From form to function.
A database approach to handle lexicon building and spotting token forms in sign languages

Reiner Konrad, Thomas Hanke, Susanne König, Gabriele Langer, Silke Matthes, Rie Nishio, Anja Regen
Institute of German Sign Language (IDGS), University of Hamburg
{reiner.konrad, thomas.hanke, susanne.koenig, gabriele.langer, silke.matthes, rie.nishio, anja.regen}
@sign-lang.uni-hamburg.de

Abstract

Using a database with type entries that are linked to token tags in transcripts has the advantage that consistency in lemmatising is not depending on ID-glosses. In iLex types are organised in different levels. The type hierarchy allows for analysing form, iconic value, and conventionalised meanings of a sign (sub-types). Tokens can be linked either to types or sub-types.
We expanded this structure for modelling sign inflection and modification as well as phonological variation. Differences between token and type form are grouped by features, called qualifiers, and specified by feature values (vocabularies). Built-in qualifiers allow for spotting the form difference when lemmatising. This facilitates lemma revision and helps to get a clear picture of how inflection, modification, or phonological variation is distributed among lexical signs. This is also a strong indicator for further POS tagging. In the long term this approach will extend the lexical database from citation-form closer to full-form.
The paper will explain the type hierarchy and introduce the qualifiers used up-to-date. Further on the handling and how the data are displayed will be illustrated. As we report work in progress in the context of the DGS corpus project, the modelling is far from complete.

Keywords: lemmatisation, type hierarchy, citation form modification, variation, qualifiers, full-form lexical database

1. Background

The aim of pre-processing language data in corpus linguistics is lemmatisation. Counting and sorting of words or word forms, part-of-speech tagging, further annotation and analysis rely on machine-readable, lemmatised corpora. Reliability as one of the quality criteria of empirical science depends on how consistently tokens are matched to lexical types.

Whereas written texts of languages with a written tradition are pre-processed more or less automatically, spoken texts have to be written down beforehand. To build a corpus in an oral language with no written tradition, one has to choose an appropriate writing and/or notation system. This is the case for sign languages that have no written tradition. Phonographic notation systems such as HamNoSys or SignWriting were developed to write down the form of a sign and are part of a transcription system. But as the International Phonetic Alphabet (IPA) in spoken languages, inventory and conventions of notation systems are not helpful for lemmatising. Therefore a coding system in the sense of Hulst & Channon (2010) is needed that allows for computerized sorting, counting and comparing of signs.

For coding tokens of sign types, glossing is the most widespread practice. Glosses are written words from the surrounding spoken language or from the researcher's language. Their meaning usually covers one of the lexical meanings of the sign. They are "relatively crude and simplistic" translations (Johnston, 2009: 91). However, a gloss neither represents the contextual meaning of a sign nor does it give any information about the sign form. There are two main reasons why glosses made their way in sign language linguistics: First, glosses are a mnemonic aid. For those having some knowledge of the respective sign language, glosses can be used as a hint to recall the sign. Second, with glosses one can communicate with ease about signs. Using glosses for literal or free translation is misleading. In corpus linguistics we are likely to deal with thousands of signs and sign variants so that the first reason is bound to fail as "it is often very difficult to know with certainty which sign form is actually being referred to by a particular gloss" (Johnston, 2009: 91). The only way to achieve consistent token-type matching is to use glosses as ID-glosses, as unique identifiers of a sign (Johnston, 2010a). This means that glosses function as if they were identifiers. Working with annotation tools like ELAN[1], where tags are not linked to a lexical database, this seems the only way to build reliable lemmatised corpora.

