

Lemmatization Rules

(for the pre-release dictionary entries)

Contributors: Gabriele Langer, Anke Müller, Sabrina Wähl, Susanne König,
Thomas Hanke, Reiner Konrad

DOI (latest version in each case):

10.25592/uhhfdm.846

Versions:

- Version 1 (2016): Original version (German)
- Version 2 (2017-04-04): Revision
- Version 3 (2020-02-26): Corrections and minor additions
- Version 4 (2020-10-30): Layout adjustments; minor stylistic changes in the introduction; German and English abstract added
- Version 5 (2024-04-17): Substitution of an example, adaptation to the current glosses and linking of the glosses to the Public DGS Corpus, minor stylistic corrections; first English edition

Abstract

This project note is part of the lexicographic manual for the dictionary (DW-DGS). It covers the process of lemma establishment for the compilation of dictionary entries, which are being made available online to the interested public as pre-release entries since 2020.

We describe the linguistic and pragmatic criteria used for deciding what comprises a sign in the lexicographic sense and therefore is to be described in one dictionary entry. The criteria are stated in form of lemmatization rules. The manual component of the sign is the starting point of our considerations, consequently one entry may describe a sign that combines with different frequent mouthings.

A sign is often realized with different (word, i.e., sign) forms, the sum of which we call its range of forms. The lexicographic description aims to give a summarizing description of each lemma sign that is based on the observed form realizations and their contextual meanings. Different lemma signs are to be described in separate entries. The lemma establishment decisions are based on the data as they are represented at the time of entry preparation in iLex, the annotational database of the DGS-Korpus project. In practice, this involves deciding which types and subtypes should be considered realizations of a given lemma sign and therefore described in one entry, and which should be excluded and described in a different entry.

Zusammenfassung (German Abstract)

Dieses Arbeitspapier ist Teil des lexikographischen Handbuchs für das Wörterbuch. Es deckt den Arbeitsschritt der Lemmakonstitution für die Erstellung der Wörterbucheinträge ab, die von 2020 an der Öffentlichkeit als Vorabeinträge zugänglich gemacht werden.

In diesem Arbeitspapier werden die sprachsystematischen, aber auch praxisorientierten, pragmatischen Kriterien festgelegt und begründet, nach denen entschieden werden kann, was eine Gebärde im lexikographischen Sinne konstituiert und damit in einem Eintrag beschrieben werden kann. Dies wird in Form von Lemmatisierungsregeln festgehalten. Ausgangspunkt unserer Überlegungen ist dabei der manuelle Anteil einer Gebärde, d.h. eine Gebärde in einem Eintrag kann verschiedene übliche Mundbilder haben.

Eine Gebärde wird in der konkreten Verwendung oft in unterschiedlichen (Wort-)Formen realisiert, die wir in der Summe als das Formenspektrum der Gebärde bezeichnen. Die lexikographische Beschreibung zielt darauf ab, das jeweilige Lemmazeichen unter Berücksichtigung seiner beobachteten Realisierungsformen und Bedeutungen in einem Eintrag möglichst angemessen zusammenfassend zu beschreiben. Verschiedene Lemmazeichen sollen in getrennten Einträgen beschrieben werden. Die Lemmakonstitution erfolgt auf Grundlage der Datenlage, wie sie sich in iLex, der Annotationsdatenbank des DGS-Korpus-Projekts, zur Zeit der Bearbeitung darstellt. Das bedeutet, dass bei der Lemmakonstitution zu entscheiden ist, welche erfassten Types und Subtypes als Realisierungsformen eines Lemmazeichens aufzufassen und gemeinsam zu beschreiben sind und welche nicht.

1. Introduction	4
2. Lemma sign establishment	5
2.1. Range of forms of a lemma sign	6
2.1.1. Word forms	7
2.1.1.1. Grammatical modifications	7
2.1.1.2. Iconic modifications (re-iconizations)	10
2.1.2. Lexicalized modifications	11
2.1.3. Realization variants	12
2.1.3.1. Determination of realization variants	14
2.1.4. Related signs	17
2.1.5. Considerations on the scope of an entry	18
2.1.5.1. Mouthing	18
2.1.5.2. Different range of modifications for different senses	19
2.1.5.2.1. Re-iconizations	19
2.1.5.2.2. Categorical indeterminacy with respect to part of speech	19
2.2. Range of meanings of a lemma sign (polysemy)	21
2.2.1. Connection via the underlying image of the sign	21
2.2.2. Semantic connection	22
2.2.3. Connection via the mouthing	22
2.2.4. Homonymy versus polysemy	23
2.3. Basic principle of lemma establishment	24
2.4. Simple cases and examples	25
2.4.1. Analysis of polysemy	26
2.5. Summary of entries (at the type level): Complex cases	28
2.5.1. Realization variants at the level of the lemma sign	29
2.5.1.1. Restricted variant sets and restricted range of meanings	30
2.5.1.2. Variant chains or networks	33
2.5.1.3. Different variant sets	33
2.5.1.4. Major differences in the range of meanings (small intersecting set)	34
2.5.1.5. Different core meanings	35
2.5.2. Different basic forms	36
3. Choice of lemma	37
3.1. Choice of the standard form	38
3.2. Choice of the basic form	38
3.3. Choice of the main variant	39
4. References	40
4.1. Linguistics and lexicography	40
4.2. Language resources and dictionaries	42
5. Definitions of terms	42

1. Introduction

These project notes are part of the lexicographic handbook for the dictionary (DW-DGS). They cover the process of lemma establishment for the compilation of dictionary entries. For the preceding step of lemma selection, see project notes *AP10-2016-02: Procedure of Analysis for (dictionary) entry writing*.

In these project notes, we do not discuss independent linguistic units below the lexeme level (e.g., how to deal with productive handshapes¹ or non-manual elements²), nor do we cover multiword expressions³ (MWE), which currently enter the dictionary as (loan) compound words as run-ons in the foot of an entry.

Ahead of the final publication of the dictionary in 2023, pre-release dictionary entries are being published from 2018 on, first with access for the focus group and since 2020 for the general public. These entries are intended to make already produced content accessible and to promote the discussion about the emerging dictionary (*Wörterbuch*, hereafter WB) with both the language and the research community.

The core of an entry/article is information on senses and variants.

Pre-release entries may still differ in form and content from the final version in the following respects:

- They may not yet contain all the units of information that will be included in the final product. For example, at present we cannot provide definitive information on “word class” due to the current state of research. This situation may have changed by 2023.
- The senses and other information are compiled based on the data (corpus, DGS-Feedback and SignHunter data⁴), and insights from follow-up elicitations will also be included if applicable. Since more data will become available before the end of the project, the dictionary entries may be supplemented or corrected at a later date with variants, senses, and other information.
- Some entry types may not yet exist, e.g., entries for multi-word expressions, for classifier handshapes, etc.
- The final layout will be designed at a later stage.
- Only basic access options will be provided, comprehensive access structures and explanatory texts will only be available in the final product.

This handbook provides the basis for the compilation of the pre-release dictionary entries and will be supplemented by further structures as the project continues.

¹ Compare Brien (1992), Johnston (1989), Centre for Sign Language (2008-2017).

² Compare Johnston/Schembri (1999, 118).

³ The term *multi-sign expression*, which would be more appropriate for sign languages, has not been established in sign language research or lexicography.

⁴ For more information concerning the DGS Feedback System and the further use of the DGS Feedback data see Wahl et al. (2018); for more information concerning SignHunter, a tool for collecting isolated signs, see Hanke et al. (2020).

2. Lemma sign establishment

Following Svensén,⁵ we speak of the *establishment of lemma signs* or, abbreviated, of *lemma establishment*. The term refers to the practice of *lemmatization* in a lexicographic sense (cf. Herbst/Klotz 2003, 178f.), which is related to, but still distinct from lemmatization in annotation, which refers to token-type matching.

Different dictionary projects weigh criteria such as etymology, meaning, and form differently to determine candidates for lemma signs, which form should serve as the lemma, and what information a corresponding entry includes. Since DGS lacks a widely used writing system and therefore written evidence from earlier language stages, no reliable etymological information is available. Instead, an iconic interpretation of the form of many signs allows for drawing conclusions about the original motivation of the sign form and possible relations to other signs. Even though the interpretation of iconicity is always to some extent subjective and thus prone to errors, we use the underlying image and image producing technique⁶ as criteria for lemma establishment (whenever they can be reliably identified), alongside form and meaning.

One sign can have several typical⁷ mouthings. When discussing whether two forms are the same or different, we always refer to the manual sign.

A sign is often realized with different (word, i.e., sign) forms, the sum of which we call its **range of forms**. The lexicographic description aims to give a summarizing description of each lemma sign that is based on the observed form realizations and their contextual meanings. Different lemma signs are to be described in separate entries.

The lemma establishment decisions are based on the data as they are represented at the time of entry preparation in iLex, the annotational database of the DGS-Korpus project. In practice, this involves deciding which types and subtypes should be considered realizations of a given lemma sign and therefore described in one entry, and which ones should be excluded and described in a different entry.

The type structure in iLex is hierarchical. Each type entry is assigned to a level (types.level). “Signs” constitute the top level (level=3) and represent lemma sign candidates. Types of the levels 2, 1, and 0 are dependent on the level 3 signs and attached to them through parent-child relationships. These dependent types are called subtypes in the following text. Level 1 subtypes (types.level=1) are called

⁵ Svensén calls this lexicographic step “... ESTABLISHMENT OF LEMMAS. This operation is not simply the same thing as lemmatization Naturally, the establishment of lemmas assumes previous lemmatization, but it also includes deciding how lexical items having identical basic forms are to be presented in the dictionary, and to what extent word elements and multi-word lexical items are to be accorded lemma status.” (Svensén 2009, 94).

⁶ For a detailed discussion of image-producing techniques, see. Langer (2005), Langer/König/Konrad (2007) or König et al. (2008).

⁷ Based on the distinction between regular and occasional sign-mouthing-combinations in Ebbinghaus/Heßmann (2001, 137f.).

“lexemes” in iLex. Lexeme types are used to group together tokens with conventionalized form-meaning relationships (usually indicated by typical sign-mouthing combinations) within a sign type.⁸ In the context of token-type matching, tokens are matched to different subtypes or to the type itself. Each type level fulfills a different grouping function: Level 1 subtypes (lexemes) group tokens according to their lexicalized uses (meanings), while Level 0 subtypes capture specific realization forms of these lexicalized uses, they group word forms, for example. There are also realization forms of types that do not correspond to any lexicalized use. These are modelled as level 2 subtypes. Each token matched to a subtype is considered evidence of the existence of this subtype and of all its superordinate subtypes or types.

During lemma sign establishment, the grouping of observed forms and uses (incl. meanings) in a synopsis of the available types and their type environments⁹ can lead to a reordering or a different allocation of (sub)types to dictionary entries than the type hierarchy structures in iLex suggest. This is because lemma establishment takes into consideration additional criteria, in particular semantic ones (cf. Langer et al. 2016). Here, language-internal criteria play a role as well as pragmatic/user-oriented considerations such as the size and complexity of the entries.

The meaning or the range of meanings of a sign (types.level=3) is the sum of the meanings of its associated lexemes (types.level=1). The range of forms results from the type’s citation form (level 3) with its range of forms as well as the range of forms of all its dependent types (level 2, 1, or 0). Lemma sign establishment always aims at an abstract lemma sign, whose scope, description, and distinction from other types must be abstracted from the forms and meanings of its concrete realizations. This will be illustrated and discussed in section 2.5.

The following rationales and criteria are developed starting with considering the corpus tokens of a subtype, i.e. a realization form with a conventional meaning. In iLex, realization forms are either represented as lexemes (types.level=1), or as so-called qualified types (types.level=0). Section 2.4 (*Simple use cases and examples*) covers lemma sign establishment at the lexeme level. The first step is to determine which word forms belong to a lexeme with a given conventional meaning and which additional realization variants (roughly phonological variants, but see Section 2.1.3) are evidenced for the lexeme under investigation.

2.1. Range of forms of a lemma sign

For most signs, no fixed and closed inflectional paradigms of grammatical word forms can be identified. The observed and possible forms can often be highly variable. The similarity of two forms (types or subtypes) – especially if they are iconically related – may, but does not have to, indicate that these are realization forms of the same lemma sign. The goal of lemma

⁸ The use of the word “lexeme” here is project-internal and should not be understood as a lexicographic or lexicological term, which can be read in different ways depending on the theoretical view taken. Compare Schläefer (2002, 16; our translation): “The understanding of the lexeme as a system word with an expressive and a content component is specified lexicologically in different ways. While on the one hand a given meaning is regarded as sign-forming and thus each lexeme has only one meaning, other views assume that lexemes can have several meanings.” The designation lexeme was chosen here more in line with the first view, but in iLex it merely represents a kind of pre-grouping (cf. also the use of “lexeme” in Svensén (2009, e.g., 95f.)). In contrast, Atkins/Rundell (2008, 162) speak of “lexical units” and “headwords” in this context: “A headword in one of its senses is a lexical unit (or LU).”

⁹ By type environment we mean those other types and subtypes in iLex that are, or potentially could be, related in some way to the form type being analyzed. This includes other word forms of the type, modifications, identified realization variants, synonymous signs, signs that have the same or a similar form, and iconically related signs.

sign establishment is to determine which forms are realizations of the same sign and to distinguish them from forms that belong to other signs.

