"Non-tokens": When Tokens Should not Count as Evidence of Sign Use

Gabriele Langer, Thomas Hanke, Reiner Konrad, Susanne König
Institute of German Sign Language (IDGS), University of Hamburg
E-mail: {gabriele.langer, thomas.hanke, reiner.konrad, susanne.koenig}@sign-lang.uni-hamburg.de

Abstract

Lemmatised corpora consist of tokens as instantiations of signs (types). Tokens usually count as evidences of the signs' use. Frequency of tokens is an important criterion for the lexical status of a sign. In combination with metadata on the signers' sociolinguistic backgrounds such as age, gender, and origin these tokens can also be analysed for regional and sociolinguistic variation. However, corpora may also contain instances of sign use that do not reflect the sign use of the person uttering them. This is particularly true for metalinguistic discussions of signs, malformed signing and slips of the hand as well as other phenomena such as copying/repeating signs of the interlocutors or from stimulus material. In our presentation we list and discuss different kinds of sign use (tokens) that should either not be counted as proof of a sign type at all or at least not as evidence of regular sign use by that particular person. Examples of these "non-tokens" are either taken from the DGS Corpus or from uploaded video answers of the *DGS Feedback*. We also discuss some implications on how to annotate these cases.

Keywords: sign language corpus, metalinguistic signing, lemmatisation, empirical status of tokens, annotation

1. Introduction

In a sign language (SL) corpus continuous stretches of signing are made accessible permanently (recorded on film) and lemmatised as running text from end to end without gaps. By this signs, linguistic phenomena and patterns can be analysed in context. Each individual manual activity that can be interpreted as a sign is tagged and has to be dealt with in annotation (lemmatisation). Occurrences of signs in signed utterances are usually taken as evidence for a sign's existence and use. The advantage of having a large corpus is that ideally one can have many tokens from different signers as evidence for one particular sign. Metadata of the signers (e.g. their regional rooting, SL acquisition age, age, gender etc.) can be used to analyse the distribution of a sign and its use or of other linguistic phenomena and patterns across different regions and groups of signers (e.g. see Langer et al. 2014).

However, some signs (tokens) in a corpus are not evidence of regular sign use by that particular person, for example when citing others in a metalinguistic comment on signs, malformed sign use or copying signs from the interlocutor or from stimulus material. Other (manual) activities appearing to be signs are not signs at all or signs from other sign languages. Depending on their number, it seems reasonable and useful to classify and tag these kinds of tokens in order to be able to include or exclude them from analyses according to the particular research question.

1.1 Background

The DGS Corpus Project is a long-term project with two major goals: building a lemmatised and annotated reference corpus of German Sign Language (DGS) and compiling a general DGS Dictionary on the basis of this corpus data. A major focus is on how to annotate a corpus in a way that best serves as a multi-purpose language resource for all kinds of research questions. At the same time there is also a strong lexicographic interest focusing on individual signs, their forms, variants, use, and distribution.

1.2 Sources of Examples

Examples for this paper are drawn from two quite different sources of data: annotated corpus material including free conversation as well as highly metalinguistic and context-dependent signed comments on stimuli presented in an online survey (called *DGS Feedback*) and recorded with a webcam.

1.2.1 DGS Corpus Material

The DGS Corpus contains over 1150 hours of footage and about 615 hours of filmed signing. A considerable portion of this data is being lemmatised, annotated and made accessible to become a reference corpus for German Sign Language (DGS) [1]. During the data collection phase (2010-2012), 330 signers where filmed in 12 different locations across Germany. The filming session for each pair of informants (signers) took place at one day and lasted about 6-7 hours (including breaks). In order to encourage the signers to keep interacting and talking to each other, a third signer (a trained deaf moderator) led through the sessions and through up to 20 different tasks. All tasks with exception of elicitation of isolated signs (see 2.1.1) were designed to record fluent natural (or near-natural) signing in context. Some tasks were more pre-structured and staged than others, several tasks involved talking about deaf-related topics or telling about one's life and personal experiences. One task is a free conversation in absence of the moderator on anything the informants liked to sign about. (For more detailed information on data collection and technical aspects cf. Nishio et al. 2010, Hanke et al. 2010).