The reason why we developed iLex (Hanke, 2002; Hanke & Storz 2008), an integrated lexical database for sign languages, is to handle large numbers of lexical types and their tokens in a consistent way. Glosses are helpful to represent sign types in the two ways mentioned above, but they are not used as identifiers in iLex. They are one value of a lexical entry amongst others as e.g. the citation form, written in HamNoSys. The identifier of each entry is a numeric code created by the database itself that guarantees its uniqueness and allows for restrictions such as one cannot create a new type entry with an already existing gloss string. In combination with token tiers that only allow for tags whose value is a type ID, the software supports the transcriber in being consistent. This support is essential in a multi-user environment, especially for quality assurance. Lemmatisation does not rely on glosses as

[1] EUDICO Linguistic Annotator; latest version and documentation are online available at: http://www.lat-mpi.eu/tools/elan/.

free text annotations, but is executed by linking tokens to type entries in a unique way.[2]

2. Data base approach and type hierarchy

With no comprehensive dictionary or lexical resource at hand, token-type matching is hard to achieve. As a bottom-up approach in transcribing each token form is far from realistic, the only way out is to build up a lexical database in parallel to segmenting and lemmatising signed utterances. This means that the transcriber constantly has to switch between top-down driven type matching and bottom-up driven adding new type entries (Konrad & Langer, 2009; König et al. 2010).

In spoken languages, lexemes are conventionalised form-meaning pairs. Instantiations (tokens) of lexemes can have different forms according to a limited set of inflected forms. Applying this lexicological and morphological model to sign languages, one has to deal with two issues:

1. Due to its iconic aspects a sign can cover a far wider range of meanings than words. It can be combined with different mouthings to express meanings that are not necessarily semantically related. Leaving aside metaphorical use and homophone calques[3], usually all these meanings are related to the same underlying image. In König et al. (2008) we refer to this process of productively combining signs and mouthings as the "iconic-combinatorial procedure".
2. Until now, no complete descriptive grammar of any sign language exists. It is an open question whether one can define complete form paradigms for different sign classes. It is one of the research directions in sign language corpus linguistics to validate assumptions on part-of-speech classification and e.g. verb modification.[4]

2.1 Type hierarchy: types and sub-types (double glossing)

Our approach to face the first issue was to take the iconicity of signs into account. Identifying lexemes by comparing token forms and meanings following the rule "same form (paradigm), same (lexical) meaning → same lexeme" does not fit the needs of sign languages. In many cases this would result in mapping the spoken language lexicon onto the sign lexicon. In changing the rule into "same form, same iconic value (+ same image producing technique[5]) → same lexeme", things look quite different. In accord with Ebbinghaus & Hessmann, their assumption that signs and words (perceptible as mouthings) contextualise each other mutually and their postulation that "[i]nformation about regular collocations with nonmanually produced units should be part of the lexicographic description of the manual lexicon of a sign language" (Ebbinghaus & Hessmann 2001: 134), we distinguish between conventional and productive sign-mouthing combinations. This procedure is operationalised by double glossing[6] and implemented as a type hierarchy in iLex. Type entries in the table "types" are linked to the table "levels" which defines type dependencies, and what kind of type information can be added. Level-3 types (in the following called types) can be parents of several level-1 types (children; in the following called sub-types). Sub-types cannot be created without a reference to a type (parent). Sub-types are conventionalised sign-mouthing combination with a lexicalised meaning. In most cases the meaning corresponds to the meaning of the mouthed word.[7] Sub-types can only be subsumed to types if they share the same underlying image and the same citation form. This information is stored in the type entry and is valid for all sub-types. Meaning is not entered in the type, but in the sub-type entry. In contrast to so-called productive signs created on the spot, corresponding to "partly-lexical signs", each type, corresponding to "full-lexical signs" (Johnston 2010a) must have at least one lexical meaning, so that each type has to be parent of at least one sub-type. If the form of a token can be identified as an instantiation of the type's citation form, and if the iconic value of the type is valid for its use in context, but the contextual meaning of the token does not correspond to the lexical meaning of a sub-type, this token will be matched directly to the type. In many cases such tokens are productive sign-mouthing combinations covering a wide range of meanings. Matching tokens either to types or sub-types helps to sort "regular collocations with nonmanually produced units" from occasional collocations. Grouping subtypes into types allows for identifying different conventionalised readings of a sign (polyseme[8]) and prevents from mapping the spoken lexicon into the sign lexicon.

[2] Cf. Johnston's "note on the use of an integrated lexical database with ELAN" (Johnston 2011: 16-17).

