

# Linking Lexical and Corpus Data for Sign Languages: NGT Signbank and the Corpus NGT

**Onno Crasborn, Richard Bank, Inge Zwitterlood, Els van der Kooij, Anique Schüller, Ellen Ormel, Ellen Nauta, Merel van Zuilen, Frouke van Winsum, Johan Ros**

Radboud University, Centre for Language Studies  
PO Box 9103, NL-6500 HD Nijmegen, The Netherlands

E-mail: {o.crasborn, r.bank, i.zwitterlood, e.vanderkooij, a.schuller, e.ormel, e.nauta, m.vanzuilen, f.vanwinsum, j.ros}@let.ru.nl

## Abstract

How can lexical resources for sign languages be integrated with corpus annotations? We answer this question by discussing an increasingly frequent scenario for sign language resources, where the lexical data are stored in an online lexical database that may also serve as a sign language dictionary, while the annotation data are offline files in the ELAN Annotation Format (EAF). There is by now broad consensus on the need for ID-glosses in corpus annotation, which in turn requires having at least a list of ID-glosses with a description of the phonological form and meaning of the signs. There is less of a consensus on standards for glossing, on practices of sign lemmatisation, and on the types of information that need to be stored in the lexical database. This paper contributes to the establishment of standards for sign language resources by discussing how two data resources for Sign Language of the Netherlands (NGT) are currently being integrated, using the ELAN annotation software for corpus annotation and an adaptation of the Auslan Signbank software as a lexical database. We discuss some of the present relations between two large NGT data sets, and outline some future developments that are foreseen.

**Keywords:** sign language resources, multimodal annotation, Signbank, lexical database ELAN annotation software

## 1. Introduction

How can lexical resources for sign languages be integrated with annotated video corpora? In this paper we aim to answer this question by discussing an increasingly frequent scenario for sign language resources, where the lexical data are stored in an online lexical database, while the annotation data are offline files in the ELAN Annotation Format (EAF).

Lexical databases for sign languages often originated from the purpose of creating sign language dictionaries (Johnston, 2001). These dictionaries were created in a variety of contexts, ranging from language technology or linguistics departments within academia to deaf associations. The varying demands and facilities have led to a diversity of proprietary databases and some open source solutions. A standard even for data structures in this domain is not within view. It is therefore extra important to document the existing solutions, as we do in this paper.

In terms of annotating and retrieving lexical signs for linguistic research, there is by now broad consensus on the need for ID-glosses (Johnston, 2008, 2010) in corpus annotation, which in turn requires having at least a list of ID-glosses with a description of the phonological form and meaning of the signs. There is less of a consensus on standards for glossing (Crasborn, Bank & Cormier, 2015), on practices of sign lemmatisation (Fenlon, Cormier & Schembri, 2015), and on the types of information that are to be stored in the lexical database.

This paper contributes to the establishment of standards for sign language resources by discussing how two data resources for Sign Language of the Netherlands (Nederlandse Gebarentaal; NGT) are currently being integrated, using the ELAN annotation software for corpus annotation (Wittenburg et al., 2006) and an

adaptation of the Auslan Signbank<sup>1</sup> software (Johnston, 2001, 2010) as a lexical database.

## 2. Two Existing Data Sets

This section describes first the Corpus NGT and then NGT Signbank. While not the only option, as we will discuss in the conclusion, this type of combination of data sets is getting more common in the domain of sign language resources.

### 2.1 Corpus NGT

The Corpus NGT (Crasborn & Zwitterlood, 2008; Crasborn, Zwitterlood, & Ros, 2008) is a collection of video and annotation data of 92 prelingually deaf signers, recorded in dyads, who retell video clips and picture stories and discuss issues related to deafness, deaf education and sign language. Annotation of the corpus is on-going; the latest (third) public release of Corpus NGT annotation that was published in June 2015 (Crasborn et al., 2015) contains over 145,000 glosses for the left and right hands. At present the production version of the corpus contains almost 370,000 annotations for different levels of transcription and analysis, from sentence-level translations to degree of thumb extension. The latest public release of Corpus NGT annotations can be viewed in and downloaded from The Language Archive<sup>2</sup>, and as a single compressed file from the corpus website.<sup>3</sup> The production version of the corpus is stored on a private SVN server. This allows working on offline copies of EAFs, with locally stored video files.