The range of forms of a sign includes the basic form selected as the citation form (see Section 3: Choice of lemma) as well as all other forms in which the sign is used while maintaining its meaning(s), or sense(s):

- systematically formed modifications that fulfill primarily grammatical functions, i.e., grammatical word forms;
- modifications that are attested for iconic signs and motivated by iconic adaptations to the visual-scenic structure of the signed utterance, e.g., re-iconizations;
- realization variants, i.e., small, conventionalized form differences that have no influence on the meaning or function of the sign in the concrete utterance.

The following do not belong to the range of forms of a sign:

- lexicalized modifications, i.e., signs that presumably originated from a modification or word form of a sign, but that in the course of language change have acquired their own, wider or more specific meaning, which no longer corresponds to the meaning of the modification and thus cannot be inferred from the base sign via systematic rules;
- iconically related signs, i.e., forms with a similar underlying image or shared image components, but different meanings.

The categories listed here are described and explained below.

2.1.1. Word forms

Word forms are either grammatical modifications or iconic modifications (re-iconizations) that can be understood through a rule-governed interpretation of their adapted forms. Both are functional modifications. They contrast with such modifications that are accompanied by a change in lexical meaning and that consequently cannot be considered to be mere word forms anymore (more on this below). The form similarities among word forms are associated with grammatical functions and contrast with form similarities that are based on phonological variation (realization variants) or with form similarities that are based on signs sharing an identical or similar underlying image or image component, but that are not related semantically.

2.1.1.1. Grammatical modifications¹⁰

Grammatical modifications do not change the meaning of a sign but they have a grammatical-textual function. They are largely systematic modifications that depend partly on the type of sign, and partly on its use in a particular context.

¹⁰ Fenlon/Cormier/Schembri (2015) take somewhat different decisions regarding word forms (their term: sign modification) vs. sign formation. For example, they do not regard number incorporation (p. 187) and specific negation forms (p. 188) as word forms but as instances of sign formation, which leads to independent dictionary entries for each.

The following list of word forms does not claim to be exhaustive and will be refined, supplemented, and expanded as part of the ongoing work process, depending on the state of processing, insights gained from transcription and detail analysis, and the current state of research.¹¹

Examples of different types of grammatical word forms relevant for the dictionary:

- Plural forms (repeated, repeated and offset): the lexemes [PERSON1](#) and [GROUP1A](#) show plural forms with repetition and movement offset; PICTURE1'phs:multiple is an example of plural with repetition only.¹²

Gloss	Form (in HamNoSys)	Meaning(s)
PERSON1	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ (basic form)	'person'
PERSON1'phs:2'offdir:rightwards_sequentially	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋	'persons'

GROUP1A	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ (basic form)	'group'
GROUP1A'phs:multiple'offdir:rightwards_sequentially	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋	'groups'
PICTURE4	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ (basic form)	'picture'
PICTURE4'phs:multiple	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋	'pictures'

- Aspect forms (different types of movement repetition or changes to movement speed¹³): iterative or durative aspect; aspect forms occur in verbal uses of a sign.

TO-WAIT1A	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ (basic form)	'to wait'
TO-WAIT1A'speed:-	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋	'to wait a long time'
BARRIER1	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ (basic form)	'barrier'
BARRIER1'phs:multiple'offdir:rightwards_sequentially	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋	'to encounter barriers over and over again'

- Intensification forms (distinctive movement shape with increased speed and/or size, or fast repetitions with smaller movements): Reinforcement, intensification of an action or quality; occurs both with signs denoting characteristics and signs used verbally. An example of a verbal use is [TO-WORK2](#) with small, fast repetitions, [BAD3B](#) is an example of an intensified sign denoting a characteristic which exhibits a distinctive movement shape.

TO-WORK2	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ (basic form)	'to work'
TO-WORK2'speed:+'phs:multiple	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋	'to work a lot or work hard'
BAD3B	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ (basic form)	'bad'
BAD3B'speed:+	⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋ ⌈ ⌋	'especially bad, the worst'

- Forms of so-called directional verbs: The form of the sign can be altered to indicate its arguments, which have been set up at different locations in the signing space. The

¹¹ Compare Schwager (2012) for DGS; for sign languages in general Pfau/Steinbach/Woll (2012), Steinbach (2012), Meir (2012), Gaurav/Rathmann (2012).

¹² In iLex, repetitions of movement are encoded with the qualifier *phases* (phs), and repetitions at different locations in space are additionally encoded with the qualifier *offset direction* (offdir).

¹³ In iLex, changes in the speed of movement are encoded with the qualifier *speed*.

arguments can take on a number of different semantic roles such as source and goal.¹⁴ The start- and/or endpoint of a sign's movement (or its location) establish(es) a deictic relation to these arguments, for instance to the conversation participants (signer, addressee). Additionally, plural modifications¹⁵ (arched, zigzag, alternating repeated) and aspect modifications can occur. The directional signs [QUESTION1](#) and [TO-VISIT-OR-TO-ATTEND1B](#) can take on different forms depending on which arguments they take and how these arguments are set up in space.

QUESTION1	$\partial_{r0} \cup X^{\pm}$ (basic form)	'to ask'
QUESTION1'src_h:middle'gol_h:signer	∂_{r0}^{\pm}	'to ask', 2. pers. → 1. pers.
QUESTION1'gol_h:mult-swipe	$\partial_{r0} \cup X^{\pm} \rightarrow \rightarrow \rightarrow$	'to ask', person A → persons B
TO-VISIT-OR-TO-ATTEND1B	$\square_{1r0} [\pm \rightarrow \rightarrow]$ (basic form)	'to visit'
TO-VISIT-OR-TO-ATTEND1B'src_h:middle'gol_h:signer	$\square_{r0} [\pm \rightarrow \rightarrow] \equiv X$	'to visit', 2.Pers → 1.Pers.
TO-VISIT-OR-TO-ATTEND1B'src_h:right'gol_h:left	$\square_{r0} \equiv [\leftarrow \rightarrow]$	'to visit', person A → person/location B
TO-VISIT-OR-TO-ATTEND1B'src_h:Signer'gol_h:mult-zigzag	$\square_{1r0} [[\pm \rightarrow \rightarrow] \rightarrow]$	'to visit', person A → persons/locations B

- For a specific group of signs with a variable place of articulation on the body, the place of articulation constitutes morphological information. In other words, these signs reference the body parts in a morphologically significant way.¹⁶ [TO-CUT2D-\\$\\$AM](#) (which, among others, includes the lexeme [OPERATION1A](#)) is such a sign: In its 'cut' and 'surgery' senses, the body location where the sign is produced indicates which part of the body is subject to surgery or where a cut is applied. This specification in meaning, which is expressed by the form adaptation, is systematic and predictable. It is thus functional in the sense that signs with a morphological body reference are underspecified for place of articulation and the interpretation of their place of articulation is rule-governed.

OPERATION1A	$[\square_{r0} \rightarrow \rightarrow] [\rightarrow \rightarrow] X^{\pm}$ (basic form)	'operation'
OPERATION1A'hd:2' bodyloc:cheek_as_eye	$\equiv [\rightarrow \rightarrow] \rightarrow X^{\pm}$	'operation on the eye'

- Alpha negation: a form of negation that is encoded by a change in the sign's movement (in the form of the Greek letter α). It occurs only in a small group of signs with predominantly predicative uses.

To illustrate, the subtypes [CAN1](#) and CAN1'alph¹⁷ have the same basic lexical meaning 'to be able to', 'to be capable of'. The movement in CAN1'alph traces the shape of the Greek letter α and fulfills the grammatical function of negation. [CAN1](#) and CAN1'alph are thus word forms belonging to the same lemma sign.

CAN1	$\square_{1r0} [\downarrow \rightarrow]$ (basic form)	'can', as in 'to be able to', 'to be capable of'
CAN1'alph	$\square_{1r0} [[\downarrow \rightarrow] \rightarrow]$	'cannot', as in 'not to be able to', 'to be incapable of'

¹⁴ In iLex, directional signs are encoded with the qualifiers *goal* (gol) and *source* (src).

¹⁵ This type of plural marking presumably arose from deictic pointing to several (imagined) referents. It differs from the plural marker *repeatedly/repeatedly offset*, which is more likely to have arisen from a model presentation (of imagined referents) in the signing space.

¹⁶ In iLex, morphological body reference is coded with the qualifier *location on the body* (bodyloc).

¹⁷ In iLex, this morphological negation is coded with the qualifier *alpha negation* (alph).

- Number incorporation: Some signs allow the morphological incorporation of the basic numbers 1-10. In these signs' forms the indication of quantity merges with the indication of what is counted. To incorporate a number into a sign, its handshape is changed to a conventional number handshape.¹⁸

MONTH1	𐄂𐄃𐄄 [↓→△] (basic form)	'month'
MONTH1'q:5	𐄂𐄃𐄄 [↓→△]	'5 months'
WEEK1A	𐄂𐄃𐄄 [→→1→1] (basic form)	'week'
WEEK1A'numinc'q:3d	𐄂𐄃𐄄 [→→1→1] [→→1→1]	'3 weeks'
OLD8B	𐄂𐄃𐄄 [X 1 𐄂] ↓ 𐄃 X (basic form)	'age'
OLD1B'q:3d	𐄂𐄃𐄄 [X 1 𐄂] ↓ X	'3 years old'

- There may be a morphological comparative: A kind of repetition that occurs with some but not all property and evaluative signs. Whether a sign takes a morphological comparative is thus lexically determined and not predictable by grammatical or phonological rules. Not all adjectivally used signs which potentially allow such a change of movement can be morphologically inflected for comparison in this way.

BAD3B	𐄂𐄃𐄄 (basic form)	'bad'
BAD3B'size:-'phs:multiple'offdir: upwards_continuously	𐄂𐄃𐄄 [𐄂+𐄃]	'worse'
WELL-KNOWN1A	𐄂 [𐄃 0 𐄄] [𐄅 𐄆 𐄇 𐄈] [𐄉 𐄊 𐄋 𐄌] [𐄍 𐄎 𐄏 𐄐] (basic form)	'well-known'
WELL-KNOWN1A'phs:2	𐄂 [𐄃 0 𐄄] [𐄅 𐄆 𐄇 𐄈] [𐄉 𐄊 𐄋 𐄌] [𐄍 𐄎 𐄏 𐄐]	'more well-known'

Certain phonological structures of the manual sign are prerequisite for the use of many of the grammatical modifications listed here. However, a given phonological structure does not predict the existence of a particular grammatical modification with certainty. For instance, number incorporation cannot be applied to all signs denoting a conceivable measurement; nor do alpha negation or forms of comparison apply across the board. Other modifications constitute groups of signs (e.g., directional verbs, signs with morphological body reference). Generally speaking, modification behavior is specific to each particular lemma sign and such information should be covered in a dictionary entry.¹⁹

2.1.1.2. Iconic modifications (re-iconizations)

In addition to grammatical modifications, word forms also include iconic modifications – i.e. modifications that produce a meaning specification that is inferable from the iconically motivated form change. In other words, the interpretation of these word forms can be predicted to a large extent from the meaning of the basic form plus the iconic modification.

In particular, signs that contain a classifier handshape which functions as a manipulator, a substitutor, or a drawing tool (in tracing signs), can often be iconically modified – provided that they exhibit the necessary iconic transparency. These iconic realizations are often not

¹⁸ In iLex, the incorporation of numbers is encoded with the qualifier *number* (q for quantity). An additional variation of the movement is encoded with the qualifier *swerve* (numinc).

¹⁹ In the pre-release entries, only limited information can be provided on lemma-specific modifications. Which modification behavior will be presented in the final dictionary entries and in which level of detail has yet to be determined.

conventionally stabilized and therefore do not form independent units (lemma signs) to be described in the dictionary.²⁰

In many signs with conventional meanings/senses, the iconicity of the underlying image can be reactivated and exploited for iconically motivated modifications. These modifications do not change the lexical meaning of the sign but rather specify, substantiate, or adapt it according to the invoked imagery. For example, tracing signs show adaptations of size and shape, signs with manipulator handshapes can adjust the movement aspect of the action (slow, fast, hesitant, large, small) – especially when used in constructed action (CA) – and signs with substitutive handshapes exhibit temporal and spatial adaptations (speed, distance, location and position in space, directions). Since the iconicity of the sign is exploited in productive ways, the process is referred to as re-iconization. The outcome of an iconic modification is a re-iconized form.

Example:

- The lexeme [TO-STIR5](#) has among its meanings ‘to mix’ and ‘to stir’. It is a strongly iconic sign with a manipulator handshape. The underlying image is of a hand holding a spoon and stirring something with it. The sign can be re-iconized as someone holding a large cooking spoon with both hands and strenuously stirring a viscous mass in a large pot.²¹

TO-STIR5	○ _{±0} ② ⁺ (basic form)	‘to stir’
TO-STIR’ <i>re-ic</i>	: ○ _{±0} ① ③ ⁺	‘to stir a large mass’ (re-iconized)

A conventional meaning of an iconic sign may be more or less directly related to the underlying image of the sign, or its relation to the underlying image can be mediated (e.g., figuratively, metaphorically, or metonymically, see below). Several established meanings of the same sign usually differ in how directly they relate to the underlying image. Often there is one meaning/sense that the underlying image depicts directly.²² Normally, it is only this sense that has re-iconized forms in its range of forms, as it is the only sense whose meaning can be specified directly with the help of the underlying image. The sign [TO-STIR5-SSAM](#), for instance, carries the iconically motivated senses ‘to mix’ and ‘to stir’, but also the mediated senses ‘to cook’, ‘(a) cook’ and ‘kitchen’. However, it is only the sense ‘to stir’ that can be re-iconized.