Some of the filmed conversations contain tokens that have to be dealt with in annotation but should not be regarded as normal sign use of this person. Some of these "non-tokens" are metalinguistic sign use and appear either spontaneously or in the conversation task *Young*

[1] A representative part of the data is published from 2015 on as a subcorpus (DGS Corpus Project, 2015-2016).

vs. old signs which was specifically designed to provoke metalinguistic utterances.

1.2.2 Answers to DGS Feedback

In the project, one intermediary step towards a corpus-based dictionary is the compilation of a preliminary basic dictionary (for more detail cf. Langer et al. 2014) In the process of compiling the basic vocabulary lemma sign candidates (including their variant forms and presumed meanings/senses) are presented in an online survey called *DGS Feedback* to be commented upon by signers. The answer options include the possibility to record a signed answer or comment via a webcam and upload it to our server. Why is this material interesting? These signed answers and comments are highly metalinguistic in content and very context-dependent in that they directly refer to the stimuli presented on the web page – often by a form of citation of or reference to the stimulus sign. On the one hand, the answers contain valuable information on sign use, and on the other hand, they are densely packed with all kinds of tokens that are some sort of "non-tokens". In order to make use of this information the uploaded signed answers have to be made accessible by some kind of annotation.[2] Whether they are selectively lemmatised and annotated or continuously as running text – in both cases one has to decide which occurrences of signs qualify to count as evidence of a sign and which do not ("non-tokens"). Therefore, this material is a valuable source, a testing ground, and ideal example for the "non-token" issue discussed here.

When using webcams for recording, the signing may be adapted to the limited field of view (and speed) of the webcam and provoke other kinds of special, non-standard tokens of signs (for this aspect see 4.2.4).

2. Influence of Elicitation Task

A good corpus should cover a variety of text types showing different uses of signs including metalinguistic ones. It is common practice in SL corpus building to have data collection sessions with more or less staged communicative events[3] to record signing. For this aim different tasks can be designed that may influence the signs' use.

2.1 Narratives and Conversation

Signed narratives, conversations or discussions can contain incidences of "non-tokens" that in corpus annotation might be useful to tag separately. In normal conversation talking about signs is just one of the many topics that can come up. Depending on introductory questions or given topics – e.g. acquisition of sign language, communication in school, family or at work, experiences with hearing persons, or communication while travelling in foreign countries – the documented signed texts are more or less likely to contain metalinguistic use. In the DGS Corpus we found examples for this kind of singing in several tasks including free conversation, young vs. old signs, and elicitation of isolated signs.

2.2 Elicitation of Isolated Signs

Eliciting individual signs by asking informants what sign they use for presented concepts is a elicitation method that has been widely used in SL research, specifically in studies on lexical variation and for compiling dictionaries. It is discussed here separately because it prompts metalinguistic responses from interviewees that often include one particular kind of tokens showing signs that the informant knows of but normally does not use in a natural signing environment, except when talking about these signs (see 3.1, category (b)). Stimuli for eliciting isolated signs are e.g. a picture, a fingerspelled word, a written word, sometimes with a short contextual hint, or a combination thereof. Ideally, one stimulus should evoke a single concept to be expressed by a (lexical) sign. The spontaneously given response to this kind of stimulus normally is an isolated sign that is used by the interviewee to express the intended concept. However, answering to prompts for isolated signs is a very unnatural communication situation in which the interviewees are highly aware of its sole purpose of collecting signs as signed representations of concepts or signed equivalents of written words. In their responses, signers sometimes show not only their own sign but also other lexical variants they know of. These (lexical or phonological) variants may or may not be used by the interviewees themselves. Sometimes informants provide explicit or implicit contextual clues to indicate whether they use these additional variants themselves or not. Often they just list variants with no indication whether they use the signs actively or not. One solution to this problem is to only consider the first – presumedly most spontaneously given – answer to a stimulus and leave out the following signs in the analysis because their use or non-use is not explicitly made clear.[4]

