more tokens as (possible) compound members and indicate the overall meaning via a controlled vocabulary (the concepts list). Another strategy for identifying sequential compounds is to start from a list of German compounds and check whether they occur as mouthings in the corpus. Mouthing tags whose value corresponds to a German compound and that span more than one token tag are good candidates for sequential compounds. A third strategy is a customized search for

[7] Note that in this example, FORK is signed first, which differs from the German phrase *Messer und Gabel* 'knife and fork'. This suggests that the combination does not constitute a borrowing from German.

[8] In fact, both *Chai* and *Tee* represent recent vs. older borrowings of the Northern Chinese *ch'a* vs. the Fukien Chinese *t'e*, respectively (Duden online: https://www.duden.de/node/180445/revision/180481). Only *Chai* is still perceived as a borrowing, however.

[9] For each sign, we provide its English gloss as it appears in the Types list of the Public DGS Corpus followed by its German gloss in parentheses. Glosses may be followed by a digit, a letter, and/or a circumflex (^). Digits stand for lexical variants, i.e. signs with a similar meaning but an unrelated form; letters represent phonological variants, and the circumflex indicates the sign level in the type hierarchy: Type glosses always have a circumflex whereas subtype glosses never have one. For further information see Konrad et al. (2020; Type hierarchy, Glossing convention). Since the circumflex is traditionally also used to link the members of a compound, you may see two adjacent circumflexes in some of these examples.

frequent collocations of a type or subtype. This is a standard procedure in the context of lexicographic analysis.

If a complex sign exhibits phonological fusion of two independent components and therefore evidence of development into a single lexeme, we classify it as a *blend compound* and annotate both components as one sign. Recent research on sign language morphology has proposed the terms *lexicalized compound* (Lepic 2015) or *semi-simultaneous compound* (Quer et al. 2016) for these cases, but we follow Plag (2015) in using the term *blend* to denote "compounds with at least one constituent having lost some of its phonological material" (2015: 2423).[10] We treat a sequence of two signs that semantically behave as a compound as a blend in case at least one of the following is true: a) The signs have assimilated in terms of number of hands; b) The complex sign does not have two separate path movements and the transition between the components is fluid; c) None of the signs are repeated.[11] Case a) is exemplified in TO-DIVIDE2^ (AUFTEILEN2^), where the second component LOCATION1A^ (ORT1A^) is signed with two hands to assimilate to the two-handed TO-SEPARATE1C^ (TRENNEN1C^). Case b) is attested in DONT-KNOW1^ (WISSEN-NICHT1^), where the single path movement of the sign leads smoothly from the final (and only) hold of TO-KNOW-OR-KNOWLEDGE2A^ (WISSEN2A^) into the arc-shaped movement of GONE-TO-LOSE-STH1^ (WEG-VERLIEREN1^).[12] Following criterion c), a positive lexeme and its negator may be annotated separately if either sign is repeated. This is the case in this combination of TO-KNOW-OR-KNOWLEDGE2A (WISSEN2A) + NO1B (NEIN1B), where NO1B is repeated.

Lastly, *simultaneous compounds* may be formed by combining the phonological parameters of two signs at the sub-morphemic level. Simultaneous compounds are discussed in Santoro (2018) and in Lepic (2015), where they are called blends. In a few cases, simultaneous compounds retain most phonological parameters of both signs because each hand signs one compound member. This is the case in DIGITAL-WAAGE1 (digital scales), where the non-dominant hand signs one tray of an old-fashioned scales moving up and down (see SCALE-OR-LIBRA1A (WAAGE1A)) while the dominant hand signs DIGITAL1A^ (DIGITAL1A^). More often, however, one or two phonological parameters of one sign are combined with the remaining parameters of the other sign:

Handshape: In AT-HOME1A^ (ZUHAUSE1A^), HOUSE1A^ (HAUS1A^) is signed with the handshape of TO-SIT1A^ (SITZEN1A^).

Location[13]: In PSYCHIATRY1^ (PSYCHIATRIE1^), the sign PILL1^ (TABLETTE1^) is signed at the location of the sign INSIDE1A^ (INNEN1A^) and maintaining the repeated downwards movement of INSIDE1A^.

Movement: In DIFFERENT5^ (ANDERS5^), the sign TO-LIST1C^ (AUFZÄHLEN1C^) is signed with the movement of DIFFERENT2^ (ANDERS2^).

---

[10] Our use of the term differs from Lepic (2015), who defines signed blends as synchronic combinations of two signs rather than as the result of a phonological fusion operation acting on two sequentially composed signs (his definition of blends is closer to what we call *simultaneous compounds*). Note further that Fradin's (2015) handbook article on blending in spoken languages characterizes the word formation process as not productive but simply creating hapax legomena. We do not wish to make such claims about phonologically fused compounds in our data. At least in the realm of positive lexeme + negator blends, compounding is a productive process in DGS.

[11] We also consider assimilation in handshape, orientation, and location, but since these also occur in phrases, we only consider them as supporting evidence of phonological fusion if any of the criteria listed in a)-c) are met.

[12] This token of DONT-KNOW1 best illustrates the fluid transition between signs.

[13] Simultaneous blends with a change only in location are rare. In contrast, incorporation of location with a change in meaning is quite a common form of sign modification in DGS. In the iLex database, these forms are either identified by using the qualifier location_on_body ('bodyloc) or lexicalized as individual type entries with cross-references to the base forms. An example of the latter is C-SECTION1^, a lexicalized form of TO-CUT2D^ and identical in form with TO-CUT2D^'bodyloc:belly (region).

Non-dominant hand: In TO-MARRY3A^ (HEIRATEN3A^), the non-dominant hand of TO-MARRY1A^ (HEIRATEN1A^) is replaced with the non-dominant hand of RING1A^ (RING1A^).

Both blend and simultaneous compounds are annotated as single signs. In iLex the respective type entry contains information about and links to each component member. For instance, in the type entry AT-HOME1A^ (ZUHAUSE1A^; Figure 3) the checkbox 'composed of' is checked and the two slots provide a link to the corresponding component types HOUSE1A^ (HAUS1A^) and TO-SIT1A^ (SITZEN1A^). The distinction between blend and simultaneous compound is marked via a metadata comment for the respective type.