2.1.2. Lexicalized modifications

We do not classify so-called lexicalized modifications as word forms.²³ A lexicalized modification is a sign that is formally equivalent to a word form of another sign, i.e., a grammatical or iconic modification of another sign, but it has been conventionalized for a

²⁰ It is conceivable that in the dictionary (final product) there will be separate entries for semantic classifier handshapes as productive elements of sign formation. Such entries are not yet planned for the pre-release entries.

²¹ The qualifier *re-iconization* (re-ic) does not yet exist in iLex, but lexicographic experience suggests that it makes sense to annotate such modifications. An exact transcription of such forms in HamNoSys is not planned.

²² *Depict* in the sense of the function pair *depicting* (*Anschauung*) vs. *naming* (*Benennung*) (see Ebbinghaus/Heßmann 2000, 62), also known as *showing* and *telling*.

²³ We are not alone with this decision, for example Johnston/Schembri write: “The signs on the left and right (OPERATE and CAESARIAN) are treated as distinct lexemes in Auslan according to the principle of ‘specificity of meaning’.” (1999, 146); “As a result, it is identified as a separate lexeme: APPENDECTOMY.” (Fenlon et al. 2015, 191). See further the entries SKOV (forest) and TRÆ (tree) in the Danish Sign Language Dictionary (Centre for Sign Language, 2008-2017) and their unpublished editorial rules (Troelsgård/Kristoffersen 2016, 3).

more specific meaning. It could be assumed that the lexicalized modification arose from the word form of the original sign. The more specific meaning is not predictable from the original sign.²⁴ The lexicalized modification is characterized by a smaller range of forms compared to the original sign; specifically, it does not exhibit the form corresponding to the basic form of the original sign. We therefore treat the lexicalized modification as an independent sign with its own basic form.

Example 1: [SURVEY1](#) vs. [QUESTION1](#)

- [SURVEY1](#) resembles the word form QUESTION1'gol h:mult-arc, i.e. an arched plural. However, unlike [QUESTION1](#), [SURVEY1](#) has no other word forms which indicate the source or goal argument of the sign. In addition, the meaning is close to 'to ask many people', but the sign more specifically denotes 'survey' and 'to carry out a survey', where a survey by its aim and methodology is more than simply an instance of asking many people.

QUESTION1	$\partial_{r0} \cup X^{\pm}$ (basic form)	'to ask', 'question'
SURVEY1	$\partial_{r0} \cup X^{[\text{Z}^c \gg \text{r}]}$ (basic form)	'survey', 'to conduct a survey'

[SURVEY1](#) is therefore a lexicalized modification with a basic form that differs from the basic form of the original sign.

Example 2:

- The lexemes [OPERATION1A](#) and [TO-CUT2D](#) represent two conventional meanings of the same lemma sign. They have the same form and both have a morphological body reference, i.e., they can be executed at different body locations (qualifier *location on the body* (bodyloc)).²⁵

OPERATION1A	$[\text{O} \text{L} \theta \times \text{O} \text{I} \text{e}] \sim X^{\text{Z} \setminus \rightarrow}$ (basic form)	'to operate', 'operation'
TO-CUT2D	$[\text{O} \text{L} \theta \times \text{O} \text{I} \text{e}] \sim X^{\text{Z} \setminus \rightarrow}$ (basic form)	'to cut', 'cut'
C-SECTION1	$\text{O} \text{L} \theta \times \text{I} \setminus \cdot \text{I} X^{\rightarrow}$ (basic form)	'C-section'
TO-VACCINATE6	$\text{O} \times \theta \setminus X^{\downarrow}$ (basic form)	'vaccination', 'to vaccinate'

- The modified form OPERATION1'bodyloc:belly_(area) has further become lexicalized to denote a specific type of abdominal operation, namely the C-section. The lexeme [C-SECTION1](#) $\text{O} \text{L} \theta \times \text{I} \setminus \cdot \text{I} X^{\rightarrow}$ is performed in exactly the same way as OPERATION1'bodyloc:belly_(area), but it cannot vary in terms of place of articulation on the body. The lexicalized meaning goes beyond the predictable meaning of the modification of OPERATION1 - a C-section is not just any surgery on the abdomen. Analogously, the form CUT2D'bodyloc:upper_arm is formally identical to TO-VACCINATE6. Like [C-SECTION1](#), TO-VACCINATE6 is not underspecified for location and its meaning 'to vaccinate' is more specific/different from 'cut on the upper arm'.

2.1.3. Realization variants

Realization variants are part of the range of forms of a sign. They are characterized by small differences in the execution of the manual sign, which do not lead to a change in meaning and can therefore be used interchangeably in typical (semantic) contexts. The form difference here has no grammatical or iconic modifying function. Form differences that can be explained by

²⁴ See Johnston/Schembri (1999, esp. 126).

²⁵ Both are lexemes that are grouped together in the type [TO-CUT2D-SSAM](#), which has further senses.

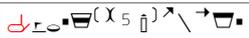
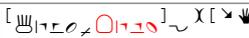
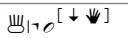
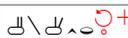
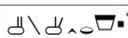
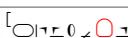
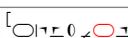
phonological processes, or by prosodic or text-structural phenomena in a given utterance are also not regarded as realization variants.²⁶

Realization variants can only be determined with respect to a specific conventional meaning (i.e., at the lexeme level), since realization variants have to have the same meaning by definition. The lexeme structure in iLex provides a sufficiently differentiated reference level for determining realization variants. When such form variants occur robustly, i.e., have become conventionalized, they are categorized as separate form types in iLex and listed as realization variants in the sign's entry.²⁷ All conventionalized forms of one sign are considered (realization) variants of each other.²⁸

For two forms to be variants of each other, they must be similar in form. This is because *realization variant* implies that there are two different realizations of the same sign/lexeme. Two forms with the same meaning but substantially different forms are not realization variants but synonyms or lexical variants, and therefore belong to separate, independent signs.

In order to decide whether two forms are realization variants of each other, one needs criteria for what counts as sufficient formal similarity.

In much of the research literature, we find the term *phonological variants*, which are defined as forms that differ in only one parameter and have the same lexical meaning. Here are some examples from DGS; the parameters in which the variants differ are highlighted:

handshape	HOW-LONG1A	
	HOW-LONG1B	
location	BLOOD1A	
	BLOOD1B	
	BLOOD1D	
movement	BLUE2A	
	BLUE18A	
direction of movement	TO-KNOW-STH2A	
	TO-KNOW-STH2B	
orientation	ROOM1A	
	ROOM1B	
passive hand	AGENCY1A	
	AGENCY1B	
number of hands	PLANT1A	
	PLANT1B	

²⁶ Instances of phonological processes are, e.g., assimilation to the form of the preceding or following sign or syllable reduction in compounds; prosodic phenomena would be changes in form due to expressivity or emphasis; text-structural phenomena would be e.g., abbreviations and contractions in repeated occurrences of the same sign(s).

²⁷ In iLex, lexeme-level variants are indicated by letters, e.g., [AGENCY1A](#), [AGENCY1B](#) or [BLOOD1A](#), [BLOOD1B](#). They are often, but not always, modeled as subtypes of the same type, e.g., [KITCHEN3A](#), [KITCHEN3B](#), [KITCHEN3C](#) are variants that are all subtypes of the type [TO-STIR5-SSAM](#).

²⁸ In practice, one form becomes the *de facto* 'main variant' because it is selected as the citation form (lemma) for an entry. A similar practice is found in the *Ordbog over Dansk Tegnsprog* (Centre for Sign Language, 2008-2017), where one variant is shown as the lemma, but the other variants are included in the entry and can be played as a movie (cf. entry ALDRIG). For information on how we select the citation form see Section 3.

Some researchers decide purely on the basis of this formal criterion whether to classify two forms as phonological or lexical variants (cf. Fenlon et al. 2015, 178f.; Hollman 2010, 141).

In earlier IDGS projects on LSP dictionaries of DGS and in the DGS-Korpus project, we consider an additional criterion: If the two form variants in question belong to iconic signs, then they can only be realization variants if they have the same underlying image (and image-producing technique). Note that realization forms whose connection to the underlying image has become less transparent through phonological processes such as reduction or stylization are still treated as realization variants.

Since we take both form and iconic relations into account, we are less strict with regard to the degree of form similarity two sign forms have to exhibit in order to count as variants of each other. Thus, we prefer to speak of *realization variants* rather than phonological variants. We consider two forms with the same meaning to be realization variants of each other even if they differ in more than one phonological parameter if they have a similar form and at the same time are linked through iconicity (or when one form has presumably evolved from the other by processes of language change²⁹). If, however, similar forms can be traced back to two different underlying images, then they are not treated as realization variants, because we take this as an indication of the signs' independent origins.

2.1.3.1. Determination of realization variants

When determining realization variants, we only take into consideration robustly occurring, conventionalized forms.³⁰ Performance-based variation, rare forms, or those resulting from the linguistic context are not considered. Below we list our rules for determining realization variants, as they follow from the notion of realization variants laid out above.

Rule: Lexemes with the same meaning and a similar form but no discernible underlying image are treated as realization variants if they differ in no more than one phonological parameter. This includes one- vs. two-handed variants.

Rule: Lexemes with the same meaning and a similar form that are transparently or robustly iconic are treated as realization variants if they share the same underlying image and possibly an image-producing technique.

Example: GLASSES

GLASSES1A	“  “	‘glasses’
GLASSES1B	“  “	‘glasses’

- The two forms differ in handshape, but both realize the same image (“glasses frame framing the eyes”) by means of the same technique (substitutive technique).

Rule: Lexemes with the same meaning and a similar form that are transparently or robustly iconic are not treated as realization variants if they have different underlying images or/and different image producing techniques.

²⁹ Non-iconic signs are assumed to undergo the same language change processes (e.g., assimilation of the shape and position of the passive hand with the active hand, a downward shift of the place of articulation on the head, etc.), which may lead to the formation of realization variants (see Frishberg 1975).

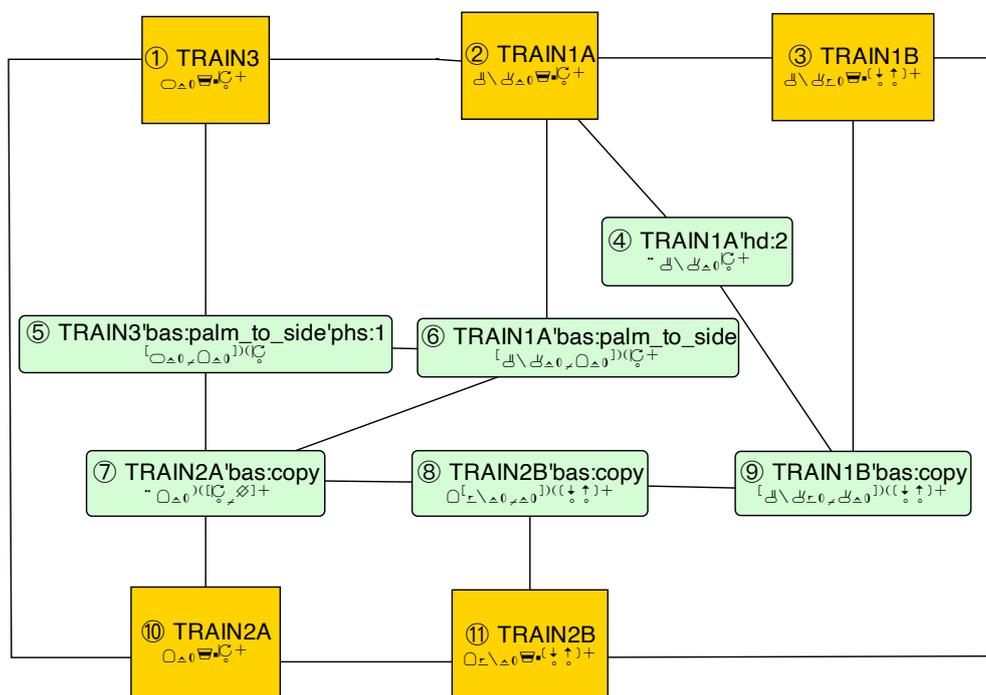
³⁰ To determine conventionalization status, we consider number of tokens and number of informants that use the form. If a form variant is used by only one informant, it may be classified as a person-specific (idiosyncratic) form.

the passive hand represents the side of a bowl or pot in [KITCHEN3B](#) and is thus iconically motivated (substitutive technique), the handshape of the passive hand in [KITCHEN3C](#) is the result of two phonological processes: assimilation of the passive hand to the handshape of the active hand and a relocation of the passive hand resulting in contact between the hands.

We assume that such related forms are realization variants even if their form differs in more than one parameter.

Example of variant chain/network:

In a chain or network of variants (where several forms have the same meaning), distant nodes may differ substantially from each other, while individual nodes differ only slightly from their immediate neighbors.



The network above illustrates connections between observed meaning-level realization forms for the meaning ‘train’.³² The forms in yellow boxes at the periphery of the diagram form a variant chain, but their various realization forms (in the green boxes) branch out in such a way that a variant network results. If we look at two nodes that are further away from each other,³³ for instance node ① (one-handed sign, fist handshape, circular movement) and node ⑨ (two-handed sign, U/V handshape, up-and-down movement, passive hand remains still) the difference is relatively large. When we consider the nodes that connect ① and ⑨, however, each only differs in one or two parameters from the next, e.g. ① $\text{O}_{\Delta 0} \text{W} \cdot \text{C}_{\text{C}}^+$ ② $\text{D} \backslash \text{D}_{\Delta 0} \text{W} \cdot \text{C}_{\text{C}}^+$ ③ $\text{D} \backslash \text{D}_{\Delta 0} \text{W} \cdot (\uparrow \downarrow)^+$ ⑨ $[\text{D} \backslash \text{D}_{\Delta 0} \text{D}_{\Delta 0}] (\uparrow \downarrow)^+$.