Although untypical for corpus data (usually aiming at documenting connected natural signing) this elicitation method has been used as one of several tasks in some SL corpus data collections[5] and has also been used in the DGS Corpus.[6]

3. Kinds of "Non-Tokens"

The issue of this paper has been called "non-tokens" but actually the signs discussed are tokens, but are special with regard to their empirical status – whether they

[2] Furthermore, such a collection of answers and comments could be regarded as a corpus of a very specific (metalinguistic) form of signing that might be worth of lemmatisation and analysis in its own right. For example, it seems to be great material to examine how metalinguistic citations and references are marked (e.g. body body shifts, facial expressions). If one were to lemmatise and annotate material of this kind the issue of how to handle different kinds of "non-tokens" would be a very prominent one.

[3] On the concept of staged communicative events cf. Himmelmann 1998, 185-186).

[4] Cf. for example Stamp (2014, 5): "Many participants produced multiple examples of signs and, as a result, either the variant stated to be the sign, or if not stated, the first variant produced, was coded."

[5] ASL Corpus: Lucas et al. (2001, 40), BSL Corpus: Schembri et al. (2013, 140).

[6] The elicitation task is only a small portion of the data: With 22 hours it accounts for 3.5 per cent of the filmed signing.

should count as proof of the sign or the sign's use or not in the light of a given research question. Whether a token is a "non-token" or not is not an on-off decision but a matter of degree and perspective.
A metalinguistic token of the sign X as in "I don't use X" might be taken as evidence that the sign X exists (token status) and that the signer knows this sign, it might also be counted in a general token count for frequency of the sign, but it may be a "non-token" of sign use by this particular person as a representative of e.g. a certain region when doing a distributional analysis of where a certain sign is used.
Malformed tokens of signs produced as slips of the hand might be "non-tokens" in an investigation of normal sign form, but might be the only tokens of interest in a study on slips of the hands.
These examples show that it would be very useful to identify, classify and label tokens with various "non-token" potential in a corpus as part of the such enriched annotation. This enables researchers to include or exclude tokens of these special kinds and thus to be more precise in the use of the corpus and avoid certain groups of noise in the data they choose for their analyses. In the following sections we will discuss different types of potential "non tokens".

3.1 Metalinguistic Reference

Any kind of metalinguistic reference to individual signs does not necessarily reflect typical sign use of the signer and should be identified and labelled.

3.1.1 Metalinguistic Reference to an Existing Sign

The first type of our catalogue of "non-tokens" is a metalinguistic reference to an existing DGS sign. The reference consists of an execution of the sign that is being referred to. Such a token does not necessarily indicate that the signer would use this sign in non-metalinguistic signing. Metalinguistic references could be either:

(a) a reference to an existing DGS sign also used by the signer, or
(b) a reference to an existing DGS sign that is normally not used by the signer.

In many cases the context will clarify which one of the above cases apply. (Think for example about the following utterances: (a) "I always sign X" vs. (b) "In Bavaria they use the sign X" (signed by a non-Bavarian signer).
Tokens of the type (a) are pieces of conscious introspective information rather than an unconsciously and spontaneously produced evidence of sign use by a particular person. These tokens still could count as evidence of sign use by that person but nevertheless they should be labelled as metalinguistic reference.[7]
Tokens of the type (b) could be interpreted as a conscious introspective piece of information on the existence and use of the sign itself but it would be misleading to include tokens of this kind for example in a distributional analyses of tokens linked with the signers' individual metadata (e.g. regional rootedness for regional distribution).
In cases where it cannot be decided whether (a) or (b) applies, one should stay on the safe side and assume (b) or open up a third category c) for these unclear cases.
Consider example 1. In the free conversation task with the moderator not being present, two female informants (A and B) discuss the use of the DGS sign WOMAN1. The sign's form is iconically motivated: The B-handshape traces the breast of a woman. Some signers, especially women, dislike the sign because they feel that it is not politically correct. First B asks A for her sign for 'woman' by fingerspelling the German word. A's answer is WOMAN3 – one example for the above-mentioned category (a). Then B starts talking about the debate showing the sign WOMAN1. A knowing about the debate anticipates the sign so that both nearly synchronically make the sign WOMAN1.