<sup>1</sup><http://www.auslan.org.au>

<sup>2</sup><http://hdl.handle.net/1839/00-0000-0000-0004-DF8E-6@view>

<sup>3</sup><http://www.ru.nl/corpusngtuk/methodology/annotation>

## 2.2 NGT Signbank

Earlier projects that aimed to link NGT corpus data with a lexical database (e.g. Crasborn, Hulsbosch & Sloetjes, 2012) involved the LEXUS lexical database. However, as the further development and support of LEXUS was discontinued in 2015, our pilot efforts in this direction had to be aborted. As an alternative, we chose to adopt the Signbank lexical database software. Signbank has originally been developed for Australian Sign Language<sup>4</sup> (Auslan; Johnston, 2001, 2010), and has since also been implemented for British Sign Language<sup>5</sup> (BSL; Fenlon et al., 2015), NGT<sup>6</sup> (Crasborn et al., 2014).

Versions of the Signbank software are currently also implemented for Finnish Sign Language<sup>7</sup> and American Sign Language (ASL). Given the fact that the sign language corpora developed in these countries are all annotated in ELAN, the discussion in this paper also pertains to the development of those Signbanks.

While the NGT Signbank software is open source and its development can be tracked online,<sup>8</sup> the NGT data set will not be published under an open access license for at least another two years. In this period, intensive annotation of the Corpus NGT leads to weekly additions to NGT Signbank. At the time of writing, the phonological description of the existing 3,200 entries is still being double-checked, and the recording of citation form videos is almost completed.

## 3. Existing Relations between Data Sets

This section describes two types of relationships between corpus and lexical database that are already implemented, while section four will focus on some further interactions between the data sets that will make exploitation of the data richer and easier. All interactions are visualised in Figure 1.

### 3.1 The Lexical Database as a Vocabulary of Gloss Types for Annotation

To facilitate video annotation in ELAN, an external controlled vocabulary (ECV; Crasborn, Hulsbosch & Sloetjes, 2012) is used. An ECV contains the full list of ID-glosses in Signbank, to label lexical signs with, as well as phonological information about those signs (e.g. handshapes, location, movement direction) and Dutch translation equivalents that serve to clarify their meaning. When deciding on an annotation, the annotator chooses an entry from the ECV to be included in the EAF file. This facilitates decision-making and reduces the occurrence of typing errors. The ECV is centrally stored on a web server (hence the E for external), allowing for central updating of the ECV with changed or added glosses, phonological information and meaning. The ECV is automatically reloaded each time

an EAF file is opened on a local computer with an internet connection. Annotation values of the glosses are then updated, if applicable, to reflect the current information in the ECV. This ensures that the annotators always work with the latest version.

As described in Crasborn & Sloetjes (2014), starting from version 4.7.0, ELAN allows for the creation of multilingual ECVs. This means that each entry in an ECV can have multiple values, one per language. Annotations based on an ECV entry will display one of these values, depending on the language selected by the user in ELAN.

A bilingual ECV is generated directly from Signbank. A nightly server-side script generates an updated ECV, including all changes made to the Signbank database in the previous day. Users can also choose to manually update the ECV, so that any changes to the database during the annotation process immediately become available to themselves and to other annotators. As was already explained above, such changes to the ECV will not be immediately visible in open EAF files: these have to be closed and reopened.

As the production version of the Corpus NGT is stored on an SVN server, all annotations are made in local copies of the corpus, which are then committed to the central server. Consequently, any changes in gloss values in Signbank (and thus, nightly or manually, in the ECV) are not propagated to the whole corpus, but only in locally opened, saved and committed files. In order to update all the gloss values (and hence provide accurate search results when doing a corpus wide search), a nightly script (currently in beta testing) changes gloss values in the corpus on the basis of their ECV-links, and also checks whether all glosses that lack an ECV-link but whose value occurs in the ECV (and thus in Signbank) receive the proper ECV-link.

