

# BSL Corpus Annotation Conventions

Version 3.0

March 2017

Kearsy Cormier, Jordan Fenlon, Sannah Gulamani and Sandra Smith

Deafness Cognition and Language (DCAL) Research Centre, University College London  
49 Gordon Square, London WC1H 0PD

Contact: [bslcorpusproject@ucl.ac.uk](mailto:bslcorpusproject@ucl.ac.uk)

<b>1</b>	<b>Acknowledgments .....</b>	<b>2</b>
<b>2</b>	<b>Introduction.....</b>	<b>2</b>
<b>3</b>	<b>Using ELAN .....</b>	<b>3</b>
3.1	<i>Opening an eaf file: Single and double video view.....</i>	<i>3</i>
3.2	<i>File naming conventions .....</i>	<i>3</i>
3.2.1	ELAN eaf filenames.....	4
3.2.2	Video filenames.....	4
3.3	<i>Basic template in ELAN .....</i>	<i>4</i>
<b>4</b>	<b>RH-ID gloss tiers and LH-ID gloss tiers .....</b>	<b>4</b>
4.1	<i>Parsing signs: deciding on the start and end points .....</i>	<i>5</i>
4.2	<i>Weak activity .....</i>	<i>5</i>
4.3	<i>ID glossing of core lexical signs.....</i>	<i>5</i>
4.4	<i>Numbers.....</i>	<i>6</i>
4.5	<i>Sign Names.....</i>	<i>6</i>
4.6	<i>Signs from foreign sign languages / other sign systems.....</i>	<i>7</i>
4.7	<i>Buoys.....</i>	<i>7</i>
4.7.1	List buoys.....	8
4.7.2	Pointer buoys.....	8
4.7.3	Fragment buoys .....	8
4.7.4	Theme buoys.....	8
4.8	<i>Conventions for non-core native and non-native signs .....</i>	<i>8</i>
4.8.1	Pointing Signs .....	8
4.8.2	Classifier/depicting signs .....	10
4.8.3	Gesture.....	13
4.8.4	Fingerspelling.....	13
4.9	<i>Errors, mistakes, uncertainties, false-starts, unknown signs.....</i>	<i>14</i>
<b>5</b>	<b>Free Translation.....</b>	<b>15</b>
5.1	<i>Segmentation .....</i>	<i>15</i>
5.2	<i>Referring to signs or fingerspelling.....</i>	<i>15</i>
<b>6</b>	<b>Appendix A: Work-in-progress basic template in ELAN as of March 2017 .....</b>	<b>15</b>
<b>7</b>	<b>Appendix B: Handshape codes for depicting signs.....</b>	<b>16</b>
<b>8</b>	<b>References.....</b>	<b>18</b>

## 1 Acknowledgments

- **UK Economic and Social Research Council (ESRC):** The UK research council which funded the BSL Corpus Project (<http://bslcorpusproject.org>, 2008-2011, RES-062-23-0825) and subsequent corpus-based projects (e.g. BSL Directional Verbs Project, 2012-2014, ES/K003364/1).
- **Digging into Data Challenge** (co-funded by ESRC, Arts and Humanities Research Council UK, Joint Information Systems Committee UK, and NWO Netherlands): Funding source for Digging into Signs project, 2014-2015, <http://www.ru.nl/sign-lang/projects/digging-signs/>
- **BSL Corpus Project:** The many team members of the original BSL Corpus Project and follow-on projects who have contributed to the building of the BSL Corpus so far (<http://bslcorpusproject.org/team>), particularly Adam Schembri.
- **Auslan Corpus Project:** The team members of the Auslan Corpus Project (<http://www.auslan.org.au/about/corpus/>), which was the inspiration for the BSL Corpus and the original basis for these current set of guidelines, particularly Trevor Johnston.
- **Corpus NGT:** The team members of the Corpus NGT 9 <http://www.ru.nl/corpusngt/>, partner in Digging into Signs project, particularly Onno Crasborn
- **DCAL:** The UCL Deafness, Cognition and Language (DCAL) Research Centre which provided partial funding for completion of translations and DCAL staff who have contributed their academic knowledge and dissemination of information.
- **BSL Corpus participants** from the British Deaf community for being involved in the BSL Corpus Project.
- **ELAN: The Language Archive of the Max Planck Institute** in continuing to develop and improve the functionality in ELAN which has supported the flexibility and ease in annotating data. We would like to thank in particular, Han Slöetjes for supporting these developments as well as always being a helpful hand for any pressing issues or queries with ELAN.