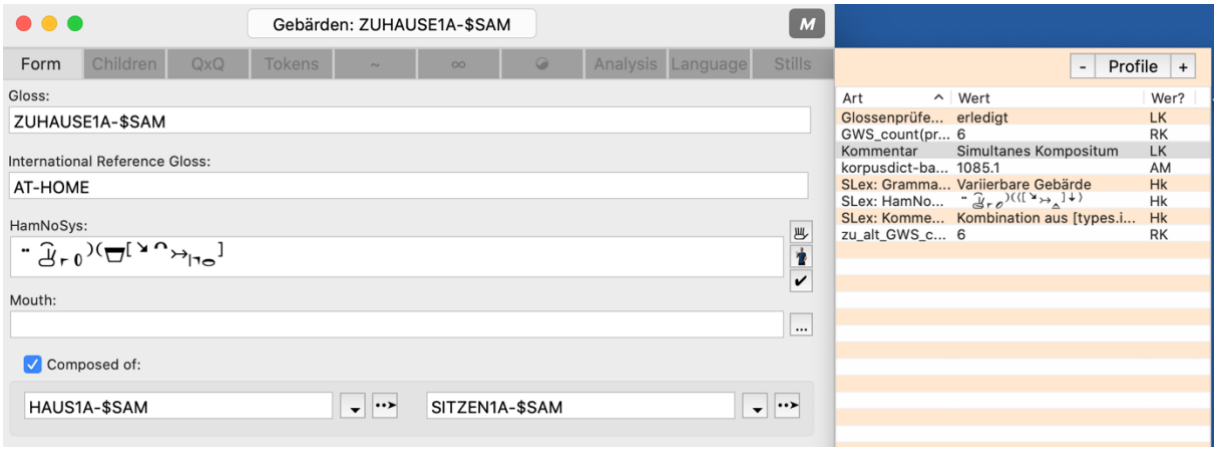

Figure 3: iLex type entry AT-HOME1A^ (ZUHAUSE1A^) with the checkbox 'composed of' and links to the blend components. On the right side, the metadata contain the comment *Simultanes Kompositum* (simultaneous compound; highlighted in grey).

2.3 Negators as affixoids

Some of the complex signs in the DGS corpus have one component that recurs over a large number of compounds with the same meaning and may therefore be in the process of developing into an affix. Some of the negators discussed in the next section might more accurately be considered affixoids: In contrast to affixes, they have free form equivalents that serve as independent negators, but like affixes, they are often phonologically reduced and can be combined productively with a number of different roots. The negators in question are GONE-TO-LOSE-STH1^ (WEG-VERLIEREN1^), GONE-TO-LOSE-STHG2^ (WEG-VERLIEREN2^), NOT1^ (NICHT1^), NO3A^ (NEIN3A^), and NO3B^ (NEIN3B^) . Since compounds and affixoids are both instances of complex signs, we represent them in the same way in iLex: If they show some degree of phonological fusion, they are annotated as single types and their components are indicated and linked in the type entry.

3 Negation

According to Pfau (2008), sentential negation may be expressed by a manual or a non-manual negator in DGS, that is sentences may be negated solely by the non-manual negator headshake (see also Zeshan 2006). Headshake has not been annotated in a systematic way in the DGS corpus yet, but where annotators note it, they transcribe it as a comment tag stretching over several token tags.[14] Schulder and Hanke (2020, 2019) plan to use automatic pose estimation

[14] In addition, non-manual negation co-articulated with a lexical sign and not spreading (cf. Quer 2012: 335) is annotated via the qualifier head_shaking ('h_s), for instance in DEAF1A'h_s 'not deaf' (TAUB-GEHÖRLOS1A'h_s; see e.g. DEAF1A* (TAUB-GEHÖRLOS1A*)).

data generated by OpenPose to detect headshakes across the DGS corpus, which will feed a more systematic annotation process. In this paper, however, we focus on the annotation of manual negators. Sentential negation is expressed with the free morphemes NO1B^ (NEIN1B^), NOT1-6 (NICHT1-6), NONE1-6 (KEIN1-6), or ZERO1-6 (NULL1-6).[15] Some of these morphemes are subtypes of signs that regularly occur with different degrees of phonological fusion with the predicate they negate: GONE-TO-LOSE-STH1^ (WEG-VERLIEREN1^), GONE-TO-LOSE-STH2^ (WEG-VERLIEREN2^), NOT1^ (NICHT1^), NO3A^ (NEIN3A^), and NO3B^ (NEIN3B^). These amalgamation and reduction processes are indicative of a productive morphological process such as compounding or affixoid formation. Hence, we have decided to annotate morphological negation systematically and plan to make these annotations available in the Public DGS Corpus. Three annotation strategies are currently used:

- The qualifier alpha_negation ('alph) is added to glosses that represent signs with a conventionalized movement pattern expressing negation.
- As a blend compound whose component parts are the positive lexeme and the negator. While the compound is annotated as a single type, its internal complexity is flagged in the gloss name, which ends in -NICHT 'not' or LOS '-less'.
- The positive lexeme and the negator are lemmatized as separate tokens if they represent calques from a German word. In these cases, the gloss of the negator carries the prefix $WORTTEIL ($MORPH) to indicate that it represents a German negative prefix or suffix. The prefixes used in these glosses are *ab-, ent-, un-*; the suffix is *-los*.[16]

We discuss each of these strategies in the context of negative morphemes that have been postulated for DGS and show how these are currently annotated in the DGS corpus project.

3.1 Negative affixes or clitics

According to Schwager (2012), DGS has two negative affixes. The alpha suffix (-NEG in Schwager's terminology) consists of a movement appended to the positive form of sign. The movement has variously been described as tracing the shape of the fish symbol[17] or the Greek letter α. Schwager claims that this affix developed historically from the sign NOT1^ (NICHT1^) (NICHT1 in Schwager's terminology), presumably because a) the sign occurs with a range of negative meanings including sentential negation and b) the sign's upwards-facing palm aligns with the final palm position of the alpha movement. Given that only (a shortened form of) the movement of NOT1^ has been retained and the transitional movement between positive form and NOT1^ has been reanalyzed as the characteristic alpha-movement pattern, this pattern is morphemic and has been analyzed as an affix or clitic because it does not have a free morpheme equivalent. In Zeshan's (2006: 49-54) terminology these forms are "irregular negatives" because of their restricted distribution.

A second negation element consists of the spread or 5-hand starting with an upward or sideward palm orientation and ending with a downward palm orientation. The sign adjusts in

[15] NICHT, KEIN, and NULL have several lexical variants each, which we have abbreviated as 1-6 here for the sake of ease of presentation. The variants are: NOT1 (NICHT1), NOT2 (NICHT2), NOT3A (NICHT3A), NOT3B (NICHT3B), NOT4 (NICHT4), NOT5 (NICHT5), NOT6 (NICHT6), NONE1 (KEIN1), NONE2 (KEIN2), NONE3 (KEIN3), NONE4 (KEIN4), NONE5A (KEIN5A), NONE5B (KEIN5B), NONE6 (KEIN6), ZERO1A (NULL1A), ZERO1B (NULL1B), ZERO2A (NULL2A), ZERO2B (NULL2B), ZERO3 (NULL3), ZERO4 (NULL4), ZERO5 (NULL5), ZERO6A (NULL6A), ZERO6B (NULL6B).