³² The figure represents the current database entries; a more detailed analysis to determine senses is still pending.

³³ The numbering of the nodes has no hierarchical implications but serves only to facilitate reference to individual nodes.

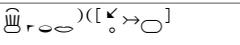
2.1.4. Related signs

Related signs exhibit form similarities and if they are iconic, their underlying images are also similar. The common form elements are related by the interpretation of their iconic motivations or through their similar usage (for non-iconic elements). Related signs of this kind are independent signs, which we discuss here only in contrast with variants and word forms.

Form elements of related signs may also carry associative meanings (e.g., a place of articulation on the head may be associated with cognitive processes, a place of articulation on the upper body/chest with emotions/psychological processes).

Example 1:

An example of iconically related signs are the two lexemes [TO-COMPREHEND1](#) and [TO-BE-SILENT1A](#).

TO-COMPREHEND1		'to grasp', 'to understand'
TO-BE-SILENT1A		'to be silent', 'to keep a secret'

They look very similar and differ only in their place of articulation (the head/forehead vs. the mouth). Their similarity is based on the fact that both signs draw on and visualize similar metaphors in the image: Grasping and holding something. In [TO-COMPREHEND1](#), the entity to be understood is 'grasped' by the mind, which is associated with the location forehead, the relevant metaphor³⁴ is FACTS ARE OBJECTS that can be grasped.³⁵ In [TO-BE-SILENT1A](#), information is, as it were, 'withheld' or 'held' in the mouth, the relevant metaphor being WORDS/INFORMATION ARE OBJECTS that can be held. The two lexemes share a handshape and movement characteristics with other signs, e.g., with [TO-GRAB1A](#) ('to grab, to catch'), where the underlying image is that of physically grasping and holding with the hand. Since [TO-COMPREHEND1](#) and [TO-BE-SILENT1A](#) use the same metaphorical source domain (physical objects) for their underlying images, it follows that they should also be formally similar. In other words, they are iconically related, but we should not consider them as lexicalized word forms that originated from a common source sign, e.g. [TO-GRAB1A](#).

Example 2:

A second example are the related signs [TO-SWITCH-OFF-HEAD1-\\$\\$AM](#), [DISCREET1-\\$\\$AM](#), [TO-SWITCH-OFF-VOICE1-\\$\\$AM](#) and [ATTITUDE-PERSONAL1-\\$\\$AM](#). They all share a common form element (handshape and a twisting motion at the wrist), which can be interpreted as the common image component of 'turning a switch'. These signs are based on the conceptual metaphor THE BODY IS A MACHINE,³⁶ which can be operated with switches and where individual functions can be turned on and off.³⁷

DGS also has a sign [TO-TURN7-\\$\\$AM](#), in which the same image component 'turn a switch or key' is present, but it is to be understood in a concrete rather than a figurative sense.

³⁴ Conceptual metaphors are usually represented in the literature by means of capital letters, which do not represent glosses here.

³⁵ see Lakoff/Johnson (2003, 10): IDEAS (or MEANINGS) ARE OBJECTS.

³⁶ see Lakoff/Johnson (2003, 28): THE MIND IS A MACHINE.

³⁷ German also makes use of this conceptual metaphor, cf. *abschalten (geistig)* 'to switch off (mentally)' vs. *abschalten (Gerät)* 'to switch off (a device)' or *Einstellung (an einem Gerät)* 'setting (on a device)' vs. *innere Einstellung* 'attitude'.

TO-SWITCH-OFF-HEAD1-\$SAM	$\text{O}^2 \text{r} \text{e} \text{o} \text{r} \text{o} \text{X} \rightarrow \text{o}$	'to switch off mentally', 'relax (recover)'
DISCREET1-\$SAM	$\text{O}^2 \text{r} \text{e} \text{o} \text{r} \text{o} \text{X} \rightarrow \text{r} \text{o}$	'discreet', 'not giving away anything', 'without mouthing', 'not to speak', 'to keep one's mouth shut'
TO-SWITCH-OFF-VOICE1-\$SAM	$\text{O}^2 \text{r} \text{e} \text{o} \text{r} \text{o} \text{X} \rightarrow \text{e}$	'(speaking) without using one's voice'
ATTITUDE-PERSONAL1-\$SAM	$\text{O}^2 \text{r} \text{e} \text{o} \text{r} \text{o} \text{X} \rightarrow \text{o}$	'attitude', 'opinion'
TO-TURN7-\$SAM	$\text{O}^2 \text{r} \text{e} \text{o} \text{r} \text{o} \text{X} \rightarrow \text{e}$	'to turn something', 'to lock'/'to unlock', 'open (not closed)', 'locked'/'closed', 'to adjust (something on a device)', 'to turn on/off', 'switch'

Since these signs differ from each other both in their forms and in their meanings (at the lexeme level), and since there is no real overlap in their meanings, the question of realization variants or word forms of a common lemma sign does not really arise here. Moreover, [TO-TURN7-\\$SAM](#) has no conventional meanings with an established morphological body reference from which the other forms could have arisen as lexical modifications.³⁸ What connects these signs is a similar use of metaphor that leads to shared form and image components, which are performed at specific body locations depending on the intended meaning. These are not spontaneous, productive word formations, but established signs, since their exact meanings cannot be predicted from their form.

2.1.5. Considerations on the scope of an entry

One of our goals is to provide a comprehensive view of the formal and semantic range of a (manual) sign within one entry. Identification and inclusion of *form* variants are determined with regard to the basic form, that is, an unmodified or unmarked form which can express all meanings/senses of a sign.³⁹ On the *semantic* side the connection between senses is established by semantic relations, through shared membership in a common semantic domain, as well as, where applicable, via the underlying image of a sign (including a shared image-producing technique).⁴⁰

Form and meaning relations are the basis for how we address a range of issues. These issues are discussed below.

2.1.5.1. Mouthing

One sign can have a number of typical mouthings (which may, but need not, belong to different senses of that sign). We do not consider a difference in mouthing a criterion for establishing separate entries. When we describe two forms as identical vs. different, we are always referring to the manual sign.

Rule: A lemma sign can be used with different mouthings. Different mouthings do not result in the creation of separate entries.

³⁸ However, a literal meaning of [TO-TURN7-\\$SAM](#) with body reference modification is possible (e.g., tightening screws in an android robot), but such uses are not conventional, at least not according to corpus data.

³⁹ For more details on the identification of the basic form, see Section 3: Choice of Lemma. Note that the basic form is possibly the basic form of a variant.

⁴⁰ Underlying image and image-producing technique can only be considered for iconically transparent signs; for semantic domains and connections in detail, see Section 2.2 on polysemy.

2.1.5.2. *Different range of modifications for different senses*

In a polysemous sign, different word forms may be associated with different senses. In particular, there may be senses whose range of possible modifications is limited compared to that of senses that can occur in all word forms. We proceed as follows:

- Senses that are close to the underlying image and potentially have richer modification possibilities (re-iconizations) are grouped in one entry with senses that are indirectly conveyed by the image.
- Nominal and verbal usages are not separated into different entries.
- Meanings that are lexicalized for the basic form and are not used with other forms can be covered as senses within the entry. (See rule 5a in Section 2.4.1.)
- Lexicalized modifications, on the other hand, are not included in the entry because they cannot be expressed by the basic form.⁴¹

Two senses having a different range of possible modifications does not automatically disqualify them from being grouped under one lemma sign.⁴²

2.1.5.2.1. *Re-iconizations*

Senses with forms that may undergo re-iconization should not be moved into separate entries. Keeping them in one entry does justice to semantic and formal relations between senses and their forms and prevents obscuring the connections between transparent iconicity and the conventionalization of meanings.

Only literal-iconic senses can undergo spontaneous re-iconization, whereas conventionalized senses that arise from metonymic or metaphorical extension do not typically exhibit re-iconization. While it is possible to determine quite accurately which senses of a sign can be re-iconized and also along which dimensions a form can be re-iconized, the re-iconized forms themselves can hardly be described as conventionalized. We mark the literal-iconic sense of a sign as ‘re-iconizable’ in the dictionary (and add an indication of the modifiable dimensions where applicable). Among other things, this means that differences concerning the property ± re-iconizable and the resulting range of forms do not influence the decision whether or not to assign two subtypes (types.level=1) to the same type (types.level=3) or whether to treat them in one entry.

Rule: Re-iconizability of individual senses should not lead to separate entries for re-iconizable and non-re-iconizable senses.

2.1.5.2.2. *Categorial indeterminacy with respect to part of speech*

Many signs are categorially underspecified. A given sense can be used verbally (as a predicate) in some contexts, nominally (as an argument) in another, and adjectivally or adverbially in yet other contexts. In some utterances, a particular use of a sign cannot be assigned a part-of-speech unequivocally given the current state of research.⁴³

To illustrate, let’s look at the lemma sign [TO-EAT-OR-FOOD2-\\$\\$AM](#), which can be used nominally as well as verbally, without a difference in form.

⁴¹ For more details see Section 2.1.2: Lexicalized modifications.

⁴² For an alternative view, see Fenlon et al. (2015, 192f.), for whom precisely this criterion is decisive for the separation of the signs CHILD and LOW, which are identical in their basic form.

⁴³ On the categorial indeterminacy of signs see Ebbinghaus (1998); Erlenkamp (2000) speaks of multifunctional signs in this context, for Polish Sign Language (PJM) see also Linde-Usiekniewicz/Rutkowski (2016).

Examples:

- nominal use: [Signer is waiting for the food they ordered. The device lights up ...] TO-LET-KNOW1A-towards_speaker I2 [TO-EAT-OR-FOOD2](#)⁴⁴ TO-GET1A, sense: food (prepared food);
- verbal use: [In the past, people in the port had bad teeth] MEAT1A [TO-EAT-OR-FOOD2](#)⁴⁵ CAN1-not, sense: to eat (to consume food);
- unclear: [description of Christmas Eve: “I always spend Christmas Eve with my son”]. [TO-EAT-OR-FOOD2](#)⁴⁶ TO-DRINK1 DONE1A; the signed utterance could be interpreted verbally: [We] eat [and] drink. [When we are] done [with that], then ... or nominally: [First there is] food [and] drink. After that ...⁴⁷.

Apart from its re-iconization potential in verbal uses ([to eat](#) – to take in food)⁴⁸, the vast majority of tokens of [TO-EAT-OR-FOOD2](#) exhibit no morphological cues as to which part of speech they are used in.

Since some signs have both nominal and verbal uses, and in some contexts, a given token’s part of speech cannot be determined morphosyntactically, distinguishing signs into a verbal and a nominal lemma sign represented by two different entries would not allow us to assign categorially indeterminate tokens to one of the two entries. Thus, one would have to create three entries for the same sense: a nominal use, a verbal use, and a nominal-verbal indeterminate use. Such a classification would not only be confusing for users of the dictionary, but it would also suggest a clear part-of-speech distinction, which is not supported by our current state of knowledge on DGS. Such a practice would not be an appropriate description of the sign’s use.⁴⁹ Therefore, we follow a different practice of lemma establishment, which is quite common in spoken language dictionaries, and describe the nominal, verbal, adjectival etc. uses of a sign in one entry (if necessary, as separate senses).

We apply this practice not only to signs that share a form across different uses, as is the case for [TO-EAT-OR-FOOD2-\\$\\$SAM](#), but also to signs that can take different kinds of modifications depending on the part of speech they are used in. Depending on the type of use a different subset of the range of forms can be realized. While verbal uses are often modified to show aspect or the source and goal of an action, typical nominal modifications include reduplication (with or without sideways displacement) for plural marking. Some modifications are attested in both nominal and verbal uses, for example modifying the place of articulation of a sign on the body where the relevant body part contributes to the meaning.

For these reasons, we opted to describe verbal, nominal, etc., and indeterminate uses of a sign in one entry even if different uses exhibit different subsets of the overall range of forms of a sign.⁵⁰

⁴⁴ [ilex://tags.id=2926674](#).

⁴⁵ [ilex://tags.id=2452856](#).

⁴⁶ [ilex://tags.id=3146531](#).

⁴⁷ DONE THEN is a very common discourse structuring device in DGS that implies a chronological sequence of events (‘afterwards’).

⁴⁸ See, for example, [ilex://tags.id=2488538](#).

⁴⁹ Svensén (2009, 100f.) discusses this issue in the chapter *The formal-grammatical approach* and concludes that a purely formal-grammatical approach to lemma establishment leads to results that clash with dictionary users’ linguistic intuitions.

⁵⁰ In the Polish Sign Language Dictionary, the issue of parts of speech in lemma establishment is treated similarly. Different uses (‘types of use’, e.g., *nominal use* etc.) are grouped in one entry (Linde-Usiekiewicz/Rutkowski 2016, 382).

Word forms or modification behavior cannot and thus should not serve as a strong or absolute criterion for the establishment of lemma signs at the present time. However, where a clear morphological grouping is discernible, we take it into account during lemma establishment (e.g., for the group of number-incorporating signs).⁵¹ For some signs, processing of the corpus data over time may suggest a specific grouping, as is conceivable, for instance, for the group of signs with morphological body reference. At present, however, we do not group into separate entries according to modification potential only, since this might lead to separating formally and semantically related realizations of a sign. Furthermore, it is unlikely that the entire modification potential of a sign is documented in the corpus.