[3] The DGS sign for 'Enkel' (grandchild) is the same as for 'Engel' (angel) because of the similarity of sound in spoken German. The mouthing of the German words is the same. This phenomenon is not restricted to DGS, e.g. in ASL you will find HUNGRY/HUNGARY.

[4] Cf. Johnston's (2010b: 141) findings on spatial modification of verbs in Auslan that support Liddell's (2003) analysis of indicating and depicting verbs in ASL.

[5] See Langer (2005) for a detailed description, König et al. (2008) for a short version.

[6] See Konrad (2011b: 145-155) for an extensive discussion; see also König et al. (2008), König et al. (2010).

[7] It happens that tokens are articulated without the corresponding mouthing. If the contextual meaning of the sign fits to the lexical meaning, this token will also be matched to the sub-type.

[8] Note that on the one hand in sign language we have to deal with far-reaching lexical ambiguity which is more context sensitive than in spoken languages, on the other hand iconicity is a valid criterion to group related meanings and distinguish lexemes (s. König et al. 2008), which is not applicable to spoken language.

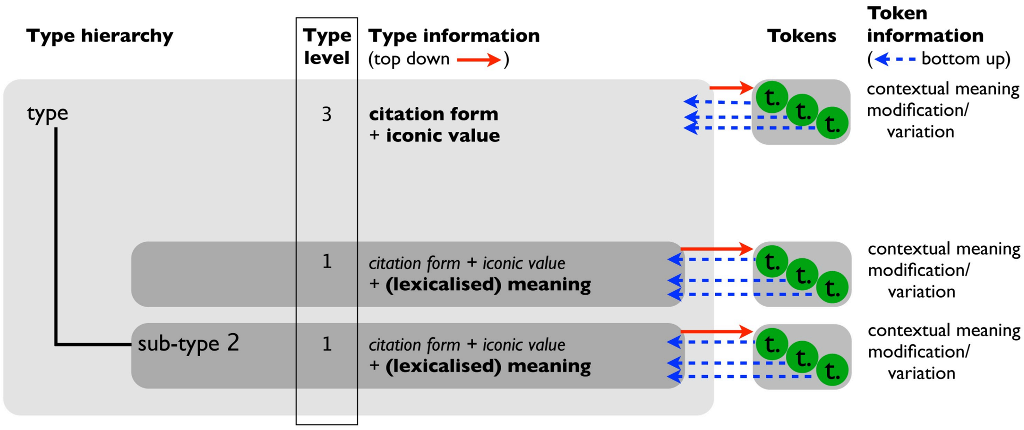

Figure 1: Type hierarchy, type information and token information

2.2 Type hierarchy and qualified types

The second issue regards the signs' potential for variation and modification in context. Differences between token form and citation form can either indicate variation that will be assigned to the phonological level or morphosyntactic patterns. When matching tokens to types, the transcriber has to compare token form and type citation form. Instead of deferring the documentation of token form differences to a second annotation pass (Johnston, 2010a: 116-117) where these differences are annotated in several tiers (orientation, citation modification, or variation tier)[9], we annotate this information to the token tag in the process of lemmatisation. These annotations are one of the main criteria to check whether the token-type matching is correct during the process of lemma revision. As in iLex tokens are linked to a type, all tokens of a type can be listed and sorted by token information such as form difference.[10]

Since 2009 we are modifying the type hierarchy in iLex in order to group different form features. Each type and sub-type can have several qualified types. Qualified types[11] are combinations of types with qualifiers. These qualifiers are form features that can have several feature values (see below). Instead of annotating the form difference to tokens, the transcriber can refer a token to an existing qualified type. This makes lemma revision easier because the tokens of one type are not only pre-sorted by conventional and productive use of signs but also by form features. Figure 2 shows the expanded type hierarchy, in figure 3 the structure is exemplified by parts of the subtree belonging to the type glossed DA1 (there).