Informant A	**1**	**2**	**3**	**4**
Gloss	YES1A	WOMAN1	YES1B	WOMAN 1
Mouthing		frau		frau
Translation	Yes, the sign WOMAN1.			
Comment		reference (b)		reference (b)

Example 1: Metalinguistic Reference (b)

Token 2 and 4 are examples of category (b), but a little later in the conversation there are also examples of category (c) when A repeats two other signs for 'woman' that B has shown before A's utterance (see example 2).

1	**2**	**3**	**4**	**5**	**6**
THERE-IS1A	AND2	ALSO1A	WOMAN4	WOMAN2	WOMAN4
es-gibt	auch		frau		
There are also the signs WOMAN4 and WOMAN2.					
			reference (c)	reference (c)	reference (c)

Example 2: Metalinguistic Reference (c)

Metalinguistic references of the type (b) also cover cases when signers talk about what signs other groups or individual signers use or have been using (for example old, young, hearing, or deaf signers, signers from certain regions, CODAs, interpreters, father, grandmother etc.).
In example 3, when discussing old versus young signs, informant A contrasts the sign of her grandmother with her own sign for 'banana'.

1	**2**	**3**	**4**
INDEX1	BANANA1A	I1	BANANA2
	banane		banane
She signs BANANA1A, I sign BANANA2.			
	reference (b)		reference (a)

Example 3: Metalinguistic Reference (a, b)

This kind of reference where the signer recalls a sign from memory can be distinguished from a reference (d) where the signer refers to a sign from his/her interlocutor as shown below in example 9, sign 2.[8]

[7] Not always does introspective information of one's own language use correspond with the unconscious language use in real life.

[8] Of course, signers like speakers adapt unconsciously their language use to their interlocutor. Often this remains unnoticed and wouldn't be labeled.

When signers metalinguistically refer to signs of a foreign sign language this is also a case of metalinguistic reference (e).[9]

1	2	3	4	5
ENGLAND3	INDEX1	YES1B	$GEST-HEAD-SHAKING	YES-CORRECT
englisch				yes
In England they don't sign YES1B, but YES-CORRECT.				
		token		reference (e)

Example 4: Metalinguistic Reference (a, b)

In this example, it clearly is not a case of regular use of a borrowed sign of foreign origin but a metalinguistic reference to the foreign sign normally not used by the signer. It would be useful to be able to distinguish metalinguistic reference to a foreign sign from regular use of a borrowed sign.
A special case of metalinguistic reference (f) is when signers talk about signs that they have been using in the past, but do not use anymore at the present time.

3.2 Un-Tokens (often contrastive)

A very common communicative pattern or strategy in DGS is to name the opposite or complementary fact or thing with negation – either by headshake and or facial expression alone or in addition by an explicit negation sign – followed by the fact or thing that one wants to express positively. This pattern works with contrasts between the negated and the positively expressed parts of the message. (The following made-up examples illustrate this pattern: CHEESE-neg NOT – SAUSAGE 'not cheese but sausage' or FREE-neg NOT – WORK 'not having spare-time but having to work'.) This strategy can also be used in metalinguistic signing about a non-sign vs. sign or non-use of a sign vs. use of a sign in a specific sense resulting in what we call *contrastive un-tokens* in the negated part.
In the *DGS Feedback* answers we find many examples for these kinds of un-tokens. The negated token can concern the sign as a whole ("non-sign"), a particular wrong execution of the sign ("non-form") or a particular use of the sign with regard to meaning ("non-use").
In example 5, taken form the *DGS Feedback* answers, the signer refers to a presented signed question item with his own sign for cheese (1, regular token), then copies the stimulus sign (3, un-token) marking it with a headshake and the following sign NOT1 in the sense of 'not used' and then repeats his own sign (6) as information of what sign he uses.