The relation between Signbank and EAF files through the ECV is currently one-way: items are added to Signbank and then displayed in ELAN. It is not yet possible to harvest new items in ELAN files and add these to Signbank, for instance, or to manually add a new item to Signbank from within the ELAN interface. Both of these options will be explored in the near future. Ideally, the ECV file would only serve as a cache for offline work, while whenever a user is online, there is a live link between Signbank and the gloss tiers in ELAN.

Currently, ID-gloss fields exist in NGT Signbank and in the ECV for Dutch and English, while the whole system is being internationalised. Both the interface and most of the data have been translated into Dutch, facilitating all team members who work with the software. Further translations can be easily made in the future. A Mandarin Chinese translation is underway.

Translation equivalents are presently only added in Dutch, and while it is in principle possible to add translation equivalents in English or another language, optimal quality would require that translation equivalents are added on the basis of the meaning of the sign, rather than by translating the Dutch translation equivalents one

<sup>4</sup><https://github.com/Signbank/Auslan-signbank>

<sup>5</sup><http://bslsignbank.ucl.ac.uk>

<sup>6</sup><http://signbank.science.ru.nl>

<sup>7</sup><https://github.com/Signbank/FinSL-signbank>

<sup>8</sup>See the Github repository at

<https://github.com/Signbank/NGT-signbank>

by one. This requires high-level knowledge of both NGT and English, which not every team member has.

NGT Signbank currently contains more than 3,200 ID-glosses. While annotating, one often encounters signs that do not yet have an ID-gloss. Several criteria are used for the introduction of new glosses in Signbank (see also Cormier et al., 2015). Basically, the phonological form of the sign has to be dissimilar from all existing forms in Signbank. However, also in case the form is similar to an existing form and the meaning of the two signs does not overlap, the sign is assigned a new ID-gloss.

of signs obtained in other ways. The first NGT dictionaries that were developed in the 1980s in the Netherlands had the explicit aim of evaluating regional variation (Schermer, Harder & Bos, 1988; KOMVA 1989). Information on the regional use came from elicitation studies (Schermer, 2012). The paper dictionaries that were published in the late 1980s and early 1990s all present information about the regions in which signs are used. While the Corpus NGT data form a broader empirical base for regional use (cf. the ideal scenario described by Schermer, Stroombergen &