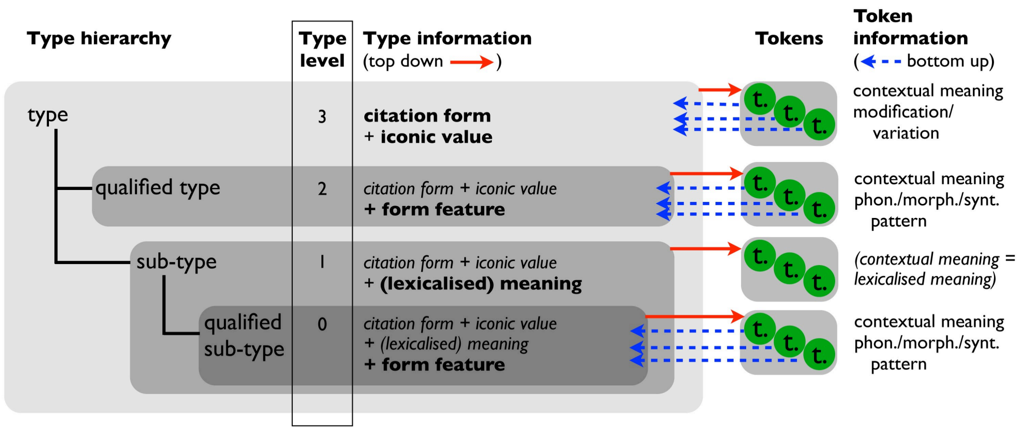

Figure 2: Expanded type hierarchy with qualified types

[9] Cf. Johnston 2011: 53-70: "Secondary processing".

[10] The process of lemma revision in iLex is described in Konrad (2011a pp. 93-96); see also König et al. (2010 and Konrad & Langer (2009).

[11] In the following all what is said about qualified types is also valid for qualified sub-types.