1	2	3	4	5	6
CHEESE2B	I1	CHEESE2A	NOT1	I1	CHEESE2B
		headshake			
käse					käse
My sign for 'cheese' is CHEESE2B, not CHEESE2A.					
regular token / reference to item		un-token / copy of stimulus			reference (a)

Example 5: Un-Token

[9] This has to be differentiated from cases where signers use signs of foreign origin non-metalinguistically in their normal signing – those cases could be either indications of borrowing or instances of code mixing.

Not all un-tokens are contrastive tokens, in principle tokens could also be negated without being followed by their positive counterpart. The point here is that the token in question is deliberately and clearly negated or declared as wrong or not being used to indicate that this in not the correct sign, sign form, or sign use.
In the following example 6 informants A and B are talking about their experiences as being deaf, in particular about their relation to and communication with their (hearing) parents. A cites a malformed sign her mother uses for 'important' – with U- instead of V-handshape – and her attempt to teach her mother the correct form IMPORTANT1:

1	2	3	…
MOTHER 2	SIGN1	IMPORTANT1*	
mutter		wichtig	
My mother signs IMPORTANT1*.			
		un-token (malformed)	

4	5	6	7	8	9
SIGN1	NORMAL1	SIGN1	IMPOR-TANT1	IN-DEX1	IMPOR-TANT1*
	normal		wichtig		wichtig
Normally one signs IMPORTANT1, but she signs IMPORTANT1*.					
			reference (a)		un-token (malformed)

Example 6: Un-Token

Example 7 is a *DGS-Feedback* comment to the stimulus WOMAN1/frau with the meaning of 'woman, female'. The stimulus sign (1) is copied as reference to the stimulus and then marked as un-token by the following comment ("I don't use this."). This is an example of an un-token with regard to meaning because the signer uses this sign, but in combination with another mouthing only to express the meaning 'breast' (example 7, sign 6).

1	2	3	4	5	6	7
WOMAN1	SIGN1A	I1	USE1	NOT1	WOMAN1	YES1A
frau		ich	benutze	nicht	busen	ja
I don't use the sign WOMAN1 for 'woman', but for 'breast'.						
copy of stimulus / un-token					metalinguistic reference (a)	

Example 7: Un-Token and Metalinguistic Reference

3.3 Copy of Stimulus (Reference)

When signed stimuli are used for prompting signers to comment about these signs – as it is done in the *DGS Feedback* – signers in their metalinguistic comments or answers often refer to these prompts either by copying the stimulus sign – here called *copy of stimulus* – or by using their own sign (see example 5, sign 1) for the same concept. Tokens that copy the presented stimulus sign for the sole purpose of referring to them are tokens that cannot count as evidence for a sign's existence or use, because they are pre-specified by the context.[10]

[10] It does not matter here whether the stimulus is a real sign or a non-existing, made-up sign form or whether an existing stimulus sign is presented correctly (e.g. with a correct meaning) or incorrectly (e.g. with an incorrect, unusual or unknown meaning). In the *DGS Feedback* we present non-existent made-up signs as distractors along with existing signs.

Often the copy of a stimulus is not only used to refer to the presented item but is also commented upon which either makes the copy a un-token sign 3 in example 5 and sign 1 in example 7 are not only un-tokens but also at the same time copies of the stimulus.

4. Non-Signs of the SL

Some manual activities within the continuous sign stream look like signs and are part of the utterances but in fact are not part of the respective target sign language.[11] However, such tokens may be part of the utterances, are needed for a correct interpretation and cannot be left out without loosing content of the utterance. This means that they cannot be skipped for annotation.[12]
For a corpus on DGS, all tokens that are not DGS sign use should be tagged. We propose the following types of non-signs or non-DGS signing respectively.