Rule: Nominal, verbal and other uses of a sign are grouped in one entry, even if they differ in their word forms (or realized modification possibilities) beyond the basic form and therefore realize different subsets of the sign's range of forms.

To what extent and in how much detail word forms will be represented in the final dictionary is yet to be determined. The pre-release entries are to contain general information on the modification potential of a sign and of its senses, provided that it can be abstracted from the data during the editorial process.

2.2. Range of meanings of a lemma sign (polysemy)

Several meanings of a sign (lemma sign) can be conventionalized. If this is the case, the sign is polysemous, and we refer to the entirety of its lexicalized meanings as its range of meanings.

All meanings of a certain sign form are considered meanings/senses of the same polysemous sign provided that they are connected to each other or to the sign form in at least one of the following ways. The connections can be of different kinds and can either be direct or indirect, but overall, we take them as indications that we are dealing with one lemma sign.

Rule: All meanings of a polysemous sign are described in one entry.

When writing dictionary entries, we only consider established, i.e., lexical meanings of a sign.

2.2.1. Connection via the underlying image of the sign

Iconic signs can be lexicalized for multiple meanings, all of which fit the underlying image of the sign, but which need not have much in common beyond that. The iconic sign form provides through the image on which it is based a plausible and appropriate contextualization for all senses/meanings and thus provides a connection between them.

Examples:

- [PLANE1-\\$\\$AM](#) $\cdot \square_{\text{flat}}^{\text{horizontal}}$ ('table', 'floor', 'carpet', 'square', 'field', 'lake': each denotes an object with a flat, horizontally oriented surface),
- [VEIL1-\\$\\$AM](#) $\cdot \square_{\text{person}}^{\text{veil}}$ ('nun', 'nurse', 'bride', 'veil': each denotes a person with a veil or the veil itself),

⁵¹ For instance, [HOUR1A/B](#) is formally and semantically close to [TIME7A/B/C-\\$\\$AM](#) but receives a separate entry because (among other reasons) it allows number incorporation, while [TIME7A/B/C-\\$\\$AM](#) does not (cf. AP10-2016-02 *Vorgehen bei der Analyse für die Artikelschreibung (Wörterbuch) [Procedure for the analysis for article writing (dictionary)]*: appendix).

mouthed German word. In that case, the sign can lexicalize for these meanings as well. The different meanings of the sign are then related to each other via a German word of similar or identical form. This means that the polysemy, homonymy or homophony of the German word (or the homomorphy, that is the similarity of the visible form of German words used as mouthings) in part motivates the use of a certain signed form for a certain meaning. This in turn justifies the grouping of corresponding meanings/senses in one entry, even if the image underlying the sign no longer illustrates this particular sense.

Example:

- The sign for *Essen* ‘food (food/nutritional intake)’ [TO-EAT-OR-FOOD2-\\$\\$AM](#) is also used for the city Essen.

Rule: If a sign has an established mouthing that corresponds to a homophonic or homographic word in German, and if the meanings of the homophonic or homographic partner also apply to the sign, the sign is considered polysemous. This is also true if the sign is robustly iconic and the image of the sign does not match all of these established meanings.

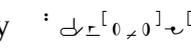
Connections via mouthing also include cases where the relevant German words display only an assumed identity of form as mouthings (i.e. homomorphy).

Example:

- The sign [WING1-\\$\\$AM](#)  is conventionally used for both the meanings ‘angel’ [German: *Engel*] and ‘grandson’ [German: *Enkel*].

Sometimes the connection is established via a partial word (simplex, morpheme) or a morphologically related word according to the rebus principle.

Example:

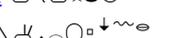
- [TO-SHOOT1B-\\$\\$AM](#)  bas:copy for ‘to shoot’ ([TO-SHOOT1D](#)) [German: *schießen*, related to *Schuss* ‘shot’] and ‘committee’ ([COMMITTEE1A](#)) [German: *Ausschuss*].

2.2.4. Homonymy versus polysemy

If none of the connections described above in the context of polysemy apply, according to our rules we do not consider a sign polysemous. Homonyms are described in separate entries.

Rule: Meanings conventionalized for the same sign form are treated as homonyms if there is no discernible connection between the meanings or between the meanings and the underlying image (as described for polysemes).

Example:

- [UNIVERSITY3-\\$\\$AM](#) 
- [HANOVER1-\\$\\$AM](#) 
- [BLUE1-\\$\\$AM](#) 

Rule: Iconic signs that share a manual form but differ in the iconic content (underlying images, image-producing technique) motivating this form are treated as independent, homonymous signs. They receive separate entries. This rule applies even if other factors favor a polysemy analysis.

Example of signs with the same (basic) form but different underlying images:

- [TO-MILK1A-\\$\\$AM](#) $\bar{\cdot} \circ_{\pm 0} (\downarrow \uparrow)^+$ (stylized milking)
- [COMPETITION1-\\$\\$AM](#) $\bar{\cdot} \circ_{\pm 0} (\downarrow \uparrow)^+$ (sometimes one person is on top, sometimes the other person)

If there is no underlying iconic image or if that image is opaque, ambiguous, or too abstract, one has to resort to the other criteria for determining homonymy vs. polysemy.⁵⁶

2.3. Basic principle of lemma establishment

Rules of lemma establishment for the dictionary are to be applied to groups of tokens that are similar with respect to a pre-defined set of criteria and which therefore have been grouped into one type/subtype during annotation. As lexicographers, we then ask: Are two lemma candidates (signs, types.level=3) or a subset of the forms and meanings identified for these lemma candidates (subtypes, types.level=2, 1, 0) groups of tokens that are best described in one entry or in separate entries?

Basic principle: An entry should cover the entire range of forms of a lemma sign, including modifications (word forms) and realization variants. In principle, all conventionalized meanings (senses⁵⁷) of this lemma sign also belong to its entry.⁵⁸

General rule: Different uses of signs (i.e., concrete realizations of signs and their meaning in context) are described in one entry, if three conditions are fulfilled: Their **forms are tightly connected** (i.e. they belong to the **same range of forms**), a **semantic or other content-related connection** exists, and no other, especially no formal, reasons stand against such a grouping.

In the following section, we present and discuss our process(es) for lemma establishment. The typical, i.e., conventional uses (i.e. lexemes) are the point of departure from which we methodically derive according to our rules a superordinate grouping and assignment to lemma signs (at the sign level/entry level), taking different aspects and a comprehensive view of the available data and the situation it presents into consideration.

In practice, however, the data are already pre-structured by the previous annotation process. Therefore, the procedure does not have to be repeated for every single lexeme in relation to every other lexeme. On the contrary, usually we start by looking at types (types.level=3) as lemma candidates and check them against their type structure and type environment to ascertain whether there are contradictions or discrepancies to the rules of lemma establishment described here. Where necessary, we make adjustments accordingly.

⁵⁶ As we discussed for the sign [TECHNOLOGY1-\\$\\$AM](#), see above in Section 2.2.2.

⁵⁷ *Sense* actually refers to the lexicographically processed, differentiated representation of meaning in the dictionary, where different uses are distinguished. *Meaning* represents a pre-analytical perspective that is independent of the dictionary. Lemma signs are established on the basis of coarser conventional meanings, as they are captured via the annotation of lexemes in iLex. Fine-grained categorization into distinct senses occurs only after lemma establishment (see AP10-2016-02). In this paper, *meaning* and *sense* are used largely interchangeably, since meaning can always be thought of in terms of the representation of meaning in the dictionary (as senses) and a sense also always describes meaning.

⁵⁸ This rule forms the basis for lemma establishment, even if the final product may not list all meanings in an entry depending on the available corpus data. The rule applies regardless of whether and in how much detail a sign's range of forms is described in the entry.

2.4. Simple cases and examples

Decision-making about merging into one entry vs. separating in different entries can be categorized in different cases. We start first from the lexeme level (types.level=1, 0). The following examples usually involve only one dimension at a time (i.e., either meanings or forms), to illustrate our decision criteria. More complex cases that involve several dimensions are discussed in Section 2.5 below.

Rule 1: Signs that have nothing to do with each other receive separate entries, even if their forms look very similar.

Case 1a:

Different forms + different meanings → separate entries.⁵⁹

If different forms have different meanings, they are two different, independent lemma signs, which are described in two separate entries. This rule intuitively makes sense for signs with clearly different forms and meanings like [UMBRELLA1](#) ('umbrella') and [INTERPRETER1](#) ('interpreter'). These signs share neither a formal nor a semantic connection. The forms of the two signs differ in several phonological parameters and have no iconic similarity.

UMBRELLA1	$[\ominus_{\tau} 0_{\neq} \ominus_{\tau} 0]^{[1]_{\neq} 5]} X^{\uparrow}$	'umbrella'
INTERPRETER1	$[\partial]_{\tau} 0_{\neq} \partial]_{\tau} 0] X^{\rightarrow \uparrow}$	'interpreter', 'to interpret'

Signs that are very similar in form (e.g. they differ only in one parameter) but are semantically unrelated are also different lemma signs and as such receive separate entries. Take, for example, [SYSTEM2](#) and [RELIGION1](#):

SYSTEM2	$\sim [\bar{\partial}]_{\tau} 0_{\neq} \bar{\partial}]_{\tau} 0] \sim (\leftarrow X$	'system'
RELIGION1	$\sim \sim [\bar{\partial}]_{\tau} 0_{\neq} \bar{\partial}]_{\tau} 0] \sim (\leftarrow X$	'religion', 'Christ', 'Jesus'

Case 1b:

Different forms + same meaning → separate entries

Signs that differ substantially in their manual form (without taking into account mouthing and mouth gestures) are considered independent signs and are given separate entries. We proceed in this way even when the signs share a conventional meaning. Since they are not connected via their form, we consider these signs synonymous, i.e., lexical variants. The different regional variants for 'woman' are an example of lexical variation:

WOMAN4A	$\sim \sim \downarrow^5 \tau 0 5] (\leftarrow +$	'woman (female adult)'
WOMAN6	$\downarrow \wedge 0 \} \cdot X^{\downarrow}$	'woman (female adult)'
WOMAN3A	$\partial \tau \rightarrow 2 X$	'woman (female adult)'
WOMAN2A	$\downarrow^5 \tau 0 \cup \cdot (X^1 \uparrow) \rightarrow 0$	'woman (female adult)'

Rule 2: Functional modifications (both grammatical and iconic) are part of the range of forms of a lemma sign and thus do not receive separate entries from the basic form.

⁵⁹ Compare Battison (2005: 240): "If two signs are made differently, and have different meanings, this is good evidence that they are separate signs".

These forms are connected via a shared semantic domain and via a shared (basic) form. Therefore, they were subsumed under a single sign (types.level=3) during the annotation process and they are described in one dictionary entry. The lemma sign is polysemous.

Rule 4b: Homonyms receive separate entries.

Rule 4c: Identical forms with distinct meanings are grouped into different homonymic lemma signs if they do not exhibit a semantic connection, nor a connection via their mouthing or their underlying image.

Case 4b/c:

Same form + different meanings → separate entries.

Example of homonymous signs:

- The forms of the lexemes [THIS-AND-THAT2](#) and [PERSPECTIVE1A](#) are the same, but they have different underlying images. [THIS-AND-THAT2](#) is a stylized pointing gesture and represents a broad spectrum or a range of different things. [PERSPECTIVE1A](#) shows in a stylized way how someone's point of view expands.

THIS-AND-THAT2	" $\square_{\Delta 0} X[\rightarrow_{\rightarrow} \uparrow_{\uparrow}]$	'a variety of things'
PERSPECTIVE1A	" $\square_{\Delta 0} X[\rightarrow_{\rightarrow} \uparrow_{\uparrow}]$	'perspective'

Because of their different meanings, and, more importantly, because of the difference in the iconic motivation of their form [THIS-AND-THAT2](#) and [PERSPECTIVE1A](#) do not belong to the same type (at the sign level) in annotation and are not subsumed under the same entry in the dictionary. They are described as two distinct and separate lemma signs in the dictionary.

Rule 4d: Different underlying images and/or image-producing techniques are a sufficient criterion for treating iconic signs as homonyms rather than as a polyseme, even if there may be other reasons that justify treatment as a polyseme.

Rule 5: Lexicalized modifications are considered independent lemma signs.

The semantic and formal connection between lexicalized modifications and their (suspected) sign origin will be recorded in the dictionary via the note "related sign".

Rule 5a: Additional meanings that are lexicalized for the *basic form* with their own limited subset of forms (this subset may consist of the basic form only) are not separated into independent lemma signs but are described in one entry with the other senses of the basic form.

Example:

- For the basic form of [TO-CUT2D-\\$\\$SAM](#) (cut on the back of the hand), the meaning 'surgeon' has also become conventionalized. This form does not relate to a specific body part as a morphologic element. However, because it shares the same basic form with the other senses of [TO-CUT2D-\\$\\$SAM](#), it is most efficiently described in the same entry together with the other senses.⁶¹

Rule 6: Signs that are merely related with regard to their iconicity are covered in separate entries.

⁶¹ Moreover, the meaning 'surgeon' is linked to the \$\$SAM-sign by regular polysemy: activity → actor.

Case 6: Similar or/and iconically closely related forms + possibly semantically close, but still different meanings → separate entries.

Closeness in form and meaning are based on coincidence or on a coincidental iconic relationship (see examples from Section 2.5.1.1.3). Iconically related signs are described in separate entries, but each entry contains a note about their relatedness.

2.5. Summary of entries (at the type level): Complex cases

The basic rules of lemma establishment are described above with clear cases and examples. Now we discuss difficulties that may arise in processing lemma candidates when we try to apply these rules to more complex cases. In particular, we look at how the rules can be applied to comprehensive type structures and environments such as may arise from the token-type mapping in iLex.