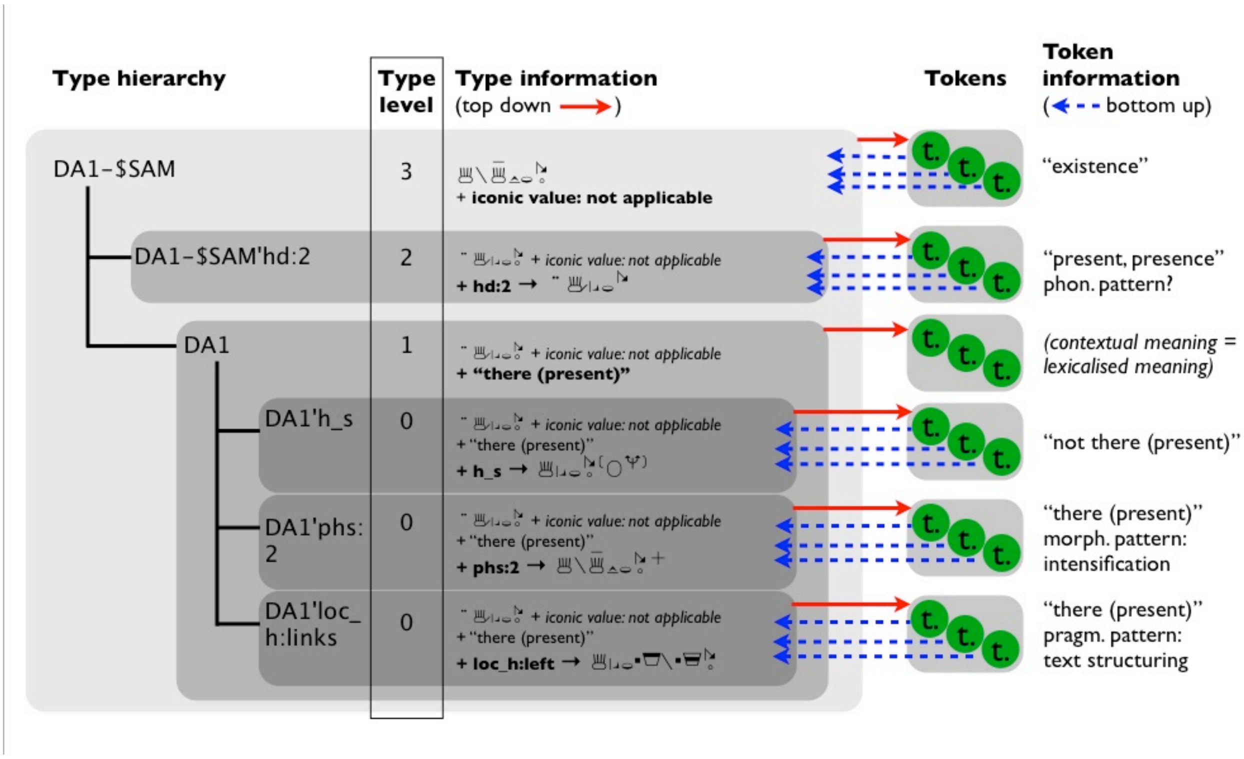

Figure 3: Expanded type hierarchy with example

In the following, the inventory of qualifiers and the handling in iLex will be described.

3. Inventory of qualifiers and feature values

As we report on work in progress, the qualifier list and the corresponding feature values are far from complete. Until now they cover some morphological patterns like inflection (location, direction, source and goal), where the form-function relation is known. Other form features like "phases" of movement (i.e. different kinds of repetition) are yet deliberately unspecified for a certain function. At the end of the lemmatisation process the distribution and frequency of these features over types and sub-types, in combination with context information, will show what phonological, morphological, and syntactic change each sign can undergo. The instantiations of qualifiers are a strong indicator for part-of-speech tagging. In addition, they will allow us to move from a citation-form lexical database closer to full-form.

The following table lists the qualifiers that are already used in iLex to specify a type or a sub-type. For most of the qualifiers feature values are pre-defined and implemented as vocabularies. The aim of these values is to get a coarser division of token form characteristics than it would be by transcribing the token form, e.g. using HamNoSys. Closed vocabularies also have values for tokens that need to be discussed (*unclear*) or that are candidates for a new feature value to be added (*leftover*). When creating a new qualified type the qualifier code is added automatically to the type/sub-type gloss. This telling gloss suffix makes it easy to understand the modification of the sign form (see explanation to figure 5 below). In general, the form of a qualified type, like the citation form of the type, is transcribed in HamNoSys. For some qualifiers this HamNoSys string can also be adapted automatically. To code tokens that show more than one form feature qualifiers can be combined so that the vocabularies can be kept concise (see below table 1, figure 4, and 5).

Except *head shaking*[12], all features refer to manual parameters. Feature values of *number of hands* include one- and two-handed and should only be used for symmetrical signs. Two-handed symmetrical signs are further qualified by movement (reversed, anti-cyclic).

The modification of the citation form by adding the nondominant hand so that a one-handed sign becomes an asymmetrical two-handed sign, or dropping the nondominant hand of an asymmetrical sign is divided into four qualifiers: *hold* and *hold resume* to identify sequences where the nondominant hand is part of a previous sign, *continued* to indicate that starting from a simultaneous sign construction the articulation of the sign of the nondominant hand is stretched over two or more signs in the dominant hand, and *base* when the nondominant hand is added (weak prop) and its iconic value can be analysed as a substitutive or manipulative image producing technique (see Langer 2005; König et al. 2008). Therefore, several feature values are provided for the basic and frequently used handshapes B-hand,

[12] This qualifier is used when the meaning of a sign is negated only by headshaking. The negation is limited to the sign and does not affect the whole phrase. It is not used when the negation is expressed by a manual form feature like *alpha negation*. In these cases headshake is additional and will be annotated in the gesture tier.

C-hand, and fist. In contrast, weak hand drop is coded by "'bas:none".

Spatial modification is divided into movement and location. Location in turn is grouped into the use of a real location of a referent or an action (qualifier *location horizontal, sagittal, vertical*) and the metaphorical location e.g. when the signing space is used to contrast two topics and therefore signs are located either on the left or on the right side (Johnston, 1991: 10-11). The text structuring and pragmatic function of location will be coded separately (qualifier *location text structure horizontal, sagittal, vertical*).[13]