[16] The corresponding type entries are: $MORPH-OFF1A ($WORTTEIL-AB1A), $MORPH-OFF1C ($WORTTEIL-AB1C), $MORPH-AWAY-OFF1A ($WORTTEIL-ENT1A), $MORPH-AWAY-OFF1B ($WORTTEIL-ENT1B), $MORPH-AWAY-OFF1C ($WORTTEIL-ENT1C), $MORPH-UN1 ($WORTTEIL-UN1), $MORPH-UN2 ($WORTTEIL-UN2), $MORPH-UN3A ($WORTTEIL-UN3A), $MORPH-UN3B ($WORTTEIL-UN3B), $MORPH-UN4 ($WORTTEIL-UN4), $MORPH-UN5 ($WORTTEIL-UN5), $MORPH-LESS1 ($WORTTEIL-LOS1), $MORPH-LESS2 ($WORTTEIL-LOS2), $MORPH-LESS3 ($WORTTEIL-LOS3).

[17] Secret symbol of early Christians, known as the "sign of the fish" or the "Jesus fish".

place of articulation to the preceding root.[18] Schwager assumes that it developed from the sign GONE-TO-LOSE-STH1^ (WEG-VERLIEREN1^) or GONE-TO-LOSE-STH2^ (WEG-VERLIEREN2^) (OHNE in Schwager's terminology). Happ & Vorköper consider the combination of this form (glossed FREI in their terminology) with a positive sign compounding; Schwager calls it an affix. We consider affixoid the appropriate term, as the sign is productively used (like an affix) but has a free morpheme equivalent (GONE-TO-LOSE-STH1/2^) and its semantics is more functional than that of (some of) the senses of the free morpheme (for a definition of affixoid, see Ruppenhofer et al. 2018).

Two further negation affixoids are based on NOT1^ and NO3A/B^ (NEIN3A/B^). After a positive lexeme, both occur with various levels of phonological fusion.

Some lexemes exhibit variation as to the negators they occur with. The concept 'weak, powerless' can be expressed by combining FORCE1A (KRAFT1A) '(physical) strength' with either $SPECIAL-NONE8^ ($SPEZIAL-KEIN8^) (the suppletive negative existential/possessive) or with GONE-TO-LOSE1^ (Happ & Vorköper 2006: 374, where the morpheme is glossed FREI).

3.1.1 Alpha negation

Most DGS grammars note that alpha negation is subject to morpho-phonological constraints (e.g. Happ & Vorköper 2006; Schwager 2012). Specifically, it applies to the modal predicates MUST1 (MUSS1) 'have to, be necessary', SHALL1 (SOLLEN1), ALLOWED1 (DARF1) 'may, be allowed', CAN1/POSSIBLE1 (KANN1/MÖGLICH1), and TO-NEED1 (BRAUCHEN1) as well as the cognitive predicates TO-KNOW-OR-KNOWLEDGE2A (WISSEN2A) 'know something or somebody', TO-KNOW-STH-OR-SB1A/WELL-KNOWN4 (KENNEN1A/BEKANNT4) 'know, be familiar' and to-BELIEVE2A (GLAUBEN2A) and the epistemic predicate RIGHT-OR-AGREED1A (STIMMT1A) 'that's right'. Happ and Vorköper (2006: 373) also mention INTEREST1A (INTERESSE1A) 'interesting' as taking alpha negation, but this form has so far not been attested in our annotations. What these signs share in common phonologically is a downward movement at the wrist (Pfau & Quer 2008). Further signs that exhibit alpha negation in the DGS corpus are TO-WANT2 (WOLLEN2), THERE-IS3 (ES-GIBT3), IT-WORKS-OUT1 (KLAPPT1), TO-ADJUST1 (ANPASSEN1) 'fit, adjust', TO-HELP5 (HELFEN5), and ALSO1A (AUCH1A) 'also, too'. Note that both TO-WANT2 (WOLLEN2) and THERE-IS3 (ES-GIBT3) have suppletive negation forms that exist in parallel with morphological negation: DONT-WANT1 (WOLLEN-NICHT1), DONT-WANT2 (WOLLEN-NICHT2), $SPECIAL-NONE8 ($SPEZIAL-KEIN8). Alpha negation is sometimes accompanied by a head shake.

In terms of its morphological status, alpha negation has been analyzed as either an affix or an (en)clitic to a modal verb (Pfau 2008). The fact that alpha negation can attach to non-predicative signs such as ALSO1A (AUCH1A) 'also, too' advocates in favor of a clitic analysis, given that clitics are freer than affixes in their co-occurrence restrictions (esp. with respect to word class) (Zwicky & Pullum 1983). The sentence-final position of lexemes with negative meaning like NOTHING1A (NICHTS1A), NOBODY1 (NIEMAND1), and NOWHERE1 (NIRGENDWO1) are also compatible with a reanalysis of the sentential negator NOT1^ (NICHT1^) as a postverbal clitic.

As part of the quality assurance process, we conducted an inspection of 555 corpus tokens that had been tagged for alpha negation, which show variation in how the alpha movement is performed (phonologically conditioned allomorphy). Aside from a classical alpha-shaped movement, which accounts for 42 per cent of the tokens (232), the following forms occurred:

- major variation (form deviates substantially from the classical alpha shape):
    - ra (ऱ) movement: the palm moves to face side-or upwards but then returns to a downward facing position, describing an arc to one side and then back again

[18] This negator follows its base in 17 types in iLex, but it is also attested as preceding the base in 2 types.

(182 tokens). The final hand orientation is annotated in a given token tag as a form deviation via the HamNoSys symbols .
    - only the first half of the alpha movement is performed (74 tokens), this is especially common in negated forms of CAN1/POSSIBLE1 (KANN1/MÖGLICH1) (41 tokens). In the token tag, this half movement is annotated as , where the citation form of, for instance, CAN1^ (KANN1^) would have .
- minor variation:
    - the palm faces sidewards rather than downwards at the end of the movement (13 tokens)
    - after completing the alpha movement, the palm returns to a downward-facing position (4 tokens)

Deaf members of the transcription team considered these forms to be allophones of the same negator rather than as conventionalized phonological variants. Moreover, at least eight signers used the classical alpha and the ɾ-movement interchangeably for the types MUST1 (MUSS1) 'must, have to', POSSIBLE1 (MÖGLICH1) and POSSIBLE2 (MÖGLICH2), CAN2B (KANN2B), RIGHT-OR-AGREED1A (STIMMT1A), and TO-KNOW-OR-KNOWLEDGE2A (WISSEN2A). These observations suggest that the above forms are phonologically conditioned allomorphs of the alpha morpheme and should be annotated in the same way: by adding the qualifier 'alph to a sign with negative meaning that exhibits one of the movement modifications described above.[19]

The qualifier 'alph is *not* used when there is a handshape change during the sign into a 1, 5, or B-handshape. The latter handshapes are characteristic of the manual negators NOT1^ (NICHT1^), GONE-TO-LOSE1^ (WEG-VERLIEREN1^), and NO3A^ (NEIN3A^)/NO3B^ (NEIN3B^) and allow the identification of a separate negation morpheme.[20] Likewise, 'alph is not used when the non-dominant hand is added during the second part of the negated lexeme – a change in the number of hands too would indicate the status of the negated form as a sequential blend compound. An example with a handshape change and a change in the number of hands is SPEECHLESS2B^ (SPRACHLOS2B^), which consists of $INDEX-ORAL1^ ($INDEX-ORAL1^) and GONE-TO-LOOSE-STH2^ (WEG-VERLIEREN2^).