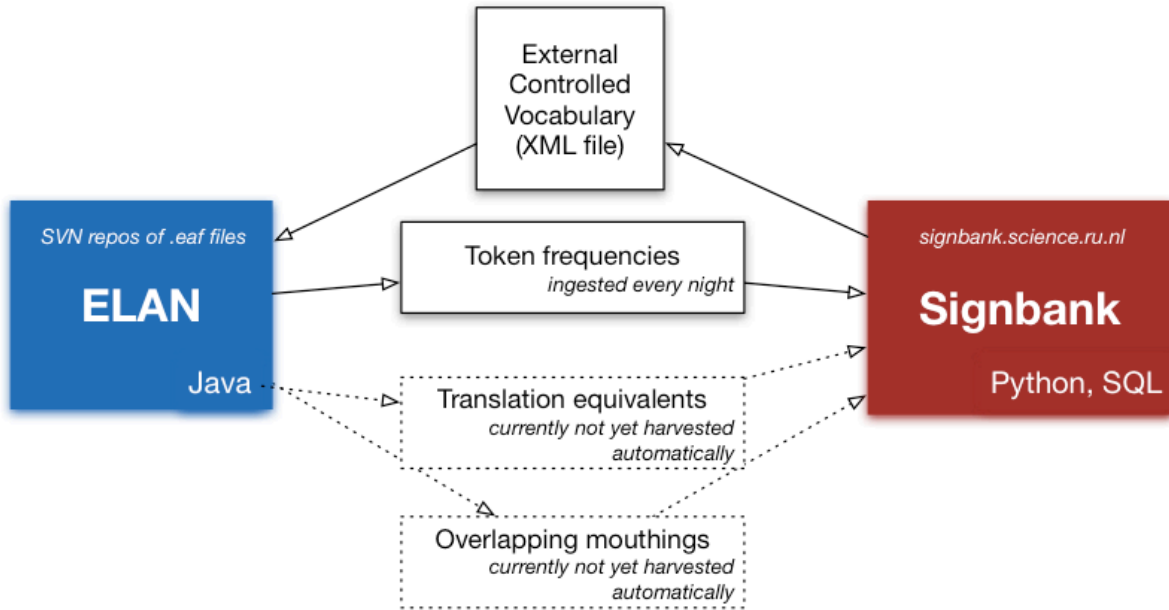


Figure 1: Overview of existing and foreseen relations between ELAN and Signbank.

### 3.2 Token Frequencies in the Lexicon

Two types of frequency data are automatically ingested in Signbank from the glosses in the Corpus NGT. First of all, there are token frequencies over the whole corpus and for each of the six regions distinguished in the metadata. These are the five traditional dialect regions in the Netherlands (Schermer, 2004), plus a rest category that includes signers with a mixed regional profile. Second, the number of signers that produce tokens of a sign is also calculated and ingested in Signbank, for the whole corpus and per region. This second type of information is particularly useful in determining how widespread the use of a sign is within a region: is it an idiosyncratic (perhaps older) form used by a single signer, or are there several people using the same sign? Together with the distinction in regions, this may help in selecting signs for inclusion in dictionaries, or even for research purposes: phonologists, for example, may not want to base their analysis on the phonology of forms that are only used by one or two signers, however frequently they may use these forms.

Token frequencies from the Corpus NGT form further empirical support for information on regional use

(Tervoort, 1984), they do face a similar limitation as the printed dictionaries of thirty years ago, in reflecting dialectal variation only at a specific moment in time. For the Corpus NGT, there was a short time window of the data recording between 2006 and 2008. Enriching token statistics by adding data from other annotated sources in the future, whether newly recorded or historic, can potentially provide a rich view on dialectal variation over time.

## 4. Data Set Interactions under Development

We presently foresee three types of data interactions that could be implemented fairly easily, and that would enrich Signbank on the basis of corpora.

### 4.1 Harvesting of Translation Equivalents from the Corpus

The meaning of signs in sign language dictionaries and lexical databases is typically represented in terms of a spoken language, by including translation equivalents and sometimes also translated sentences illustrating typical use of signs. NGT Signbank lists translation equivalents in Dutch, as was already mentioned in

section 3.1. At present, these translation equivalents are added based on the knowledge of annotators and researchers. The latter will often overlap with the meaning of actual uses of those signs in the corpus, but mismatches in both directions are observed: Signbank also lists translation equivalents that are not observed in the corpus, and not all possible translations of signs in the corpus are (yet) present in Signbank. These translations can be specified with each ID-gloss on a separate tier named Meaning in the corpus. While this is not done systematically for every gloss at this moment, annotators are requested to specify the meaning of a sign in each case where it does not overlap with the meaning encoded in the ID-gloss. This workflow has been adopted more than a year ago, and there remains a large set of older files (with an estimated 100,000 glosses) for which meanings have only occasionally been specified.

By harvesting the meaning annotations that are specified for ID-glosses, translation equivalents can be generated in Signbank in a corpus-based way. By including frequencies of different translation equivalents, the quality of semantic information in Signbank can be enriched. At the same time, the corpus or corpora on which sign language lexical resources are based are restricted, due to the immense time effort that the manual annotation process takes up. It is therefore desirable to store both translation equivalents based on signer intuitions and corpus-based translation equivalents.

## 4.2 Harvesting of Mouthings from the Corpus

The ubiquitous use of mouthings and their presumed role in the interpretation of NGT (Bank, 2015) calls for its systematic annotation in sign language corpora. Crasborn & Bank (2014) propose an annotation scheme for its annotation. The study of mouth actions in relation to signs continues to raise many questions. The mechanisms behind the variation found in the use of either mouthings or mouth gestures with signs, for instance, is not yet fully understood. Inclusion of corpus-based information on mouthings in the lexical database can help us to better understand the relation between manual signs and mouthings, and as for translation equivalents, frequency information on mouthings can aid in the determination of the semantics of signs.