Covering …	**Qualifier**	**Code**	**Feature**	**Vocabulary**
Number of hands	number of hands	hd	Number of hands	closed
Nondominant hand	hold	h	hold	closed
	hold resume	hres		
	continued	cont		
	base	bas	base	closed
Location	location horizontal	loc_h	location horizontal	closed
	location sagittal	loc_s	location sagittal	closed
	location vertical	loc_v	location vertical	closed
	location on body	bodyloc	location on body	open
	location text structure horizontal	loc_ts_h	location ts horizontal	closed
	location text structure sagittal	loc_ts_s	location ts sagittal	closed
	location text structure vertical	loc_ts_v	location ts vertical	closed
Movement …	source	src_h	source+goal	closed
	source	src_v	location ts vertical	closed
	goal	gol_h	source+goal	closed
	goal	gol_v	location ts vertical	closed
Movement	phases	phs	phases	closed
	reverse	rev		
	offset direction	offdir	offset direction	closed
	alpha negation	alph		
Handshape	assimilation	assim		
Nonmanuals	head shaking	h_s		
Sub-system manual alphabet	fa one-handed	1	alphabet	open
	fa two-handed	2	alphabet	open
	fa tracing	sk	alphabet	open
	fa tracing on hand	skh	alphabet	open
	fa ligature	lig		
Sub-system numbers	quantity	q	quantity	closed
	number	n	number	closed
	m out of n	of	quantity	closed
	detour	dt		

Table 1: Inventory of qualifiers

[13] Of course, metaphorical use of location to express temporal aspects like locating signs on a horizontal, vertical, or sagittal time line will also be covered by a separate location qualifier.

Real location is coded by three features following the three dimensions in space. *Location horizontal* has five values (head, throat, upper chest, chest, belly), *location vertical* has also five values (left, diagonal left, front, diagonal right, right), and *location sagittal* has three values (close, near, far). These values can be added using a diagram that displays the spatial relations. In the same way the features for the text structuring use of location follow the three dimensions, but their vocabularies are smaller (high and low for the horizontal plane, left and right for the vertical plane, and front and back for the sagittal plane). Another qualifier helps to code all tokens that were modified by a specific body location (*location on body*). Due to anatomical facts and the more or less specific use of body parts the corresponding feature list can be quite large and is implemented as an open vocabulary.

For movement modification, following the well-known inflection of directional verbs, the qualifiers *source* and *goal* are used. Each of them are coded by two features with respect to the horizontal and vertical dimension (*source h, source v, goal h, goal v*). The horizontal plane is divided into left, right, middle, and signer, whereas for the vertical plane the vocabulary of the feature *location text structure vertical* can be used. In addition the sweeping and the zigzag movement that are morphological features of the distributional aspect of some verbs are coded separately. These movement modifications also involve change in palm orientation and/or direction of the fingers.

The qualifier *phases* covers repetition of movement. It turned out that for signs which already have repeated or repeated circular movement in their citation forms it was not sufficient only to label one up to three repetitions. Further on reversed movement is covered by a separate qualifier (*reverse*) just as repeated movement with simultaneous change in direction (*offset direction*). The combination of these three features allows for coding different movement patterns. Finally, in DGS some verbs can be modified by changing their movement as if the hand would trace the Greek letter alpha (α) in the air. The semantic function of this movement pattern is negation and will be covered by the qualifier *alpha negation*.

In order to get a clear picture of the variety of signs for manual alphabet and numbers, we have defined several qualifiers. For the manual alphabet (*alphabet*) we differentiate between one-handed (*fa one-handed*) and two-handed signs (*fa two-handed*), tracing signs in the air (*fa tracing*) or on the nondominant hand (*fa tracing on hand*). If the fingerspelled letters are connected by a slight movement, this feature can be added by choosing the qualifier fa ligature. Signs for numbers are covered by the qualifier *number*. For number incorporation the qualifier *quantity* is provided. It can be combined with the feature *detour* for movements with an additional slightly curved path. If the nondominant hand shows a quantity from two to five and serves as a list for the dominant hand pointing to any finger but not the one representing the maximal number, a separate feature *m out of n* is used.[14]

4. Handling: Attributing qualifiers and data retrieval

The main task in the process of lemmatisation is to find the right type a token should be matched to. In comparing type citation form and token form the transcriber should be able to document his findings in a quick and easy way. The simplest way is to mark the token that there is a form difference. In the second pass of lemma revision where all tokens of one type are checked, this piece of information is relevant. A more efficient way for lemma revision is to note the salient feature in which the token is different from the citation form, e.g. in HamNoSys. Sorting all tokens of one type according to these annotations helps to get a quick overview and to find tokens with the same kind of modification or variation. This is what we did before implementing qualifiers in iLex and what we still will do when the qualifier and its feature values do not cover all the token features. So lemma revision is not a singular pass, but has to be done several times.