## 2 Introduction

This document aims to describe the annotation conventions as they have been used in the BSL Corpus, building on the first and second releases v.1.0 and v.2.0. The main addition in v.2.1 is English translation on the Free Translation tier.

The BSL Corpus is a collection of around 125 hours of signing by deaf native and near-native BSL signers from 8 regions around the UK (Schembri, Fenlon, Rentelis, & Cormier, 2014; Schembri, Fenlon, Rentelis, Reynolds, & Cormier, 2013). It was published as a partly open-source, partly restricted-access video collection in 2011, and is hosted by UCL CAVA (Human Communication Audio-Visual Archive for UCL). The narrative and lexical elicitation data are open access, while the conversation and interview data are restricted to registered researchers only. Further information about the movies, the annotations and the restrictions can be found on the BSL Corpus web site, <http://www.bslcorpusproject.org/cava/>. Both CAVA and a version of this Corpus for a general audience can be found from our Data page: <http://www.bslcorpusproject.org/data/>.

To date, there are around 100 files that have been annotated primarily for manual activity at the lexical level (on right hand and left hand tiers) and that are available on CAVA: 25 each from Birmingham, Bristol, London and Manchester from the conversation data. A substantial part of this annotation work has been carried out for a lexical frequency study (Fenlon, Schembri, Rentelis, Vinson, & Cormier, 2014) with the remainder done as part of a study on directional verbs (Fenlon, Schembri, & Cormier, 2014). Additionally, under the Digging into Signs project, an additional 50 files were annotated at the lexical (ID gloss) level: 25 each from Belfast and Glasgow from the narrative data.

These guidelines are based largely on annotation guidelines for the Auslan Corpus as they existed in 2010, with some adaptations including those which bring the BSL Corpus annotations closer in line with Corpus NGT and updates on the Auslan Corpus annotation guidelines since 2010 (see Johnston 2014). ID gloss annotations for lexical signs link to BSL SignBank which initially began as a lexical database and has been developed into an

online dictionary (<http://bslsignbank.ucl.ac.uk>). For advanced access to BSL SignBank, which provides access to ID glosses, register on SignBank as a researcher (staff or student); follow links for providing Feedback. For more on ID glosses, see section 4.3.

The preferred minimum number of annotation tiers for any sign language corpus is three: two ID gloss tiers and also a translation tier (Johnston, 2014). So far translations have been completed for some of the BSL Corpus data (see release notes accompanying v.2.1); further translations will be made available at a later date depending on funds.

### 3 Using ELAN

ELAN allows the inserting of annotations with a precise time-alignment referring to the video(s) in display, using multiple tiers to display different kinds of information about a particular sign in a particular time-length/annotation length. This makes it possible to easily match annotations with the raw video data and allows for flexibility of having different kinds of information annotated on multiple tiers in one file. We use the latest version (and do so consistently on all the computers accessing the corpus) in order to exploit and take advantage of the latest functionality tools in ELAN as well as avoid errors, file conflicts, inconsistent practices and old versioning layouts in comparison to new versions if using several different versions. We found that by taking this approach, it enabled a smoother workflow and files within the team and also it improved some function ability in ELAN, in particular the multiple file processing. Some functions in ELAN are particularly useful for corpus data where several eaf files are produced and can be manipulated; such as multiple search, find and replace, multiple file processing, importing and exporting.