4.1 Gestures of Hearing Persons

Sometimes signers talk about gestures used by hearing persons either in communication attempts with deaf or among themselves. In the following example informant A tells informant B about her colleague who had learned the gesture for 'tasty' during his trip to Italy:

1	2	3	4	5
TASTY3	ITALY1	BELONG-TO1	SAY1	$GEST-TASTY
lecker	italien			lecker
In Italy this gesture means 'tasty'.				
regular token				cited gesture

Example 8: Copy of Gesture

In the course of the conversation, both signers use the gesture and also use their own signs for 'tasty'. Informant A marks the difference between her own sign TASTY3 and B's sign TASTY1:

1	2	3	4	5
INDEX1	TASTY1	ICH1	TASTY*	TASTY3
	lecker		lecker	lecker
You sign TASTY1, I sign TASTY*, TASTY3.				
	reference (d)		slip of the hand	reference (a)

Example 9: Metalinguistic Reference (a, b)

Interestingly, just before showing her own sign TASTY3 (5), A produces a mixed form (4) with the location of TASTY3 (belly) but the circling movement of TASTY1 (instead of repeated tapping). It is a malformed metalinguistic reference and interpreted as a slip of the hand (see below 4.2.2) that is immediately followed by a self-correction (5).

4.2 Malformed or Deviated Sign Forms

The next group consists of tokens that are in some way malformed or have deviated sign forms. They appear in normal (i.e. non-metalinguistic) signing contexts. Sometimes these tokens slip by unnoticed by the signer (cf. sign 5 in example 9), sometimes they are noticed and self-corrected.

4.2.1. False Starts and Aborted Tokens

Some tokens are instances where the signer starts to produce the sign and the hands move to the beginning of the sign but do not finish executing the sign completely. Sometimes one can guess at the intended sign, sometimes not.
Unfinished signs can be false starts or aborted signs. Often false starts are indicated by a facial expression or a subsequent headshake when the signer becomes aware of the near-mistake and produces a correction. Annotation guidelines have to specify how to treat false starts and aborted tokens and whether to tag them at all. If they are tagged, they should be labelled in a specific way in order to search for them separately or to exclude them from token counts and analyses.

4.2.2. Slips of the Hand

Sometimes signers accidently use the wrong sign or execute a sign with a wrong parameter. These instances are known as slips of the hand. They can pass by unnoticed or they can be self-corrected by the signer. Sign 4 in Example 9 is an example of slip of the hand followed by the correct sign.
While slips of the hand are an interesting topic of investigation in their own right (e.g. cf. Leuninger et al. 2004), for most other analyses one would rather exclude them from all counts and therefore need to label them accordingly.

4.2.3. Trying out or Searching for Signs

Sometimes in conversation but even more so in elicitation tasks for single signs, a signer is searching for the right sign or sign form he/she wants to use – visibly thinking "out loud" and in the process trying out several different signs or slightly different sign forms before arriving at the searched-for sign. These forms should not be regarded as tokens in their own right and be labelled accordingly.
In the conversation task *Young vs. old signs* informant A and B (from example 3) try to remember an old sign for 'parents'. B first shows a form like scratching her cheek with the fingertips of her 4-bent-handshape. Immediately, B reacts in raising her arm and wiggling with her index and middle finger. While B is in doubt and shows her own sign for 'parents', A continues to try out, moving both fingers to her cheek, then making small movements downwards while bending the fingers slightly as B did with four fingers. B copies the sign from A and shows a negative facial expression while A confirms that she at least knows this sign.

4.2.4. Adapted Tokens (Limited View of Webcam)

In cases where the camera field of view is limited to a small window – for example when recording with a webcam – signers sometimes subconsciously or consciously adapt their signing to fit into the window. They may for example change the place of articulation. In one feedback answer the signer showed his sign for 'hunger' twice: The first token of HUNGER was executed at the stomach – the normal place of articulation for this sign –

[11] As opposed to other manual activities that are not part of the linguistic utterance such as scratching one's nose, pushing one's hair back, rubbing or tapping nervously one's fingers or shaking the hand to loosen tense muscles. These manual activities are generally not annotated.
[12] As one would probably do in a selective spot transcription coding only tokens of interest, leaving out all others.

and therefore almost completely outside the view of the camera, the second token of HUNGER was executed at the chest – wrong place of articulation – to show handshape and the movement within the view of the camera. The second token of HUNGER is an example for an adapted token. Another example was the answer to the stimulus 'cheese' already discussed in example 5. The answer continued after a short break with another token of the sign CHEESE2B but this time adapted in form (both orientations changed in relations to each other) to show the handshapes clearly in the view of the camera and make sure that the handshape is visible from the front. These adapted sign forms should not be treated as normal instances of the sign with regard to form but should be labelled as adapted or annotated as deviant forms when lemmatising.