In general, our aim is to present the dictionary entries as comprehensively and informatively as possible without losing clarity, and to give language users an overview of the uses and the semantic environment of a sign.

One difficulty in lemma establishment is that in some cases the data can only be described by a simultaneous combination of several simple cases, i.e., they involve multiple dimensions. Assessing the data then requires weighing a number of criteria or factors against each other, and any decision on whether or not to create one, two, or multiple entries is informed by what importance is given to each factor.

To illustrate, when checking for polysemy we take into consideration the variant sets of different senses - which therefore are assumed to have already been determined. In turn, when we assess the variant status of two forms, we look at how many of their meanings overlap, which in turn presupposes that a polysemy check has already been performed. Both steps are mutually dependent and presuppose each other, therefore both must be considered at the same time when assessing the data.

Given that several criteria are weighed against each other in complex cases, different rules and criteria may suggest different decisions (separation vs. merging of forms/meanings) in each case and, if necessary, may lead to a re-grouping of forms and meanings in the light of the whole situation presented by the data.

Moreover, corpus data cannot be expected to be complete. Two senses may not have the exact same range of forms because of data gaps for one sense, or two forms may differ in their range of meanings because not all meanings are attested in the corpus for one of the forms. The question then arises whether the differences in form or meaning range are basic and so important that this fact should result in separating the forms/meanings into two entries, or whether the gaps are accidental, can therefore be neglected and need not be reflected in a given entry.

The different dimensions and perspectives on the data which one must consider simultaneously when making a lemma establishment decision include the following:

- Variant set: Starting from one conventional meaning, we consider which realization variants are attested for this meaning.

- Range of meanings: Starting from one form, we consider which meanings it conventionally expresses and whether these meanings can be subsumed under a polyseme entry or whether they belong to different homonyms (polysemy analysis).
- Synopsis and evaluation: The forms and meanings attested for the subtypes of the lemma candidate are compared and checked for correspondences. We take into consideration the degree of overlap between the ranges of meanings of the different variants and the overlap between the variant sets attested for different conventional meanings.

2.5.1. Realization variants at the level of the lemma sign

Realization variants can only be determined for individual meanings (see Section 2.1.3), since variants have to share the same meaning in order to be considered realization variants (criterion: interchangeability in typical contexts without a change in meaning).

On the sign level, however, it also makes sense to speak of realization variants if two signs do not differ significantly in their range of meanings, i.e., in the majority of their (core) senses and how these are used. In this case, it makes sense to describe these signs in one dictionary entry with several form variants.

If one wants to abstract from the variants of each individual conventional meaning (at the lexeme level) to variants of the entire lemma sign that should be listed in the entry, one finds that often, not all variant candidates cover all meanings/senses, or that meanings either do not occur with all variant candidates or with different variant sets.

So, there are cases where meanings/senses

- have the same variant set compared to other senses (the variant sets of sense 1 and sense 2 are identical),
- have a restricted variant set compared to other senses (the variant set of sense 1 is a subset of the variant set of sense 2),
- have a different but overlapping variant set compared to other senses (the variant set of sense 1 overlaps with the variant set of sense 2).

Based on the form, a realization variant can

- have the same range of meanings as another realization variant (the range of senses of variants 1 and 2 are identical),
- have a restricted range of meanings compared to another realization variant (the range of senses of variant 1 is a subset of the range of senses of variant 2),
- have a different range of meanings compared to another realization variant (the range of senses of variant 1 and variant 2 overlap).

A polysemy analysis of a variant form A may suggest grouping a set of lexemes into one polysemous entry. By looking at the individual meanings and their realization variants, one may find that there are different variant sets for each meaning. If one zooms in on one of those variants (B) and determines its range of meanings via a polysemy analysis, one may find that A and B overlap considerably in their meanings, which would argue in favor of them being variants of the same lemma sign at the sign level. However, it is also possible that both A and B have several meanings that they do not share, which would point towards them not

being fully interchangeable at the sign level. Things get even more complicated when there are more than two potential variants, so that different senses really have distinct variant sets.

When abstracting away from individual lexemes or variants to the lemma sign as a whole, individual details should not be over-generalized. At the same time, entries must not become cluttered and unwieldy by detailing every exception and restriction on usage for individual variants or senses. It is therefore necessary to decide on a case-by-case basis which grouping is the most practicable and user-friendly.

2.5.1.1. Restricted variant sets and restricted range of meanings

Let's look at the following example: [TO-EAT-OR-FOOD2-\\$\\$SAM](#) (628 tokens) and [TO-EAT-OR-FOOD1-\\$\\$SAM](#) (376 tokens).

Well-documented established meanings (selection)		Form 1	Form 2
		$\text{TO-EAT-OR-FOOD2-}\$ \SAM	$\text{TO-EAT-OR-FOOD1-}\$ \SAM
'food (nutrition)', 'meal', 'to eat (to take in food)'	A	TO-EAT-OR-FOOD2	TO-EAT-OR-FOOD1
'breakfast'	B	BREAKFAST2	BREAKFAST1
'(at) noon'	C	LUNCH-OR-NOON9	
'Essen (City in Germany)'	D	ESSEN-CITY1	

- All meanings of form 1 are interrelated, 'food' (A) and 'breakfast' (B) are semantically close, 'at noon' (C) is metonymically close to 'food' (A) as the typical time for eating lunch. The 'city Essen' (D) relates to 'food' (A) via its mouting. Considered by itself, form 1 would be described as a polyseme with all four lexemes covered in one entry. The same is true for form 2 and its two lexemes.
- Forms 1 and 2 differ only in handshape and share the same image (holding something and putting it into one's mouth) and the same image-producing technique (manipulative). They share important core senses (A and B). For A and B, the difference in form does not correspond to a difference in meaning, so at the lexeme level, forms 1 and 2 are realization variants of each other. We then have to determine whether these two forms should also be considered realization variants for the overall dictionary entry.
- According to the corpus data, form 1 has more (attested) established meanings than form 2 (namely 'noon, midday' (C) and 'the city of Essen' (D)). Form 2 therefore either has a more limited range of meanings, or the data for form 2 are simply incomplete.⁶²
- Senses D and probably C have a restricted variant set, because they do not usually occur as meanings of form 2.

What is the best strategy for dealing with this data when composing a dictionary entry? There are several options:

- Option 1:

⁶² 68 deaf people provided feedback on these forms by March 2017. Of these, 29% were familiar with Form 1 for 'noon/midday', but only 7% reported familiarity with Form 2 for this sense. This suggests that Form 2 – 'midday' – is rather not established. We did not ask about the city of Essen; however, 13 out of a total 100 participants stated that they use Form 1 to refer to the city. Nobody commented that they used Form 2 in this sense. This suggests that only Form 1 is lexicalized for the city of Essen.

One entry with two realization variants: Since senses C and D are attested for form 1, we assume that they should in principle also be available for form 2 (generalization). The fact that C and D are not attested for form 2 is treated as an accidental gap in the data.

<p>Entry 1 Form variants: Form 1, Form 2 Senses: A B C D</p>

- Option 2:
 One entry with two realization variants, but the entry contains information about usage restrictions for form 2 (form 2 + C, form 2 + D are not attested). In this case, we assume that the data reflect actual usage accurately and that certain senses only occur with certain variants, or certain variants do not cover a particular sense. Given the gaps in our data set, such explicit statements are difficult, since they represent the data in terms of absolutes.

<p>Entry 1 Form variants: Form 1, Form 2 Senses: A B C only attested for Form 1 D only attested for Form 1</p>

Another possible solution is to create two entries to represent the data more accurately. A number of options are available:

- Option 3:
 The criterion *non-identical range of meanings* serves to divide the two entries by form (vertical division) which results in entries with similar forms which overlap in a part of their meanings.⁶³

<p>Entry 1 Form variants: Form 1 Senses: A B C D</p>	<p>Entry 2 Form variants: Form 2 Senses: A B</p>
---	---

- Option 4:

⁶³ In the final product, the entries would receive a synonym reference for the senses in question.

The criterion *non-identical variant set* serves to divide the two entries based on meanings or groups of meanings that have different variant sets (horizontal division).⁶⁴ This division results in several entries that share at least one form variant.

<p>Entry 1 Form variants: Form 1, Form 2 Senses: A B</p>
<p>Entry 2 Form variants: Form 1 Senses: C D</p>

In this particular example, a horizontal division would trigger a second horizontal division. The senses ‘the city of Essen’ (D) and ‘noon’ (C) would no longer be housed in a joint entry, since they would no longer be linked via sense A ‘food’ in the same entry:

<p>Entry 1 Form variants: Form 1, Form 2 Senses: A B</p>	<p>Entry 2 Form variants: Form 1 Senses: C</p>	<p>Entry 3 Form variants: Form 1 Senses: D</p>
---	---	---

We have decided to proceed as follows:

Rule: If some senses exhibit a restricted variant set or some variants have a restricted range of meanings, they will still be described in one entry as long as the overlap in core senses is convincing enough.

Where the core senses overlap and some senses exhibit a restricted variant set, we describe both forms with all their meanings in one entry. It’s at the lexicographer’s discretion to decide where generalizations (option 1) are justifiable and where the overall picture⁶⁵ suggests noting usage restrictions in the entry (option 2). In the example at hand, feedback from deaf signers would lead us to favor option 2 and add the reservation “only attested for Form 1” to sense D ‘city of Essen’.

Some reasons for opting for joint entries wherever possible are:

- In principle, we would like to show the use of a lemma sign (including all of its forms and meanings) in a clear, straightforward way, which can be achieved by presenting the information in one entry. Hence, dividing forms and/or meanings pertaining to one

⁶⁴ Compare this practice with an editorial rule for the Danish Sign Language dictionary: “Another principle is that a sign described as having several variant forms can include only senses that can all be expressed through one particular variant (shown as the citation form in the dictionary). Any variant-specific sense gets its own independent sign entry in the dictionary.” (cited in Langer et al. 2016, 149)

⁶⁵ We also draw on data from previous projects, feedback data, further consultations with informants, and the intuition of our deaf employees to reach a decision.

lemma sign into several entries has to be well-motivated. We wish to avoid a large number of entries based on the same form.

- In cases where we have a restricted variant set, options 1 and 2 offer sufficient space to provide the required corresponding information in a nuanced way. There is no need for separate entries as long as a joint entry does not get too complex.
- We do not wish to divide lemma signs based on potentially incomplete data. A division would be more serious than a comment accompanying the relevant sense and would suggest several independent signs.
- If there is a large overlap in the range of meanings, the shared senses need only be described once for both forms.

2.5.1.2. Variant chains or networks

Sometimes it is advisable to divide forms vertically (by form) into two or more entries for structuring or legibility purposes. This may be the case for variant chains or networks, where several forms differ along several formal parameters - e.g. some differ in handshape and others differ in movement.

Rule: Depending on the overall assessment, variant chains or networks (division according to forms) can be divided into several entries (by form) if a single entry would be too cluttered and difficult to read.

For the variant network shown in Section 2.1.3, one would also have to consider the usefulness or necessity of creating several entries. This decision would have to be based on weighing different rules and criteria against each other based on the data. For ‘train’, such a decision can only be made after lemma revision.

2.5.1.3. Different variant sets

Different variant sets are a reason for us to divide potentially form- and meaning-related subtypes into different entries according to their meanings (horizontally).

Rule: Different variant sets for different established meanings/senses result in these meanings/senses being described in separate entries, preferably together with their given realization variants.⁶⁶

If one variant set is a subset of another, they do not count as different. Only if both senses have variants that are not shared by the other sense do we consider variant sets as different.

Example:

Let’s illustrate this rule using the form $[\text{O}_{\text{L}} \text{O}_{\text{R}} \text{O}_{\text{H}} \text{O}_{\text{D}}]$. Its conventional meanings are, among others:

- ‘(government) agency’: the lexeme [AGENCY1B](#) with two realization variants: [AGENCY1A](#) and [AGENCY1C](#). The variant using

⁶⁶ A variant set for iconic signs probably results from alternatives that all visualize the same image, as well as the sign form changing over time. Therefore, iconic signs with different underlying images are likely to have different variant sets. Non-iconic signs may also have different variant sets (presumably because of language change processes acting on the original sign and the form). Therefore, the criterion of having different variant sets applies to all signs (iconic and non-iconic) and could partially replace the criterion of “having different underlying images” in iconic signs (this being one of several interpretations for types with similar forms), or support the latter as an observable criterion.

two fists has presumably developed as a stylized form of [AGENCY1A](#) and can thus be traced back to the underlying image “to stamp”.