One of the biggest challenges in the automated harvesting of mouthings, however, is temporal alignment. Mouthings do not necessarily align with the signs they accompany: they can spread over adjacent signs, or a sign can co-occur with multiple mouthings, and all the variations inbetween. This is illustrated in Figure 2. Even when a stretch of connected signing co-occurs one-on-one with corresponding mouthings, annotation alignment is necessarily noisy, due to the complexity of the phonetic signal.

Time	GlossL S1 [65]	GlossR S1 [198]	Mouth S1 [25]	Mouth_Transl.S1 [7]
00:02:20.500		3-B	'drie'	'three'
00:02:21.000		KEER	'maaf'	'times'
00:02:21.500	JAAR-B	#PER JAAR-B	'per'	'per'
00:02:22.000		NAAR HUIS-A	'jaar'	'year'
00:02:22.500			'naar'	'to'
			'huis'	'house'

Figure 2: Partial alignment (JAAR with *jaar* 'year') and non-alignment (HUIS with *naar* 'to') of manual signs and mouthings.

The solution we aim for is to list for each sign all mouthings that co-occur with that sign, including those that only partly overlap. In addition, two distinct values may be calculated and stored in relation to overlapping mouthings. First of all and most importantly, for each mouthing type, it should be calculated how often it occurs with a sign, just as for the translation equivalents discussed in the previous section. Second, the average amount of overlap of a mouthing type with a sign could be computed. This provides a way of determining whether a mouthing should be seen as co-occurring with a sign (in case of large overlap) or as coincidental (in case of small overlap, see e.g. in Figure 2 the overlap of #PER with the mouthing 'maal' that co-occurred with the previous sign but didn't really spread). Consequently, only mouthings with large overlap ratios (at least 60% overlap) should be included in the lexical database. The two numbers – frequency and overlap ratio – together provide a clear and concise measure of co-occurrence with sign types.

## 4.3 Use of Corpus Examples in the Lexical Database

A third possibility for enriching a corpus-based lexical database like Signbank would be to include information on the use of signs in their context. This can help in providing a richer view of the lexical semantics and pragmatics of signs, as well as form a solid basis for a learner dictionary in the long term (but see Hunston, 2009, on some of the complexities involved in presenting corpus data to learners).

Presenting information on signs in their context could take the form of collocations, highlighting the most frequent left and right collocates of signs. However, given that a lot of information in signed utterances is realised simultaneously, by means of features of facial expression and two-handed constructions to name but two prominent aspects, sequential collocates are less informative for sign languages than for many spoken languages (Crasborn et al. 2013). It may therefore be more informative to aim to include full sentences in a way that maximises the use of basic level annotations like the glosses of the two hands, mouth actions, and sentence level translations. A presentation like in the alignment view in ELAN (Crasborn & Sloetjes, 2014) could be ideal for this, yet may be difficult to implement. A basic list of sentence translations as used in some sign language dictionaries could be a simple start.

Both for scientific uses and for dictionary users, one may wonder, however, whether a full list of all occurrences is really desirable. Perhaps it would be better to manually or automatically select typical examples. This is clearly an area that needs further exploration.

## 5. Where to Access the Corpus Metadata?

The metadata for the Corpus NGT have been described in the IMDI standard when the Corpus NGT was first archived in 2008. IMDI files in The Language Archive can be accessed by online searches in the archive; this makes it possible to first select a set of constraints (age, gender etc.) before searching the content of the EAFs.

To date, however, there is no straightforward way to refer to the available metadata while searching in ELAN. For instance, restricting searches to certain regions or age groups is not possible. Some metadata are stored in ELAN files ‘natively’, as part of the XML specification of EAF files: the participant property and the annotator property of tiers. This enables for instance searching for sign X overlapping with mouthing Y by participant Z. However, more options are desired for many types of searches that are now impossible.