3.1.2 Spread hand negation (GONE-TO-LOSE1/2^)

In contrast to alpha negation, the second negation morpheme attested in the corpus is specified for all phonological parameters and has a free morpheme equivalent. It generally follows the predicate it negates but in two cases precedes it (e.g. NOT-PRESENT-OR-HERE1^ (DA-NICHT1^) and NOT-BAD1^ (SCHLECHT-NICHT1^)). Given the independent morpheme status of this negator, we do not annotate it via a qualifier, which typically marks changes to the base form of a

[19] We are aware that a potential consequence of this decision is that we lump together at least two different negative morphemes under the umbrella alpha negation. Specifically, ɾ-movements could have a different origin than the classic alpha movement. Given their final downward-facing palm orientation, ɾ-movements may derive from NO3A^ (NEIN3A^) or NO3B^ (NEIN3B^), which both have negative meanings ('no, not, never') and a downward-facing palm moving from the center of the signing space towards the ipsilateral side. Evidence for this claim comes from compound-like tokens combining e.g. TO-KNOW1B (KENNEN1B) and NO3A^ (NEIN3A^) in a fluid ɾ-movement (SEE TO-KNOW-STH-OR-SB1B* (KENNEN1B*). However, synchronically, the two forms have the same meaning and are perceived as instantiations of the alpha movement by DGS signers, which is why we annotate them in the same way.

[20] Note that some positive signs exhibit these handshapes (e.g. TO-BELIEVE2B (GLAUBEN2B), TO-KNOW-STH-OR-SB1A (KENNEN1A)) as well. If they are negated with an alpha-shaped movement, they receive the qualifier 'alph, although alpha movement in these cases looks identical to a compound of, e.g., TO-KNOW-STH-OR-SB1A (KENNEN1A) and GONE-TO-LOSE1^ (WEG-VERLIEREN1^).

positive lexeme. Instead, there are two ways to annotate GONE-TO-LOSE1/2^ (WEG-VERLIEREN1/2^):

1) as part of a blend compound, one of whose components is GONE-TO-LOSE1/2^ (WEG-VERLIEREN1/2^)
2) as a separate sign

Which annotation strategy is selected depends on a) the presence of mouthing on the negator and b) whether we observe phonological reduction of the positive base form and/or the negator (e.g. loss of movement repetition, assimilation in place of articulation). If the mouthing *los* ('less') is present, the negative element receives its own gloss $MORPH-LESS1 OR $MORPH-LESS3 ($WORTTEIL-LOS1 or ($WORTTEIL-LOS3), which signals that the complex sign in question is a calque from a German word ending in the negative suffix *-los* ('-less'). Often, such calques can be identified via their mouthings in the corpus, as these end in *-los* and span at least two token tags. Other than that, the gloss name $MORPH ($WORTTEIL) signals that the negator and the preceding sign form a complex sign.

If *los* is not present as a mouthing and there is evidence of phonological reduction in the base or in the negator, the token is matched to a blend compound type that comprises the base and the negator. This results in the following parallel annotations:

Compound component	**Separate negator**
GONE-TO-LOSE-STH1^ (WEG-VERLIEREN1^)	$MORPH-LESS1 ($WORTTEIL-LOS1)
GONE-TO-LOSE-STH2^ (WEG-VERLIEREN2^)	$MORPH-LESS3 ($WORTTEIL-LOS3)

The manual form of the two $MORPH-LESS types in the right column are identical to their corresponding blend components GONE-TO-LOSE1/2^. We opted to keep the parallel annotation scheme for two main reasons. On the one hand it facilitates identifying German calques such as *arbeits-los* ('unemployed'), annotated as two tokens and lemmatized with TO-WORK1 (ARBEITEN1) and $MORPH-LESS1 ($WORTTEIL-LOS1). On the other hand it helps to avoid an inflated number of qualified types resulting from phonological reduction in compounding. Take the hypothetical TO-HEAR'phs:0[21] | $MORPH-LESS1'assim_loc 'deaf', where the first sign has been phonologically reduced (it consists only of a hold) and the second has assimilated to the first in place of articulation (the ear instead of a location in neutral space). By annotating this as part of a blend compound or affixoid instead, we do not inflate the number of tokens of the kind TO-HEAR'phs:0. This is what we want, because decisions on the base form of a type are in part based on the frequency of occurrence of a variant, and we would not want to claim that the base form of TO-HEAR has no movement component.

3.1.3 NICHT1^ negation and NEIN3A/B^ negation

Lastly, some signs are negated with the help of negators NOT1^ (NICHT1^) or NO3A^ (NEIN3A^)/NO3B^ (NEIN3B^), that is, alpha negation coexists with its likely origin, negation via NOT1^. Here too, the negator may be annotated as part of a compound/affixoid or as a separate sign, depending on the presence of phonological reduction in the base form or negator as well as *los* mouthing (see above) or *un* ('un') mouthing (articulating a German word starting with the negative prefix *un-* ('un-')):

Blend/compound component	**Separate negator**

[21] Qualifier phases ('phs) with feature value = 0 means that there is no movement.

NOT1^ (NICHT1^)	$MORPH-LESS2 ($WORTTEIL-LOS2)
NO3A^ (NEIN3A^)	$MORPH-UN3A ($WORTTEIL-UN3A)

As it stands, the corpus distinguishes three different annotation strategies for different degrees of fusion between the predicate and its negator. Depending on the form of the negator, the same positive sign may be annotated in three different ways:

'alph: TO-KNOW-OR-KNOWLEDGE2A^'alph (WISSEN2A^'alph)
complex sign: DONT-KNOW2A^ (WISSEN-NICHT2^) (blend compound of TO-KNOW-OR-KNOWLEDGE2A^ (WISSEN2A^) and NO3A^ (NEIN3A^)
separate tokens: TO-KNOW-OR-KNOWLEDGE2A (WISSEN2A) followed by NO3A^ (NEIN3A^) as in TO-KNOW-OR-KNOWLEDGE2A NO3A (WISSEN2A NEIN3A)

Tokens lemmatized by using a type or subtype with alpha qualifier as well as tokens matched to a complex blend compound both show some level of phonological reduction. In contrast, segmenting and lemmatizing sign and negator as separate tokens, this is warranted by separate mouthings for each sign (e.g. *weiß* and *nicht*), the addition of the non-dominant hand for NO3A^ (NEIN3A^) and the fact that there was no phonological reduction.