For more information about ELAN and its features, the manual can be downloaded here: <https://tla.mpi.nl/tools/tla-tools/elan/>

#### 3.1 Opening an eaf file: Single and double video view



Figure 1. eaf file for Participant X.

Each eaf file consists of information about one participant (Participant X); using LH-IDgloss and RH-IDgloss as the two main tiers for annotating depending on which hand(s) are being used by Participant X. The single view showing only Participant X is presented first as a master media which is the first video media file to be linked to the eaf file (this is what ELAN will ask you to find first when you open the eaf, followed by the double view). Next to the single view of Participant X is the double view which includes Participant X's signing partner, Participant Y. Annotating data always follows the single video view (Participant X). The start of annotating starts with the first sign that is produced. The double view is used both as a backup source of information in case the single view production of a particular sign by Participant X is not clear, and also to allow reference to Participant Y for content of information to support annotating the data.

#### 3.2 File naming conventions

It is important for corpus files to be named appropriately in order to enable a quick scan of information and for files to be easily located.

### 3.2.1 ELAN eaf filenames

ELAN eaf files are named with the following demographics in this order: Region, Participant number, Gender, Age, Ethnicity, Deaf/Hearing family, Task.

#### Possible values

Region: LN (London), BM (Birmingham), BL (Bristol), BF (Belfast), N (Newcastle), C (Cardiff), GW (Glasgow)

Participant number within region: A number between 01 and 36 (for most regions this is a number between 01 and 30 but there are higher numbers in a few regions)

Gender: M or F

Age: Age in years

Ethnicity: A (Asian), B (Black), W (White)

Deaf/Hearing family: D or H

Task: L (lexical elicitation), I (interview), C (conversation), N (narratives), or NC (combined narrative/conversation)

#### Example

BM07F35WHC

Region	Participant number	Gender	Age	Ethnicity	Deaf/Hearing Family	Task
BM	07	F	35	W	H	C

### 3.2.2 Video filenames

Video files are named with a slightly different convention from the eaf files as video files consist of either a view of a single participant or a view of two participants. Video filenames consist of a code for Region followed by Participant number(s) followed by Task.

#### Possible values

Region: L (London), BM (Birmingham), BL (Bristol), BF (Belfast), N (Newcastle), CF (Cardiff), G (Glasgow)

Participant number within region: A number between 1 and 36 (for most regions this is a number between 1 and 30 but there are higher numbers in a few regions)

Task: l (lexical elicitation), i (interview), c (conversation), n (narratives), n-c or n+c (combined narrative/conversation)

#### Examples

BM7c: Birmingham participant 7 (single view) in combined narrative/conversation task

G13+14l: Glasgow participants 13 and 14 (double view) in lexical elicitation task

## 3.3 Basic template in ELAN

The basic template for the 2.1 release of BSL Corpus annotations has three tiers: RH-ID gloss, LH-ID gloss, and Free Translation. Our work-in-progress template uses 49 tiers that are arranged in a hierarchy and controlled vocabularies have been assigned to some tiers. See Appendix for a list of work-in-progress tiers.

## 4 RH-ID gloss tiers and LH-ID gloss tiers

These tiers represents all manual material articulated on either the right or left hand. For first release, we only annotated lexical signs on the dominant hand. For example, the two-handed DRIVE is only annotated on the tier reflecting the signer's dominant hand (whether it is the left or right hand). *In the second release, for some files, two-handed signs are annotated on both the RH and LH tiers, with the start and endpoint determined by the*

*dominant hand (see release notes for details.)* If each hand was producing separate meaningful units (e.g. a lexical sign on one hand and a pointing sign on the other) then these were annotated on both the RH and LH tier as appropriate.