- ‘to do’, ‘to work’, ‘to make’: the realization variant of [TO-WORK2](#) that does not have movement repetition (TO-WORK2- $\$SAM$ 'phs:1, types.level=2). The sign occurs with these meanings with and without repetition, but repetition is common.⁶⁷ This form has the realization variant: [WORK1- \$\\$SAM\$](#) $[\text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \uparrow \text{C}^{\text{X}} +$ ⁶⁸
- ‘hard’, ‘stone’, ‘concrete (material)’: the sign [HARD2A- \$\\$SAM\$](#) with the following realization variants: [HARD2B- \$\\$SAM\$](#) $[\text{O}_{\text{I}} \text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \leftarrow \text{V} (\uparrow \uparrow \text{X})$ (‘stone’), [HARD1B- \$\\$SAM\$](#) $[\text{O}_{\text{I}} \text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \uparrow \downarrow \text{X} \rightarrow \text{E}$ (‘hard’), [HARD1A- \$\\$SAM\$](#) $[\text{O}_{\text{I}} \text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \uparrow \downarrow \text{X}$ (‘hard’, ‘stone’), [HARD1C- \$\\$SAM\$](#) $[\text{O}_{\text{I}} \text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \sim \downarrow \text{X}$ (‘hard’).
- ‘done’, ‘end(ing)’, ‘already’: the sign [END1- \$\\$SAM\$](#) $[\text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \uparrow \uparrow \text{X}$ (‘end(ing)’, ‘stop’, ‘done’, ‘already’) and the lexeme [DONE1B](#) $[\text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \uparrow \sim \text{X}$ (‘end(ing)’, ‘stop’, ‘done’, ‘already’...)⁶⁹.
- ‘can’, ‘must’, ‘possible’: the realization variant [CAN2B- \$\\$SAM\$](#) $[\text{O}_{\text{I}} \text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \uparrow \sim \text{X}$ (‘can’, ‘possible’) and [CAN1- \$\\$SAM\$](#) $[\text{O}_{\text{I}} \text{O}_{\text{E}}] \uparrow \rightarrow \text{A}$ / $[\text{O}_{\text{I}} \text{O}_{\text{E}}] \uparrow \rightarrow \text{A}$ (‘can’, ‘possible’).
- ‘nevertheless’: sign with and without repetition (regional: East (Leipzig)): no variants.

The different variant sets attested for each of the established meanings of this form point towards dividing the different senses into separate entries rather than treating them as one polyseme. For almost all of these senses, additional criteria can be adduced to show that they do not belong to the same lemma sign, for instance different underlying images, a lack of semantic coherence, or a different basic form.

2.5.1.4. Major differences in the range of meanings (small intersecting set)

Rule: Two forms are not considered realization variants if the set of senses they share (especially core senses) is not large enough. In that case, they are described in two separate entries, even if they can be traced back to the same or to related underlying image(s).

Two forms are best split into two entries if their range of meanings is substantially different, especially if their status as realization variants at the sign level would only be founded on one shared sense. The case becomes even clearer if the one shared sense is marginal and each form encodes further, more central meanings that it does not share with the other form. This holds true even if the forms are based on the same or on similar images, and especially if the only shared sense is the literal-iconic meaning for which these signs can be re-iconized.

Example 1:

The literal-iconic reading ‘to ring’, ‘to toll’ and ‘bell’ respectively is not sufficient for establishing variant status.

- [CHAIRPERSON2- \$\\$SAM\$](#) $[\text{O}_{\text{I}} \text{O}_{\text{E}} \text{O}_{\text{X}} \text{O}_{\text{U}}] \uparrow \rightarrow \text{L} +$
 - image: ringing a handbell or hitting a table with a hammer;

⁶⁷ Another criterion for a division into different entries is having a different basic form (see below).

⁶⁸ Here the presumed image underlying the sign is the tapping of e.g., a hammer against an object. Thus, the signs would have different images, which would be another criterion for separating the forms.

⁶⁹ The variants with fists vs. with flat hands have approximately the same range of meanings – therefore it makes sense to describe them in one entry (see Langer et al. 2016).

- conventional meanings: ‘chairperson’, ‘minister’, ‘chancellor’, ‘president’, ‘colleague’;
- literal-iconic meaning: ‘to ring a (hand) bell’, ‘hand bell’.
- [TO-RING-BELL1-\\$SAM](#) 〇² ㄹ^ㄷ
 - image: ringing the clapper of a bell that is hanging higher up by using a string;
 - conventional meanings: ‘alarm’, ‘alarm clock’, ‘break’;
 - literal-iconic meaning: ‘to ring a bell’, ‘bell’ (?).⁷⁰

Example 2:

The literal-iconic sense ‘to put up one's feet’ and the derived meaning ‘to rest’, ‘to take a break’ are not sufficient to warrant variant status.

- [TO-LIE-LEG1A-\\$SAM](#)^{hd:2x} ㄷ\ㄷㄹ\ㄷ0[11 ㄹ5]X[ㄷ>ㄹ]
 - image: to put up one’s legs (with legs crossed);
 - conventional meanings: ‘break’, ‘rest’, ‘to rest’, ‘passive’, ‘pension’, ‘retirement’, ‘pensioner’;
 - literal meaning: ‘to put up one's legs’, figurative meaning: ‘to take a break’, ‘to rest’, ‘to be idle’.
- [TO-LIE-LEG1A-\\$SAM](#) ㄷ\ㄷㄹ\ㄷ0[ㄷ>ㄹ]
 - image: to put up one’s legs (parallel side by side);
 - conventional meaning: ‘bathtub’ (core sense);
 - literal meaning: ‘to lie down (with one’s legs up)’, figurative: ‘to take a break’, ‘to rest’, ‘to be idle’ (?).

2.5.1.5. Different core meanings

Rule: If potential realization variants share a core sense but differ in other core senses, if they therefore have clearly different foci of meaning, the data are potentially better presented in two separate entries.

Example:

Entry	Form	Senses			
		1	2	3	4 ...
Entry 1	HEART1A-\$SAM ㄷㄹ0•ㄱ)(±X	‘heart (organ)’	‘Hartz IV’ (marginal)		
	HEART1C-\$SAM ㄷㄹ3 ㄹ0•ㄱ)(±X	‘heart (organ)’	‘Hartz IV’ (marginal)		
Entry 2	HEART1B-\$SAM ㄷㄹ\ㄹ0•ㄱ)(±X+	‘heart (organ)’		‘the heart beats’ (marginal)	‘heart’ (in the figurative sense of ‘being well-disposed towards someone’, ‘warm-hearted’, ‘sensitive’, ‘empathetic’), ‘temper/disposition’, ‘good-natured’, ‘soft-hearted’, ‘conscience’

⁷⁰ Aside from the fact that people may differ in their opinions about how closely the images are related (holding a hand bell and ringing it vs. using a rope to set in motion the clapper of a large overhead bell), the marginal and iconic-literal meanings ‘to ring (bell)’ and ‘bell’ (if conventional at all – which is still debatable) would not be sufficient to warrant a common entry. The core senses and ranges of meanings are different and the forms differ as well.

- [HEART1A-\\$\\$AM](#), [HEART1B-\\$\\$AM](#), [HEART1C-\\$\\$AM](#) have a common underlying image: pointing to the heart (image-producing technique: indicating). The forms differ only in one parameter, namely handshape; they have the sense ‘heart (organ)’ in common. This points towards treating them as realization variants. The range of meanings for [HEART1A](#) and [HEART1C](#) is also the same: both mean ‘heart (organ)’ and ‘Hartz IV’. These two forms should be described as variants in one entry. In contrast, while [HEART1B](#) also has the sense ‘heart (organ)’ and can further indicate that the organ ‘*heart is beating*’, these senses seem to be rather marginal. The sign is mainly used for ‘heart’ in a figurative sense: ‘well-disposed towards someone’, ‘radiates warmth’, ‘affectionate’, ‘emotional’, ‘sensitive’, and related meanings: ‘disposition’, ‘good-natured’, ‘soft-hearted’, ‘conscience’. Only [HEART1B-\\$\\$AM](#) is attested with these senses, which moreover seem central to this sign. Therefore, we treat the form $\square_{\text{r}} \setminus \langle \text{0} \cdot \square \rangle (\downarrow \chi +)$ as a separate entry (criterion: too little overlap in ranges of meanings). [HEART1A/C-\\$\\$AM](#) and [HEART1B-\\$\\$AM](#) differ considerably in their range of meanings, even though one sense of [HEART1B-\\$\\$AM](#) constitutes the core sense of the other two forms. A single entry for all three forms would be very complex and would have to list exceptions for all variants and/or senses. Two separate entries provide a clearer, more user-friendly structure.

2.5.2. Different basic forms

Rule: If two meanings/senses have a large overlap in their range of forms but different basic forms (forms that are central to a particular sense),⁷¹ then the data is best represented by two different entries.

Example 1:

- $[\square_{\text{r}} \setminus \langle \text{0} \cdot \square \rangle (\downarrow \chi +)]$ occurs with and without movement repetition with the meanings ‘done’, ‘completed’ and ‘work’, ‘to work’. For [DONE2](#) (‘done’, ‘completed’, ‘end’, ‘conclusion’), frequency considerations suggest that the non-repeated variant is central and therefore basic. [DONE2](#) also has quite a few tokens with repetition, but these could be influenced by the number of syllables in the corresponding mouthing, e.g. *fer-tig* ‘finished’, *en-de* ‘end’. In contrast [TO-WORK2](#) (‘work’, ‘to work’) is most frequently signed with repeated movement, suggesting that this form is the more common or central one. It may occur in a shortened form, i.e., without repetition, especially in compound-like sign combinations. In the dictionary, we would divide [TO-WORK2](#) and [DONE2](#) into separate entries whose lemmas (citation forms of the sign) are similar, but not identical in form.⁷²

Example 2:

- Within the type hierarchy of the lemma sign candidate [\\$COLOURS8-\\$\\$AM](#) $\downarrow^5 \text{r} \square \cdot \uparrow$ we find two commonly used, i.e. lexicalized forms $\downarrow^5 \text{r} \square \cdot \uparrow$ (shaking) and $\downarrow^5 \text{r} \square \cdot \rightarrow$ (single flip of orientation) both used to express two meanings or ranges of meanings respectively: Meaning A covers the color of ‘yellow’, and range of meanings B relates to the ‘color and material of gold’ and metonymic and metaphorical meaning extensions of ‘gold’.

⁷¹ For criteria for choosing the basic form, see Section 3.

⁷² The meanings ‘to work’ and ‘done’ could be judged as semantically close enough to be considered two senses of one polyseme. However, the difference in basic forms is not the only criterion militating for separation into two entries. Different variant sets also suggest separate entries.

Range of meanings (A): 'yellow' (color)	Range of meanings (B): 'gold', 'golden' (color) 'gold', 'golden' (material) 'gold', 'golden' (several metonymic, metaphorical extension of gold)
---	--

Both forms are attested for both meaning groups. However, when choosing the basic form to represent the whole entry and taking frequency into consideration the suitable basic form for 'yellow' (A) would be the shaking variant while for 'gold' (B) it would be the single flip variant.

	𐀓 ⁵ 𐀔𐀕 𐀖𐀗𐀘𐀙 (shaking)	𐀓 ⁵ 𐀔𐀕 𐀖𐀗𐀘𐀙𐀚 (single flip of orientation)
A 'yellow'	104 tokens	30 tokens
B 'gold'	6 tokens	47 tokens

The basic form of the main variant of an entry is also displayed in the German index describing the above data in one entry would result in 'gold' being listed with the shaking-form. In the interest of clarity and user-friendliness, we would represent the data in the dictionary as two lemma signs with different basic forms in two entries. Each entry contains also a minor variant that is identical to the basic form of the other entry's main variant (comp. Entries 18 and 1682).

3. Choice of lemma

The lemma sign is represented in the dictionary by one of its realization variants (in entries and in cross references to entries). The realization variant selected for this purpose is called *lemma* or *citation form*.

Rule: One of the attested realization variants of a lemma sign is selected as its citation form (lemma).⁷³

The following criteria for determining the citation form of a lemma sign are to be understood as a provisional starting point, which can be expanded and rendered more precise as our understanding of the data advances.

There are several aspects to consider when choosing a lemma:

- Range of modifications: We choose one representative form among the range of attested modifications (the word forms) for a variant. We call this form the *basic form*.
- Differences in realization within a realization variant (without any grammatical function, e.g., one-handed vs. two-handed forms or \pm repetitions): One of potentially several forms has to be chosen as the representative form. Since there is no orthographic standard, we have to select a form. In the following section, this form is called the *standard form*.
- Variant set: A variant is selected from the variant set that is representative of the entire variant set. We call this variant the *main variant*.

⁷³ This is common in German lexicography but does not apply to all languages. The lemma can also be a constructed form, which is unattested in the data but contains important information about the lemma sign.

In the dictionary, we use the standard form of the basic form of the main variant as the lemma for an entry. In the isolated studio recording, the lemma is shown without a mouthing, since it is meant to represent the entire lemma sign and not only one of its senses. Adding mouthing might pick out only one sense.

3.1. Choice of the standard form

The choice of standard form is mainly about the questions of whether a one-handed vs. two-handed realization variant and whether a form with or without repetition should be set as the standard form. A standard form already exists in the transcription database. Deviations from this form are either deemed irrelevant, in which case the relevant token may sometimes receive a note ‘deviation’, or they constitute frequent deviations and are annotated by using qualifiers.⁷⁴ Looking at frequencies in the annotated data then allows us to confirm or correct previously made conclusions.

3.2. Choice of the basic form

There is no standard in the lexicography of DGS yet for selecting a *basic form* from the range of modified forms of a variant. In the type-structure in iLex, one form has already been chosen as the basic form. The choice of this form is usually based on the intuition of deaf annotators and may also be influenced by which realization variant was first annotated in the annotation process. If we take the intuitive decisions seriously, we see that the basic form is not always selected on purely formal criteria, because different lemma signs with a similar range of forms are treated differently. Since different lemma signs often do not exhibit uniformity with respect to their forms and allow different modifications, we consider it prudent to determine basic forms on a case-by-case basis by looking at the data in context. At least at the current stage, this practice seems preferable to treating all lemma signs according to the same formal rules.

Some criteria for an intuition-based selection of the basic form:

- Which form is perceived as the most neutral, unmarked form?
- How is the sign signed in isolation, i.e., when it is not embedded in an utterance? (e.g., in response to the question: *Which sign do you use for [elicit sign via finger alphabet or a paraphrase in DGS]? or in meta-linguistic use, i.e., citation of the sign.)*
- Which form of the sign is most likely to be perceived as prototypical?