4 Directional verbs

Space serves a number of grammatical functions in sign languages, most importantly tracking discourse referents and indicating the grammatical relations between a predicate and its argument(s) (see Sandler & Lillo-Martin 2006 and references therein). Referential determiner phrases (DPs) can be anchored in a particular location in neutral space (a referential locus) by, for instance, a pronominal pointing sign or by signing the DP itself at the referential locus. Predicates can then make use of referential loci to indicate who did what to whom by setting their initial and/or final locations to where their agent/source, patient, or goal argument is anchored in space. In signs without path movement, the predicate may be signed at an argument's referential locus or its palm orientation may be directed towards that locus. Predicates that indicate their argument(s) in this manner have variously been called *agreeing verbs*, *directional verbs*, and *indicating verbs* (see Mathur & Rathmann 2012 for an overview). We opt for the theory-neutral term *directional verb* and do not follow the classical distinction between agreeing and spatial verbs, as both index their arguments by changing their initial and/or final location and only differ in the semantic role of the arguments they index (agent/patient vs. locations).

In order to decide whether a given token is used predicatively and indexes its arguments, one would need two additional types of annotations that are not currently present in the corpus. On the one hand, clause- or proposition-like units need to be annotated in order to establish which overt DPs enter into grammatical relations with a given predicate. On the other hand, the locations in neutral space at which these DPs have been set up need to be annotated in order to ascertain which argument the starting or end location of a given predicate token is indexing. Clause-level annotations were discussed at a corpus-internal workshop in May 2020 and form the subject of Elena Jahn's dissertation project, but have as of yet not been implemented systematically. Annotation of DPs will be discussed in section 5 (Reference tracking).

In light of these facts, we have opted for a two-pronged annotation strategy for directional verbs. Any predicate token whose starting and/or end point(s) deviate from the citation form will be marked with a qualifier that specifies the starting and/or end location of the sign. That way potentially directional verbs are identified by their form, but predicate tokens whose initial

or final location deviate from the citation form for independent reasons are also included. Existing qualifiers that may be adopted for this function are source ('src), goal ('gol), movement_direction ('movdir), location ('loc), and palm_orientation ('palm_orient). The qualifiers 'src and 'gol[22] were originally designed to mark only those deviations in starting and endpoint of a sign that function to index arguments, while 'movdir was conceived as a means of annotating any other deviation in movement direction. However, a purely formally oriented annotation of deviations from starting and endpoint should be consistent. If 'src and 'gol are to be used for this purpose, their values need to be expanded to include the sagittal axis. It is not advisable that 'movdir takes on this function, since a large set of values to mark initial locations would have to be added ('movdir currently only specifies directions towards a final location). For signs whose entire place of articulation differs from the citation form, 'loc may be used, and 'palm_orient can serve to annotate deviations in palm orientation, which often mark object agreement in predicates that do not have a path movement. The second annotation strategy for directional verbs requires more in-depths analysis of argument structure and reference tracking and will only be applied in exemplary fashion to subparts of the corpus. For this purpose, we developed a reference tracking annotation schema and workflow that we present in the next section.

5 Reference tracking

To describe accurately how a given sign may be used in predication, we need to know how many and which types of arguments a sign occurs with and whether and how these arguments may be indexed on the predicate. In addition to annotating changes to the initial and final (or overall) location of the predicate itself, which may be used to index arguments, we need to annotate the arguments themselves in a given stretch of discourse.

Some initial ideas were discussed but then discarded: Each referent that serves as argument to a predicate should receive their own ID. This ID cannot be provided by the token-tag ID of the predicate itself, since predicates may introduce more than one argument but each referent should receive a unique ID. Secondly, the same referent would potentially receive different IDs if it serves as argument to more than one predicate. Further, there needs to be a way to tag referents that are not explicitly mentioned but that can be inferred from the discourse (e.g. a waiter in a restaurant scenario). We opted to introduce the new tiers “References_A/B” for for each participant in all transcripts, which allow creating a maximum of four tags within one token-tag on the subtype/type-gloss tier (“Lexeme/Sign”). These argument tags contain plain text descriptions of the referents associated with each argument. They further indicate information about the referent’s activation status for reference tracking purposes. The annotation schema is exemplified in the table below. Both the arguments themselves and the corresponding predicates are annotated.

Codes	**Explanation**
Annotation on the argument	
#father	Referent is first **introduced.** Note that the predicate itself may first introduce the referent, in which case it also receives a #.

[22] In fact, we distinguish between the horizontal and the vertical plane for both, source and goal, in order to keep the value lists clear. For specific starting and ending points in a 3-dimensional signing space 'src_h and 'src_v and/or 'gol_h and 'gol_v were combined (cf. Appendix 1).

@father	A referent that has been (or will be) introduced into the discourse is **anchored/localized** in space.
#@son	Simultaneous **introduction and localization**: The place of articulation of a nominal sign is modified to localize it.
Annotation on the predicate	
arg1_act:father, arg2_gol:son, arg3_pat:book	The general semantic role of the argument is specified (act(or), pat(ient), go(a)l/rec(ipient)).

We also plan to introduce a second set of tiers "Indexed_by_A/B", which allows entering information on the means by which an argument is indexed on the predicate, e.g. via modification of the predicate's initial and/or final location (src/gol), via eye gaze, mouthing, or Constructed Action, or implicitly via the context.

The DGS Corpus project currently lacks the resources to provide this level of detail in argument/referent tracking for the entirety of the corpus data. However, we believe that the annotation scheme may prove useful for future users of the corpus researching e.g. reference tracking, and plan to make sample annotations available as part of the Public DGS Corpus.

6 Repetition

A form variation that is associated with a number of lexical and morphological processes is repetition. At the lexical level, signs may vary between a single movement vs. one or more movement repetitions. Repeated variants tend to occur more often with multisyllabic mouthings in BSL and LSF (Woll 2001; Sallandre 2003; Fenlon et al. 2015), but given exceptions to this trend, there may be other conditioning factors as well as some free variation. In contrast, variation in the number of movements is systematic in word formation processes. When compounds become lexicalized, for instance, movement repetition in each compound member is often lost due to prosodic pressure to create monosyllabic signs (Hohenberger 2008). Movement repetition can also be morphemic e.g. as a derivational morpheme creating nouns from verbs (Supalla & Newport 1978, Cuxac 2000, Johnston 2001, Kimmelmann 2009) or as an inflectional morpheme denoting e.g. habitual or iterative aspect or intensification or duration on verbs (Fernald et al. 2000; Cuxac 2000, Notarrigo & Meurant 2019) or pluralizing nouns (Costello 1995; Steinbach 2012). In light of the range of functions assumed by movement repetition, it is desirable to annotate deviations from the number of movements of the citation form and make these annotations available in the Public DGS Corpus. Furthermore, student-generated annotations for number of movements can be verified by automated pose estimation analysis such as offered by OpenPose, ensuring a reliable quality assurance mechanism for this aspect of detailed transcription.