5. Consequences for Annotation

The various examples in this paper show that some tokens in running signed texts should not qualify as evidence for a sign's existence or for the use of a particular sign by the particular signer. It would be wise to label cases that are coming to one's attention accordingly. This allows for conscious decisions on whether to include or exclude tokens of certain kinds from analyses. One could argue that the percentage of non-tokens is not large enough to worry about them but our impression is that the percentage might be much larger than originally thought. SL corpora are much smaller than written text corpora and rely on only a relatively small number of individual signers. Therefore, even a single discussion like that described in example 1 and 2 can influence or distort results of analyses when not labelled properly and "non-tokens" not being excluded from certain kinds of analyses.

For the time being, we annotate all cases of "non-tokens" with metadata to the token tags, using an open vocabulary corresponding to the categories presented in sections 3 and 4. The vocabulary also contains some grouped values in order not to complicate the annotation too much for the first-pass annotators in cases where a categorisation is not straightforward. This approach is a preliminary version only. While it allows us to keep track of the cases already identified and to do some analysis, it does not yet allow iLex to automatically integrate the "non-token" flags into account when computing token counts. For this, a more sophisticated solution needs to be implemented once we have enough evidence that the current categorisation is both stable and manageable for the annotators.

6. 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.

7. Bibliographical References

DGS Corpus Project (2015-2016): Subcorpus. Available at: http://meine-dgs.de.

Hanke, T., Storz, J., Wagner, S (2010). iLex: Handling Multi-Camera Recordings. In *Workshop Proceedings. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies Language Resources and Evaluation Conference (LREC) Valletta, Malta, May 2010,* pp. 110--111.
[URL: http://www.lrec-conf.org/proceedings/lrec2010/workshops/W13.pdf].

Himmelmann, N. (1998). Documentary and descriptive linguistics. *Linguistics,* 36, pp. 161--195.

Langer, G., König, S., Matthes, S. (2014). Compiling a Basic Vocabulary for German Sign Language (DGS) – lexicographic issues with a focus on word senses. In A. Abel, Ch. Vettori and N. Ralli (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus, July 15-19 2014 in Bolzano/Bozen – Italy*, pp. 767--786.
[URL: http://euralex2014.eurac.edu/en/callforpapers/Documents/EURALEX_Part_2.pdf].

Langer, G., König, S., Matthes, S., Groß, N., Hanke, T. (2016). Variation of DGS lexical items. What sign language lexicography can gain from a mixed method approach: Corpus data supplemented by crowd sourcing. Poster presented at the *Theoretical Issues in Sign Language Research (TISLR) 12 Conference, Jan 4-7, 2016 in Melbourne, Australia*.
[URL http://www.tislr12.org/posters/; no. 117].

Leuninger, H., Hohenberger, A., Waleschkowski, E., Menges, E., Happ, D. (2004). The impact of modality on language production: Evidence from slips of the tongue and hand. In T. Pechmann, C. Habel (eds.). *Multidisciplinary approaches to language production*. Berlin: de Gruyter, pp. 219--278.

Lucas, C., Bayley, R., Clayton, V. (2001). *Sociolinguistic Variation in American Sign Language.* Washington, DC: Gallaudet University Press.

Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke, T., Rathmann, C. (2010). Elicitation Methods in the DGS (German Sign Language) Corpus Project. In *Workshop Proceedings. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies Language Resources and Evaluation Conference (LREC) Valletta, Malta, May 2010,* pp. 178--185. [http://www.lrec-conf.org/proceedings/lrec2010/workshops/W13.pdf].

Schembri A., Fenlon J., Rentelis R., Reynolds S., Cormier K. (2013). Building the British Sign Language Corpus. *Language Documentation and Conservation* 7, pp. 136--154.

Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., Cormier, K. (2014). Lexical Variation and Change in British Sign Language. *PLoS ONE* 9(4): 094053, pp. 1--14.
[doi: 10.1371/jpournal.pone.0094053].