The reason why we moved to qualifiers and qualified types is a practical one. Instead of annotating several times the same salient form feature to tokens and in a second pass grouping these tokens together, the transcriber can do this in the first instance of the lemmatising process. After linking the token tag to a type, e.g. by dragging an item from the type list and dropping it over the token tier (Konrad & Langer, 2009), one can use the context menu (right mouse click) to display type, sub-types, and qualified types of the chosen item. Figure 4 shows the type and all the qualified types of the sub-type glossed as DA1.[15]

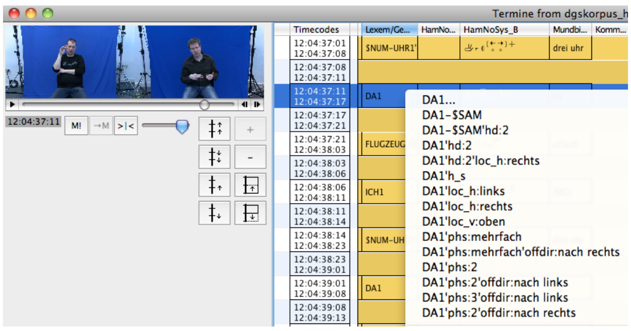

Figure 4: Context menu displaying type and qualified types of the sub-type DA1

If the token form does not match any of the already existing qualified types, the transcriber can either annotate the form difference or create a new qualified type by choosing "DA…" on the head of the listed items.

[14] See Liddell 2003, Johnston 2011: list buoys.

[15] To see all the sub-types of the type glossed as DA1-$SAM, one first links the token to the type and then re-uses the context menu.

Note that even though the database creates new type entries, this procedure is always bottom-up. Qualified types are derived from already existing types. The type hierarchy is used here to simplify token annotation. As with lexicon building, in the beginning one has to invest more in creating new qualified types, but once they are there, they will be listed and can easily be picked to spot token form features.
Instead of running searches over a multitude of tiers to find tokens that match the search criteria, the type hierarchy allows for having all spotted token form features of one type at a glance. Figure 5 shows the existing qualified types of the sub-type DA1. There are tokens with headshake ('h_s), two-handed articulation ('hd:2), different location ('loc_h:links (left)), combination of two-handed and location feature ('hd:2'loc_h:rechts (right)), repetition ('phs:2, 'phs:mehrfach (multiple)), and combination of repeated movement with simultaneous change in direction (e.g. 'phs:2'offdir:-nach links (to the left)). Double-clicking on one of the items opens the qualified type entry where all tokens are displayed.[16]

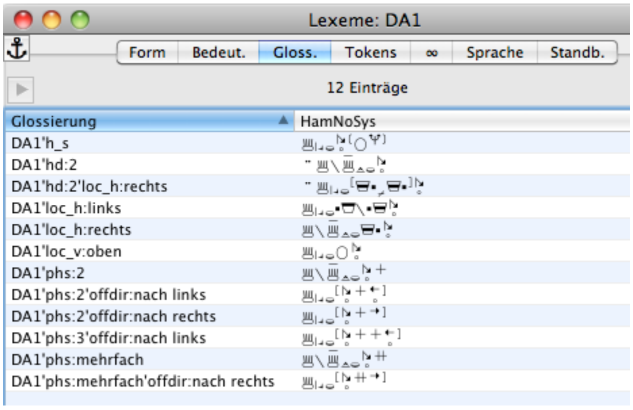

Figure 5: Qualified types of sub-type DA1

5. Acknowledgements

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

6. References

Crasborn, O., Zwitserlood, I. (2008). *Annotation of the video data in the Corpus NGT*. Nijmegen: Radboud University. [Online available: http://www.ru.nl/aspx/download.aspx?File=/contents/pages/515436/corpusngt_annotationconventions.pdf].