Deviations in the number of path movements are annotated using the qualifier phases ('phs), whose values indicate the number of performed movements (phases) with a range from "0" (no path movement) to "4" (three repetitions) plus the value "multiple" for an unspecified number of more repetitions. One complete movement phase consists of either a) a single one-directional movement (e.g. no2a (nein2a)), b) a complex back and forth movement (e.g. to-decide1a (entscheiden1a)), or c) an alternating movement executed first with the dominant and then with the non-dominant hand (e.g. BOREDOM1 (LANGEWEILE1)). For each deviating token we indicate how many phases it has in total rather than in how many phases it differs from the citation form. In other words, a token annotated as DISAPPOINTED5'phs:2 (ENTTÄUSCHT5'phs2)

has two complete movements, not three. A token that does not exhibit the path movement of its citation form receives the value “0”, for instance TO-RING-UP1'phs:0 (ANRUFEN1'phs:0).

Plural formation in nouns and distributive verbal aspect often require path movements to be repeated at slightly different locations in signing space. Displacement along the horizontal axis is indicated by a separate qualifier offset_direction ('offdir, e.g. CHILD2'phs:2'offdir: rightwards_sequentially / CHILD2* (KIND2'phs:2'offdir:nach rechts_seq / KIND2*) and TO-VISIT-OR-TO-ATTEND1A 'phs:3'offdir: leftwards_sequentially / TO-VISIT-OR-TO-ATTEND1A* (BESUCHEN1A 'phs:3'offdir:nach links_seq / BESUCHEN1A*)).

7 Number of hands

Signs with a one-handed citation form may sometimes occur with two hands, where the non-dominant hand mirrors the articulation of the dominant hand (Nilsson 2007). At the same time, weak drop can lead to two-handed signs being produced with only one hand (Padden & Perlmutter 1987). The number of hands in a sign may change over time based on a number of different factors. Perceptual considerations can lead to one-handed signs becoming two-handed when they are located near the edge of the visual field e.g. lower on the body Siple (1978). In contrast, two-handed signs near the face or center of the visual field may undergo weak drop. Johnston and Schembri (1999: 157-171) dicuss semantic factors, subordinate incorporation and deletion leading to singling and doubling in Auslan. Lepic et al. (2016) mention semantic factors that may influence handedness, specifically lexically inherent plurality. Looking at ASL, SSL, and ISL, the authors note that two-handed signs often denote (spatial) relations between paired entities or between corresponding parts of a single entity (e.g. the roof and walls of a house). The hands may also encode the wider concept of ‘more than one’ as it occurs in collective or mass entities such as rain or snow. In that case, the double articulators may represent the boundaries of an entity’s shape or volume.

Our annotation scheme distinguishes between symmetrical and asymmetrical two-handed signs. Symmetrical signs share the same phonological parameters but may have opposite movement directions. The qualifier number_of_hands ('hd) is used to mark symmetrical two-handed tokens with a one-handed citation form ('hd:2) as well as one-handed tokens of a symmetrical two-handed sign ('hd:1). For instance, the citation form of REQUEST1A (ANTRAG1A) is two-handed and symmetrical, but we find one-handed tokens and annotate them as REQUEST1A'hd:1 (ANTRAG1A'hd:1). Likewise, a two-handed token of the one-handed sign ALERT1 (ALARM1) would be annotated as follows: ALERT1'hd:2 (ALARM1'hd:2). The values of 'hd further allow specifying whether a two-handed token uses an alternating movement, the two hands have the same vs. opposite movement directions, or the movement onset for each hand is different.

In asymmetrical two-handed signs, the non-dominant hand typically serves as a place of articulation and/or takes on a classifier handshape. Deviations from the citation form are indicated with the help of the qualifier base ('bas), which shows that an asymmetrical two-handed sign exhibits weak drop or that a one-handed sign is produced with a passive (base) hand. The values of 'bas allow specifying which part of the non-dominant hand/arm serves as place of articulation as well as the handshape of the non-dominant hand. For example, the one-handed PRESENT-OR-HERE1 (DA1) may be signed with a C-shaped passive hand: PRESENT-OR-HERE1'bas:c-hand_palm_side (DA1'bas:C-Hand-Seite). An example of weak drop would be TO-READ-BOOK1A (LESEN-BUCH1A) signed without the non-dominant hand representing the entity that is read: TO-READ-BOOK1A'bas:none (BUCH-LESEN1A'bas:keine).

8 References

Costello, Elaine. 1995. Signing: How to speak with your hands. New York: Bantam Books.

Cuxac, Christian. 2000. La Langue des Signes Française (LSF). Les voies de l'iconicité. Paris: Ophrys.

Fenlon, Jordan, Cormier, Kearsy & Schembri, Adam. 2015. Building BSL SignBank: The lemma dilemma revisited. In: *International Journal of Lexicography 28 (2)*, 169–206. DOI: 10.1093/ijl/ecv008.

Fernald, Theodore B. & Napoli, Donna Jo. 2000: Exploitation of morphological possibilities in signed languages: Comparison of American Sign Language with English. In: *Sign Language & Linguistics 3 (1)*, 3–58. DOI: 10.1075/sll.3.1.03fer.

Fradin, Bernhard. 2015. Blending. In Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan, Rainer, Franz (eds.). *Word-Formation: An International Handbook of the Languages of Europe*, vol. 1 (HSK Handbücher zur Sprach- und Kommunikationswissenschaft), 386–412. Berlin: Mouton de Gruyter. DOI: 10.1515/9783110246254-023.

Hanke, Thomas / König, Susanne / Konrad Reiner / Langer, Gabriele. 2012: Towards tagging of multi-sign lexemes and other multi-unit structures. In In Crasborn, Onno et al., Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. Paris: ELRA, European Language Resources Assoc., 67–68. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/12029.html.

Happ, Daniela & Vorköper, Marc-Oliver. 2006. *Deutsche Gebärdensprache: Ein Lehr- und Arbeitsbuch*. Frankfurt a.M.: Fachhochschulverlag.

Hohenberger, Annette. 2008. The word in sign language: Empirical evidence and theoretical controversies. *Linguistics 46 (2)*, 249–308. DOI: 10.1515/LING.2008.011.