Choice of the basic form:

- If a lemma sign occurs in only one (grammatical) form, this form is the basic form.
- If there is an unmodified form, this is usually chosen as the basic form if there are no other reasons that might be prioritized (unmodified means non-directional, without

⁷⁴ Especially in the case of iconic signs, different perspectives may be considered when using formal criteria for choosing the standard form or the main variant form. Either one starts from the iconically most complete form (and thus prefers a two-handed variant) or from the most basic and thus one-handed form. If the latter is to be considered the core of the sign, then the passive hand may add further elements to its underlying image. It seems more reasonable, however, to choose the more frequently occurring form – assuming it is the more common one – for a given lemma rather than applying the same formal rules to all (iconic) signs. The same guideline applies to the decision which standard form should be selected, the one-handed or two-handed version, or a form with or without repetition.

plural or aspect marking or the like, not placed meaningfully in a particular location in the signing space or on the body, etc.).

- For directional signs, a signer-centered form is typically preferred, unless usage or semantics suggest otherwise.
- For signs that relate to varying body parts, the basic form is usually signed on the passive hand.
- For number-incorporating signs, we choose the neutral form (without incorporation) if there is one. If there is no neutral form, a default digit will be used across number-incorporating signs. This digit has yet to be determined; at the moment we favor the number 3, because as of now, it exhibits the least variation.

In addition, viewing the data in context can inform the choice of a basic form:

- Are there several senses with restricted modification behavior?
- If so, then the basic form is the one that can express all senses.

3.3. Choice of the main variant

If there are several variants, one has to be selected as the main variant. As a lemma, the main variant represents the entire entry. If possible, the variant that is most common is chosen. To identify this variant, we use three criteria.

- Scope of meaning: The variant with the largest scope of meaning (i.e., the variant that covers all senses) is preferred, so that it can serve as a frame of reference to which other variants with fewer senses can be related.
- Frequency: Preference is given to the variant with the most tokens.
- Regional distribution: Preference is given to the variant that is the most frequent nation-wide or the one that covers the largest regional area of all variants.

If applying these criteria results in different competing forms as candidates for the main variant, then the criteria are weighed against each other taking into consideration the entire lemma sign and the main variant is determined pragmatically.

4. References

4.1. Linguistics and lexicography

- Atkins, B.T. Sue / Rundell, Michael (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Battison, Robbin (2005): Signs Have Parts. A Simple Idea. In: C. Valli, C. Lucas, & K.J. Mulrooney (eds.): *Linguistics of American Sign Language. An Introduction*. 4th ed. Washington, DC: Gallaudet University Press, pp. 230–241.
- Ebbinghaus, Horst / Heßmann, Jens (2000): Leben im Kommunikationskonflikt. Zur Ungleichsprachigkeit Hörender und Gehörloser. In: Hess-Lüttich, Ernest W. B. / Schmitz, H. Walter (eds.): *Botschaften Verstehen. Kommunikationstheorie und Zeichenpraxis. Festschrift für Helmut Richter*. Frankfurt a.M.: Lang, pp. 47–66.
- 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.
- Ebbinghaus, Horst (1998): Warum deutsche Wörter wesentliche Bestandteile der Deutschen Gebärdensprache sind. Part 1 in: *Das Zeichen* 45, pp. 443–451; part 2 in: *Das Zeichen* 46, pp. 594–611.
- Erlenkamp, Sonja. 2000: *Syntaktische Kategorien und lexikalische Klassen. Typologische Aspekte der Deutschen Gebärdensprache*. München: Lincom Europa.
- Fenlon, Jordan / Cormier, Kearsy / Schembri, Adam (2015): Building BSL Signbank: The Lemma Dilemma revisited. In: *International Journal of Lexicography* 28 (2), pp.169–206. DOI: <https://doi.org/10.1093/ijl/ecv008>.
- Frishberg, Nancy (1975): Arbitrariness and iconicity: Historical Change. In: *American Sign Language*. Language 51, pp. 696–719.
- Hanke, Thomas / Jahn, Elena / Wähl, Sabrina / Böse, Oliver / König, Lutz (2020): SignHunter – A Sign Elicitation Tool Suitable for Deaf Events. In: Eftimiou, Eleni et al. (eds.): Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives. 12th International Conference on Language Resources and Evaluation (LREC 2020), Marseille, Frankreich, 16.5.2020. Paris: European Language Resources Association (ELRA), pp. 83–88. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/20030.pdf> (last accessed 2024-04-17).
- Herbst, Thomas / Klotz, Michael (2003): *Lexikografie*. Paderborn: Schöningh.
- Hollman, Liivi / Sutrop, Urmas (2010): Basic Color Terms in Estonian Sign Language. In: *Sign Language Studies* 11 (2), pp. 130–157. DOI: <https://doi.org/10.1353/sls.2010.0011>.
- Johnston, Trevor / Schembri, Adam (1999): On Defining Lexeme in a Sign Language. In: *Sign Language & Linguistics* 2 (2), pp. 115–185. DOI: <https://doi.org/10.1075/sll.2.2.03joh>.
- 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, pp. 379–404.
- Lakoff, George / Johnson, Mark (2003): *Metaphors we live by*. Chicago: Chicago University Press. [First publication: 1980].
- Langer, Gabriele (2005): Bilderzeugungstechniken in der Deutschen Gebärdensprache. In: *Das Zeichen* 70, pp. 254–70.

- Langer, Gabriele / Troelsgård, Thomas / Kristoffersen, Jette / Konrad, Reiner / Hanke, Thomas / König, Susanne (2016): Designing a Lexical Database for a Combined Use of Corpus Annotation and Dictionary Editing. In: Efthimiou, Eleni / Stavroula-Fotinea, Evita / Hanke, Thomas / Hochgesang, Julie / Kristoffersen, 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. 143–152. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/16014.pdf> (last accessed 2024-04-17).
- Langer, Gabriele / König, Susanne / Konrad, Reiner (2007): Hintergrundinformationen. In: Konrad, Reiner / Langer, Gabriele / König, Susanne / Schwarz, Arvid / Hanke, Thomas / Prillwitz, Siegmund (Hg.): *Fachgebärdenlexikon Gesundheit und Pflege*. 2 vols. Seedorf: Signum.
- Linde-Usiekniewicz, Jadwiga / Rutkowski, Paweł (2016): The division into parts of speech in the Corpus-based Dictionary of Polish Sign Language. In: *Proceedings of the XVII EURALEX International Congress*. Tbilisi, 2016. URL: <http://euralex2016.tsu.ge/publication2016.pdf> (last accessed 2020-10-30).
- Mathur, Gaurav / Rathmann, Christian (2012): Verb agreement. In: Pfau, Roland / Steinbach, Markus / Woll, Bencie (ed.) (2012): *Sign Language. An international Handbook*. (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK); 37). Berlin [...]: de Gruyter, pp. 136–157. DOI: <https://doi.org/10.1515/9783110261325.136>.
- Meir, Irit (2012): Word classes and word formation. In: Pfau, Roland / Steinbach, Markus / Woll, Bencie (eds.) (2012): *Sign Language. An international Handbook*. (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK); 37). Berlin [...] de Gruyter, pp. 77–111. DOI: <https://doi.org/10.1515/9783110261325.77>.
- Pfau, Roland / Steinbach, Markus (2012): Plurality. In: Pfau, Roland / Steinbach, Markus / Woll, Bencie (eds.) (2012): *Sign Language. An international Handbook*. (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK); 37). Berlin [...]: de Gruyter, pp. 112–136. DOI: <https://doi.org/10.1515/9783110261325.112>.
- Pfau, Roland / Steinbach, Markus / Woll, Bencie (2012): Tense, aspect, and modality. In: Pfau, Roland / Steinbach, Markus / Woll, Bencie (eds.) (2012): *Sign Language. An international Handbook*. (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK); 37). Berlin [...]: de Gruyter, pp. 186–204. DOI: <https://doi.org/10.1515/9783110261325.186>.
- Schlaefter, Michael (2002): *Lexikologie und Lexikographie. Eine Einführung am Beispiel deutscher Wörterbücher*. Berlin: Erich Schmidt Verlag.
- Schwager, Waldemar (2012): Morphologie: Bildung und Modifikation von Gebärden. In: Eichmann, Hanna / Hansen, Martje / Heßmann, Jens (eds.) (2012): *Handbuch Deutsche Gebärdensprache. Sprachwissenschaftliche und anwendungsbezogene Perspektiven*. Seedorf: Signum, pp. 61–110.
- Svensén, Bo (2009): *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge [...]: Cambridge University Press.
- Troelsgård, Thomas / Kristoffersen, Jette (2016): *The Danish Sign Language Dictionary – Editorial rules*. [English translation of unpublished, internal editorial rules, last update: Jan 2013.]
- Wähl, Sabrina / Langer, Gabriele / Müller, Anke (2018): Hand in Hand - Using Data from an Online Survey System to Support Lexicographic Work. In: Bono, Mayumi et al. (eds.): *Workshop Proceedings. 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community. Language Resources and Evaluation Conference (LREC)*, Miyazaki, Japan, 12 May 2018. ELRA, pp. 199–206. URL: <https://www.sign-lang.uni-hamburg.de/lrec/pub/18025.pdf> (last accessed 2024-04-17).

4.2. Language resources and dictionaries

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

Centre for Sign Language, University College UCC (2008–2017): *Ordbog over Dansk Tegnsprog*. URL: www.tegnsprog.dk (last accessed 2024-04-17).

Konrad, Reiner / Langer, Gabriele / König, Susanne / Schwarz, Arvid / Hanke, Thomas / Prillwitz, Siegmund (Hg.) (2007): *Fachgebärdenlexikon Gesundheit und Pflege*. Hamburg: Universität Hamburg, Institut für Deutsche Gebärdensprache und Kommunikation Gehörloser. URL: www.sign-lang.uni-hamburg.de/glex/ (last accessed 2024-04-17).

Johnston, Trevor (1989): *AUSLAN Dictionary. A Dictionary of the Sign Language of the Australian Deaf Community*. Maryborough, Victoria: Australian Print Group.

Łacheta, Joanna / Czajkowska-Kisil, Małgorzata / Linde-Usiekiewicz, Jadwiga / Rutkowski, Paweł (eds.) (2016): *Korpusowy słownik polskiego języka migowego*. [Corpus-based Dictionary of Polish Sign Language]. Warsaw: University of Warsaw, Faculty of Polish Studies. URL: <http://www.slownikpjm.uw.edu.pl> (last accessed 2024-04-17).

Stoesesand, Bernhard (1970): *Tausend Taubstummengebärden*. Berlin: Selbstverlag.

5. Definitions of terms

Gloss name: The part of a gloss that consists of a German or English word or a combination of German or English words (without digits, letters, or the addition of qualifiers).

Number incorporation: The lemma sign can be realized with different handshapes and can be one- or two-handed depending on which number it integrates. The lemma sign can integrate a handshape for the numbers from 1-10 into the sign form, which then corresponds to the amount represented by this handshape. There are signs whose handshape always represents a semantically relevant number (e.g., the actual number signs, glossed as e.g., \$NUM-ONE ...) as well as those that have a neutral handshape with an independent meaning but can also integrate a handshape with quantifying meaning, e.g., WEEK (as in NEXT WEEK vs. WEEK-1, WEEK-2, WEEK-3 ...). We count both types of signs among the number-incorporating signs.

Polyseme: A sign form that is used for different related meanings; i.e., the meanings need to share a transparent connection, or the meanings and the form need to share such a connection.

Realization variants:

- Realization variants on the level of meaning/on the lexeme level are slightly different forms which – if they are iconic – go back to the same underlying image (including having the same image-producing technique). They have the same meaning, are interchangeable in typical contexts without any change in meaning, and coexist with a certain degree of stability (criterion: conventionalization).
- Realization variants on the level of (lemma) signs are slightly different forms which – if they are iconic – can be traced back to the same underlying image (including having the same image-producing technique), they coexist with a certain frequency (criterion: conventionalization) and they have a sizeable overlap in their range of meanings, i.e., they have essential meanings/senses (core senses) in common as established meanings.

Type environment: By type environment, we mean those other types and sub-types in iLex that are or could potentially be related in some way to the form type being analyzed. In particular, this includes other word forms of the type, modifications, identified realization variants, synonymous signs, signs that are identical in form, and signs that are similar in form or iconically related.

Variant chains: Several forms that may all have the same meaning; immediate neighbors in the chain differ only slightly (e.g., only in one parameter), while the endpoints of the chain may have entirely different forms.

Variant set: We call a variant set the group of very similar sign forms used for a particular meaning (either a sense or, in iLex, a group of senses: still unspecified meaning/concept marked on the lexeme) and for which the following applies:

The differences in form are judged to be relevant and stable enough that a separate description as realization variants seems reasonable (as opposed to minimal individual differences in how the variant is signed). Some criteria are:

- attestation in the corpus: frequency, distribution;
- linguistic intuition: relevance of the difference, documented by independent glosses at the lexeme level (A/B) – as opposed to deviations or qualified types.

The difference in form cannot be attributed to different underlying images; in iconic signs, the forms are based on the same image. The different forms are interpreted as different realizations of a lexeme with a particular meaning. The variant set can also be realized as a variant chain. If we are dealing with a polysemous sign, we can assume that the variant sets of the conventional meanings/senses are essentially very similar, so that a common variant set at the sign level (valid for the entire range of senses) can be inferred from the variant sets at the lexeme level.