Hanke, T. (2002): iLex. A tool for Sign Language Lexicography and Corpus Analysis. In González Rodriguez, M. & Paz Suarez Araujo, C. (Eds.), *Proceedings of the Third International Conference on Language Resources and Evaluation. Las Palmas de Gran Canaria, Spain*; Vol. III. Paris: ELRA, pp. 923-926.

Hanke, T., Storz, J. (2008). iLex - A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In Crasborn, O., Efthimiou, E., Hanke, T., Thoutenhoofd, E.D., Zwitserlood, I. (Eds.), *LREC 2008 Workshop Proceedings. W 25: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora.* Paris: ELRA, pp. 64-67. [Online available: http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf].

Hulst, H. van der, Channon, R. (2010). Notation systems. In Brentari, D. (Ed.), *Sign Languages.* Cambridge University Press, pp. 151-172.

Johnston, T. (1991). Transcription and Glossing of Sign Language Texts: Examples from AUSLAN (Australian Sign Language). In *International Journal of Sign Linguistics*, 2(1), pp. 3-28.

Johnston, T. (2009). Creating a Corpus of Auslan within an Australian National Corpus. In Haugh, M., Burridge, K., Mulder, J., Peters, P. (Eds.), *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*. Somerville, MA: Cascadilla Proceedings Project, pp. 87-95.

Johnston, T. (2010a). From archive to corpus: Transcription and annotation in the creation of signed language corpora. In *International Journal of Corpus Linguistics*, 15(1), pp. 106-131.

Johnston, T. (2010b). Adding value to, and extracting of value from, a signed language corpus through secondary processing: implications for annotation schemas and corpus creation. In Dreuw, P., Efthimiou, E., Hanke, T., Johnston, T., Martínez Ruiz, G., Schembri, A. (Eds.), *LREC 2010 Workshop Proceedings. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pp. 137-142. [Online available: http://www.lrec-conf.org/proceedings/lrec2010/workshops/W13.pdf].

Johnston, T. (2011). *Auslan Corpus Annotation Guidelines.* [Last updated 30 November 2011. Online available: http://www.auslan.org.au/video/upload/attachments/AuslanCorpusAnnotationGuidelines30November2011.pdf].

König, S., Konrad, R., Langer, G., Nishio, R. (2010). How Much Top-Down and Bottom-Up do We Need to Build a Lemmatized Corpus? Poster presented at the *Theoretical Issues in Sign Language Research Conference (TISLR 10), Sept 30 - Oct 2, 2010 at Purdue University, Indiana, USA*. [Online available: http://www.purdue.edu/tislr10/pdfs/Konig%20Konrad%20Langer%20Nishio.pdf].

König, S., Konrad, R., Langer, G. (2008). What's in a sign? Theoretical Lessions from Practical Sign Language Lexicography. In J. Quer, (Ed.), *Signs of the Time. Selected Papers from TISLR 2004*. Hamburg: Signum, pp. 379-404.

Konrad, R. (2011a). *Die Erstellung von Fachgebärdenlexika am Institut für Deutsche Gebärdensprache (IDGS) der Universität Hamburg (1993-2010).*

[16] By changing the list statement more information, e.g. the number of tokens in this view, can be displayed.

Corrected Version. [Online available: http://www.sign-lang.uni-hamburg.de/projekte/mfl/konrad_2011_fachgeblexika.pdf]

Konrad, R. (2011b). *Die lexikalische Struktur der Deutschen Gebärdensprache im Spiegel empirischer Fachgebärdenlexikographie. Zur Integration der Ikonizität in ein korpusbasiertes Lexikonmodell.* Tübingen: Narr.

Konrad, R., Langer, G. (2009). Synergies between transcription and lexical database building: The case of German Sign Language (DGS). In Mahlberg, M. / González-Díaz, V., Smith, C. (Eds.), *Proceedings of the Corpus Linguistics Conference (CL2009). University of Liverpool, UK, 20-23 July 2009.* [Online available: http://ucrel.lancs.ac.uk/publications/cl2009/346_FullPaper.doc].

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

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