Johnston, Trevor. 2001. Nouns and Verbs in Australian Sign Language. In: *Journal of Deaf Studies and Deaf Education 6 (4)*, 235–257. DOI: 10.1093/deafed/6.4.235.

Johnston, Trevor. 2011. Auslan Corpus Annotation Guidelines. [Last updated 30 November 2011. URL: http://www.auslan.org.au/video/upload/attachments/AuslanCorpusAnnotationGuidelines 30November2011.pdf.

Johnston, Trevor. 2019: Auslan Corpus Annotation Guidelines. August 2019 version. [Online resource; URL: https://media.auslan.org.au/django-summernote/2020-09-04/69f180d0-04bf-4831-b59c-0b0b8f07cbb5.pdf.

Johnston, Trevor / Schembri, Adam. 1999. On Defining Lexeme in a Sign Language. In: Sign *Language & Linguistics 2 (2)*, S. 115-185. DOI: 10.1075/sll.2.2.03joh.

Kimmelman, Vadim. 2009. Parts of speech in Russian Sign Language: The role of iconicity and economy. In: *Sign Language & Linguistics 12 (2)*, 161–186. DOI: 10.1075/sll.12.2.03kim.

König, Susanne, Reiner Konrad & Gabriele Langer. 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, 379–404.

Konrad, Reiner, Hanke, Thomas, Langer, Gabriele, König, Susanne, König, Lutz, Nishio, Rie, & Anja Regen. 2020. Public DGS Corpus: Annotation Conventions / Öffentliches DGS-Korpus: Annotationskonventionen (Version 3). DOI: 10.25592/uhhfdm.1860.

Konrad, Reiner, Hanke, Thomas, König, Susanne, Langer, Gabriele, Matthes, Silke, Nishio, Rie & Anja Regen. 2012. From form to function. A database approach to handle lexicon building and spotting token forms in sign languages. In Crasborn, Onno et al., Proceedings of the LREC2012 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. Paris: ELRA, European Language Resources Assoc., 87–94. URL: https://www.sign-lang.uni-hamburg.de/lrec/pub/12023.html.

Konrad, Reiner, Langer, Gabriele, Müller, Anke & Sabrina Wähl. in press. Where frozen signs reclaim iconic ground – iconic modification in German Sign Language (DGS). Proceedings of the 12th International Symposium on Iconicity in Language and Literature. Lunds Universitet, 3–5 May 2019.

Lepic, Ryan, Börstell, Carl, Belsitzman, Gal & Sandler, Wendy. 2016. Taking meaning in hand: Iconic motivations for two-handed signs. *Sign Language & Linguistics* 19(1), 37–81. DOI: 10.1075/sll.19.1.02lep.

Mathur, Gaurav & Rathmann, Christian. 2012. Verb agreement. In Pfau, Roland, Steinbach, Markus & Woll, Bencie (eds.), Sign language: An international handbook. Berlin: Mouton de Gruyter, 136–157. DOI: 10.1515/9783110261325.136.

Nilsson, Anna-Lena. 2007. The non-dominant hand in a Swedish Sign Language discourse. In Vermeerbergen, Myriam, Leeson, Lorraine & Crasborn, Onno (eds.), *Simultaneity in Signed Languages: Form and Function*, 163–185. Amsterdam: John Benjamins. DOI: 10.1075/cilt.281.08nil.

Notarrigo, Ingrid & Meurant, Laurence. 2019. Conversations spontanées en langue des signes de Belgique francophone (LSFB): Fonctions et usages de la répétition. *Lidil* [en ligne] 60, DOI: 10.4000/lidil.7002.

Padden, Carol & Perlmutter, David M.. 1987. American Sign Language and the architecture of phonological theory. *Natural Language & Linguistic Theory* 5. 335–375. DOI: 10.1007/BF00134553.

Pfau, Roland. 2008. The grammar of headshake: A typological perspective on German Sign Language negation. *Linguistics in Amsterdam* 1: 37–74.

Pfau, Roland & Quer, Josep. 2008. On the syntax of negation and modals in Catalan Sign Language and German Sign Language. In Perniss, Pamela, Pfau, Roland & Steinbach, Markus (eds.), *Visible Variation: Comparative Studies on Sign Language Structure*. Berlin, New York: de Gruyter Mouton, 129–162. DOI: 10.1515/9783110198850.129.

Plag, Ingo. 2015. English. In Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, vol. 4 (HSK Handbücher zur Sprach- und Kommunikationswissenschaft), 2411–2426. Berlin: de Gruyter Mouton. DOI: 10.1515/9783110379082-002.

Quer, Josep, Cecchetto, Carlo, Donati, Caterina, Geraci, Carlo, Kelepir, Meltem, Pfau, Roland & Steinbach, Markus, 2016. SignGram Blueprint: A guide to sign language grammar writing. Berlin: de Gruyter Mouton. DOI: 10.1515/9781501511806.

Quer, Josep. 2012: Negation. In Pfau, Roland, Steinbach, Markus & Woll, Bencie (eds.): *Sign Language. An International Handbook.* Berlin, New York: de Gruyter Mouton, 316–340.

Ruppenhofer, Josef, Wiegand, Michael, Wilm, Rebecca & Markert, Katja. 2018. Distinguishing affixoid formations from compounds. Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, August 20-26, 2018, 3853–3865. URL: https://aclanthology.org/C18-1325.pdf.

Sallandre, Marie-Anne. 2003. Les unités du discours en Langue des Signes Française. Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité. 2 vol. Paris: Université Paris 8.

Schulder, Marc & Hanke, Thomas (2020). OpenPose in the Public DGS Corpus. Project Note AP06-2019-01. Version 2. Hamburg, Germany: DGS-Korpus project, IDGS, Hamburg University. DOI: 10.25592/uhhfdm.842.

Schulder, Marc & Hanke, Thomas. 2019. Identifying negation in the DGS corpus. Paper presented at the SignManuals 2 International Workshop in Graz.

Schwager, Waldemar. 2012. Morphologie. Bildung und Modifikation von Gebärden. In Eichmann, Hanna, Hansen, Martje & Heßmann, Jens (eds.), *Handbuch Deutsche Gebärdensprache*. Seedorf: Signum Verlag.

Siple, Patricia. 1978. Visual constraints for sign language communication. *Sign Language Studies 19*. 95–110. DOI: 10.1353/sls.1978.0010.

Steinbach, Markus. 2012. Plurality. In: Pfau, Roland, Steinbach, Markus & Woll, Bencie (eds.): Sign *Language: An International Handbook*. Berlin, New York: de Gruyter Mouton, 112–136. DOI: 10.1515/9783110261325.112.

Supalla, Ted & Newport, Elissa L. 1978. How many seats in a chair? The derivation of nouns and verbs in ASL. In: Siple, Patricia (ed.): *Understanding language through sign language research*. New York, San Francisco, London: Academic Press, 91–132.