One way to tackle this problem would be to include metadata in the EAF itself, on separate tiers. This is what we plan to do as long as search options in ELAN cannot refer to information in metadata files. We aim to include one extra tier per signer, containing metadata information like age or age group, gender, region and type of video content (e.g., fable, free discussion etc.). The tiers will contain a single annotation that covers the entire length of the EAF, to allow for searches on overlapping annotations (e.g. all tokens of AMSTERDAM by signers from Amsterdam). Each annotation will need to contain a string of metadata values that are preceded by a string, letter, or number identifying the metadata field.

## 6. Conclusion

In this paper we discussed several links between Corpus NGT annotations made in ELAN and the lexical database NGT Signbank. While the implementation of the links brings along some software development particular to the design of the two tools, the nature of the information is of a more general nature and has clear linguistic motivations. Information on lexical items stored in a lexical database is needed for a proper use of ID-glosses in the annotation of manual signs in sign language corpora. The frequency data, semantics and contextual information from corpora all form important additions to a lexical database. They can ultimately lead to corpus-based dictionaries (see also Hanke, 2006 for discussion).

The scenario we describe here is of course not the only one currently in use – but there are not too many alternatives. Hanke (2002), Konrad & Langer (2008), and Hanke & Storz (2009) describe the integrated iLex environment, where type and token data as well as metadata are integrated in a single database. This solution has also been adopted in Poland and Denmark, among other countries. One advantage of the use of ELAN over iLex is independent of the interaction with Signbank, namely the possibility of offline use of a corpus. This advantage is of course becoming less prominent as ELAN documents interface more directly

with the online database Signbank. Together with the scenario described in this paper, these two seem to be the only solutions world-wide that have a substantial number of users, both in terms of the sign languages covered and the number of research groups working with them. For those corpus research groups that choose or chose ELAN as their annotation tool, the ECV/Signbank scheme described here provides a great enhancement in workflow control.

## 7. Acknowledgments

We thank Trevor Johnston and Steve Cassidy for letting us use the Auslan Signbank software as the basis for the NGT Signbank, and Wessel Stoop for his work in implementing NGT Signbank. We further acknowledge NWO grants 640.006.011 (Digging into Signs) and 360.70.500 (Form-Meaning Units).

## 8. References

- Bank, R. (2015). *The ubiquity of mouthings in NGT. A corpus study*. Utrecht: LOT.
- Crasborn, O., & Bank, R. (2014). An annotation scheme for the linguistic study of mouth actions in sign languages. In O. Crasborn, E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Hochgesang, J.H. Kristoffersen, & J. Mesch (Eds.), *Beyond the manual channel. 6th Workshop on the Representation and Processing of Sign Languages* (pp. 23-28). Paris: ELRA.
- Crasborn, O., Bank, R., & Cormier, K. (2015). Digging into Signs: Towards a gloss annotation standard for sign language corpora. Nijmegen: Radboud University & London: University College London. [http://www.ru.nl/publish/pages/723853/dis\\_annotation\\_guidelines\\_4may2015.pdf](http://www.ru.nl/publish/pages/723853/dis_annotation_guidelines_4may2015.pdf)
- Crasborn, O., Bank, R., Zwitserlood, I., Van der Kooij, E., Ormel, E., Ros, J., ... Vonk, M. (2014). NGT Signbank. Nijmegen: Radboud University, Centre for Language Studies.
- Crasborn, O., Hulsbosch, M., & Sloetjes, H. (2012). Linking Corpus NGT annotations to a lexical database using open source tools ELAN and LEXUS. In O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen, & J. Mesch (Eds.), *Interactions between Corpus and Lexicon. Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages. Language Resources and Evaluation Conference (LREC)* (pp. 19–22).
- Crasborn, Onno, Hulsbosch, Micha, Lampen, Lari, & Sloetjes, Han. (2013). *New Multilayer Concordance Functions in ELAN and TROVA*. Paper presented at TiGeR 2013: Tilburg Gesture Research Meeting, Tilburg, the Netherlands.
- Crasborn, O., & Sloetjes, H. (2014). Improving the exploitation of linguistic annotations in ELAN. In ELRA (Ed.), *LREC 2014: 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland* (pp. 3604–3608). Paris: ELRA.
- Crasborn, O., & Zwitserlood, I. (2008). Annotation of the video data in the Corpus NGT. Nijmegen:

- Radboud University, Centre for Language Studies. [http://www.ru.nl/publish/pages/527859/corpus\\_ngt\\_annotationconventions.pdf](http://www.ru.nl/publish/pages/527859/corpus_ngt_annotationconventions.pdf)
- Crasborn, O., Zwitserlood, I., & Ros, J. (2008). Corpus NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands (Video corpus). Nijmegen: Radboud University, Centre for Language Studies
- Crasborn, O., Zwitserlood, I., Van der Kooij, E., Sáfár, A., Ros, J., Nauta, E., ... De Meijer, A. (2015). Corpus NGT gloss annotations, 3rd release. Nijmegen: Radboud University, Centre for Language Studies. <http://www.ru.nl/corpusngtuk/methodology/annotation>
- Fenlon, J., Cormier, K., & Schembri, A. (2015). Building BSL Signbank□: The lemma dilemma revisited. *International Journal of Lexicography*, 28(2), 169–206.
- Hanke, T. (2002). *iLex - A tool for sign language lexicography and corpus analysis*. Paper presented at the LREC 2002 conference, Las Palmas de Gran Canaria, Spain.
- Hanke, T. (2006). Towards a Corpus-based Approach to Sign Language Dictionaries. In *Lexicographic matters and didactic scenarios: Second Workshop on the Representation and Processing of Sign Languages (LREC2006)* (pp. 70–73).
- Hanke, T., & Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In O. Crasborn, T. Hanke, E. D. Thoutenhoofd, I. Zwitserlood, & E. Efthimiou (Eds.), *Construction and exploitation of sign language corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages (LREC)* (pp. 64–67). Paris: ELRA.
- Hunston, S. (2009). The usefulness of corpus-based descriptions of English for learners. The case of relative frequency. In Karin Aijmer (Ed.), *Corpora and language teaching* (Vol. 33, pp. 141-154). Amsterdam/Philadelphia: John Benjamin Publishing Company.
- Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics*, 4(1), 145–169.
- Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In O. Crasborn, T. Hanke, E. D. Thoutenhoofd, I. Zwitserlood, & E. Efthimiou (Eds.), *Construction and exploitation of sign language corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages (LREC)* (pp. 82–87). Paris: ELRA.
- Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15(1), 106–131.
- KOMVA. (1989). Basis-gebarenschat: een beknopt gebarenwoordenboek [Basic sign language vocabulary: a concise sign dictionary]. Amsterdam: NSDSK.
- Konrad, R., & Langer, G. (2009). Synergies between transcription and lexical database building: The case of German Sign Language (DGS). In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference, CL2009*. Liverpool, UK.
- Schermer, T. (2004). Lexical variation in Sign Language of the Netherlands. In M. v. Herreweghe & M. Vermeerbergen (Eds.), *On the lexicon and beyond* (pp. 91-110). Washington D.C.: Gallaudet University Press.
- Schermer, T. (2012). Sign Language Planning in the Netherlands between 1980 and 2010. *Sign Language Studies*, 12(4), 467-493.
- Schermer, T., Stroombergen, M., & Tervoort, B. (1984). How to make a lexicon of sign? In F. Loncke, P. Boyes Braem, & Y. Lebrun (Eds.), *Recent research on European sign languages* (pp. 105-112). Lisse: Zwets & Zeitlinger.
- Schermer, T., Harder, R., & Bos, H. (1988). *Handen uit de Mouwen: Gebaren uit de Nederlandse Gebarentaal in kaart gebracht* [Hands at work: Mapping signs from Sign Language of the Netherlands]. Amsterdam: NSDSK/Dovenraad.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556–1559).