Woll, Bencie. 2001. The Sign That Dares to Speak Its Name: Echo Phonology in British Sign Language (BSL). In: Boyes Braem, Penny & Sutton-Spence, Rachel (eds.): *The Hands Are the Head of the Mouth*. Seedorf: Signum, 87–98.

Zeshan, Ulrike. 2006. Negative and interrogative constructions in sign languages: A case study in sign language typology. In: Zeshan, Ulrike (ed.): *Interrogative and negative constructions in sign languages*. Nijmegen: Ishara Press, 28–68.

Zwicky, Arnold & Pullum, George K. 1983. Cliticization vs. inflection: English n't. *Language 59*: 502–513. DOI: 10.2307/413900.

Appendix 1

Qualifiers used to spot token form modification or variation in the DGS corpus, grouped by parameter and function.
Some qualifiers are outdated: The Qualifier *hold* was introduced in the experimentation phase, but abandoned and replaced by indication of form deviation ("h") in the token tag. Likewise, the qualifiers *location_text_structure_horizontal/..._sagittal/..._vertical* were replaced by *location* (without feature values). The qualifier *movement on surface* was replaced by *plane* in combination with *movement_direction*.

Tokens	Qualifier	Code	Feature	Parameter	Others	Function *Origin*
39688	number_of_hands	hd	number_of_hands	no/hands		
6132	base	bas	base	no/hands		
199	base_fixed	bas_fix		no/hands		
620	hold	h	hold	no/hands		
2510	assimilation	assim		hs		phonotact
59	assimilation_location	assim_loc		loc		phonotact
2129	location	loc		loc		
903	location_on_body	bodyloc	location_on_body	loc		
1957	location_text_structure_horizontal	loc_ts_h	location_text_structure_horizontal	loc		
56	location_text_structure_sagittal	loc_ts_s	location_text_structure_sagittal	loc		
179	location_text_structure_vertical	loc_ts_v	location_text_structure_vertical	loc		
5966	movement_direction	movdir	movement_direction	mov		
179	axis	ax	axis	mov		
467	plane	p	plane	mov		
417	movement on surface	mos	movement on surface	mov		
9	halt	halt	kind of movement	mov		
41	tense	tense	kind of movement	mov		
419	continued	cont		mov		
495	inclusive	1incl		mov		
156	reverse	rev		mov		
699	size	size	size & speed	mov		
96	speed	speed	size & speed	mov		
15	timeline	t		mov		morph/sem
1711	alpha_negation	alph		*[mov]*	affix	morph/sem
4745	goal_h	gol_h	source+goal	mov		morph/synt
2302	source_h	src_h	source+goal	mov		morph/synt
120	goal_v	gol_v	location_text_structure_vertical	mov		morph/synt
54	source_v	src_v	location_text_structure_vertical	mov		morph/synt
52717	phases	phs	phases	mov		
1810	offset_direction	offdir	offset_direction	mov		
207	sketching movement	sk_mov	sketching movement	mov		
68	distalization	dist	distalization	*[mov]*	var_paradigm	
49	proximalization	prox	distalization	*[mov]*	var_paradigm	
161	extended_finger_direction	ext_fing_dir	extended finger direction	ori		
158	palm_orientation	palm_orient	palm orientation	ori		
796	head_shaking	h_s		*[nonman]*	nonmanual	morph/sem
4	fa_bsl	2_bsl	alphabet			alphabet
4655	fa_one-handed	1	alphabet			alphabet
88	fa_tracing	sk	alphabet			alphabet
128	fa_two-handed	2	alphabet			alphabet
46	fa_ligature	lig		*[mov]*		alphabet
435	cued speech	cs	cued speech			*articulation*
2815	number	n	number			number
393	detour	numinc		*[mov]*		number
2806	m_out_of_n	of	number_handshape			number
21550	quantity	q	number_handshape			number
109	quantity1	q1	number_handshape			number
105	quantity2	q2	number_handshape			number
2	num_tracing	sk_n	number_written form			number

Appendix 2

Qualifiers used to spot token form modification or variation in the DGS corpus,, sorted by number of tokens:

Tokens	Qualifier	Code	Feature	Parameter	Others	Function *Origin*
52717	phases	phs	phases	mov		
39688	number_of_hands	hd	number_of_hands	no/hands		
21550	quantity	q	number_handshape			number
6132	base	bas	base	no/hands		
5966	movement_direction	movdir	movement_direction	mov		
4745	goal_h	gol_h	source+goal	mov		morph/synt
4655	fa_one-handed	1	alphabet			alphabet
2815	number	n	number			number
2806	m_out_of_n	of	number_handshape			number
2510	assimilation	assim		hs		phonotact
2302	source_h	src_h	source+goal	mov		morph/synt
2129	location	loc		loc		
1957	location_text_structure_horizontal	loc_ts_h	location_text_structure_horizontal	loc		
1810	offset_direction	offdir	offset_direction	mov		
1711	alpha_negation	alph		*[mov]*	affix	morph/sem
903	location_on_body	bodyloc	location_on_body	loc		
796	head_shaking	h_s		*[nonman]*	nonmanual	morph/sem
699	size	size	size & speed	mov		
620	hold	h	hold	no/hands		
495	inclusive	1incl		mov		
467	plane	p	plane	mov		
435	cued speech	cs	cued speech			*articulation*
419	continued	cont		mov		
417	movement on surface	mos	movement on surface	mov		
393	detour	numinc		*[mov]*		number
207	sketching movement	sk_mov	sketching movement	mov		
199	base_fixed	bas_fix		no/hands		
179	axis	ax	axis	mov		
179	location_text_structure_vertical	loc_ts_v	location_text_structure_vertical	loc		
161	extended_finger_direction	ext_fing_dir	extended finger direction	ori		
158	palm_orientation	palm_orient	palm orientation	ori		
156	reverse	rev		mov		
128	fa_two-handed	2	alphabet			alphabet
120	goal_v	gol_v	location_text_structure_vertical	mov		morph/synt
109	quantity1	q1	number_handshape			number
105	quantity2	q2	number_handshape			number
96	speed	speed	size & speed	mov		
88	fa_tracing	sk	alphabet			alphabet
68	distalization	dist	distalization	*[mov]*	var_paradigm	
59	assimilation_location	assim_loc		loc		phonotact
56	location_text_structure_sagittal	loc_ts_s	location_text_structure_sagittal	loc		
54	source_v	src_v	location_text_structure_vertical	mov		morph/synt
49	proximalization	prox	distalization	*[mov]*	var_paradigm	
46	fa_ligature	lig		*[mov]*		alphabet
41	tense	tense	kind of movement	mov		
15	timeline	t		mov		morph/sem
9	halt	halt	kind of movement	mov		
4	fa_bsl	2_bsl	alphabet			alphabet
2	num_tracing	sk_n	number_written form			number