#### 4.1 Parsing signs: deciding on the start and end points

The start point for a sign is when the hand or hands appear to start moving away from articulating the previous sign. This is signalled by a change in direction, orientation, and/or handshape. The end point for a sign is when the hand appears to start moving towards articulating the following sign. Again, this is signalled by a change in direction, orientation, and/or handshape.

A sign sequence is considered to be finished normally when the hands begin a return to a rest position (e.g., folded arms, hands on hips, laps, or some supporting surface or object, or arms resting at the side of the body). Signers may, however, maintain their hands in a signing position without actually signing in order to signal their desire for a turn, or to hold the conversational floor.

For the first release, the two hands in signing do not always move in the same way and some handshapes can spread beyond a lexical sign. For this reason, the start and end point for a sign may be different if one is looking at either the right or left hand. Annotations conducted on the RH and LH tiers are therefore done independently from the other hand. *For the second release: The RH and LH tiers remain independent if there are two different signs or other constructions e.g. classifier constructions being produced at the same time or overlapping in time. In the second release, for two-handed signs or for dependent combinations (such as buoys and their counterpart on the dominant hand), for some files, the start and end of the sign follows the dominant hand (these are signs which have been “double glossed” - see release notes for details).*

There are small gaps (2 frames) between each annotation partly for historical reasons (in previous versions of ELAN, annotations which touch each other were problematic for exporting).

#### 4.2 Weak activity

Weak activity is not annotated. Weak activity is defined as instances where the hand is relaxed and/or partially forming the handshape or partially copying the movement and where there is no discernible addition to meaning or communicative intent.

#### 4.3 ID glossing of core lexical signs

- All lexical signs are annotated using an identifying gloss (*Annotation ID gloss*) from BSL SignBank if one exists (if one does not, see 6.4). An ID gloss is an English gloss (always in upper case, e.g., SISTER) that is consistently used with a unique sign (or ‘lemma’) to represent the sign whether in citation form or any phonological or morphological variant. If a sign needs more than one distinct English word to gloss it, hyphens are placed between the words (spaces are not used), e.g. PULL-APART not PULL APART OR PULLAPART. The ID gloss for each sign (in citation form) can be found in the BSL SignBank and is usually the same as one of the keywords associated with the sign. It is important that BSL SignBank is consulted to ensure that the right ID gloss is used at all times.
- As work on BSL SignBank is ongoing, signs are frequently encountered that have not yet been added to this resource. These signs are annotated as ‘ADD-TO-SIGNBANK’ as a placeholder followed by a suggested ID gloss in parentheses (e.g., ADD-TO-SIGNBANK(RED3)). These ADD-TO-SIGNBANK tokens are regularly discussed by the team to resolve any potential lemmatisation issues and then, when an appropriate ID gloss is agreed, these are added to SignBank and the placeholder replaced in the ELAN files. Further information about ID glossing and lemmatisation can be found in Fenlon et al. (under review).
- If a sign appears to be a lexical sign but is not known by the annotator, this is glossed as ADD-TO-SIGNBANK(UNKNOWN).

- Lexical variants have the same or similar/related meanings but (unlike phonological variants) they generally differ in two parameters or more from each other. Lexical variants are distinguished using a numeral tag (e.g., BROWN, BROWN2 and BROWN3). Note that the first lexical variant is not indicated with a number (e.g., BROWN not BROWN1). (This is simply to aid in the speed of annotation, eliminating the need to find/replace all tokens of e.g. BROWN with BROWN1 in ELAN and the same change made in BSL SignBank, when BROWN2 is encountered.)
- Phonological/morphological variants are slowly being introduced into BSL SignBank. These are marked with a small letter suffix – b, c, d, etc. in the ‘ID gloss’ field in SignBank. Examples include BOOKEDb, LOVELY2b and BIRTHDAY3b. Note that the citation form is not indicated with a letter a (e.g. BOOKED not BOOKEDa). Phonological and morphological variants are marked in BSL SignBank with tags (phonology:variant and morphology:variant). Current practice is to annotate Annotation ID gloss only (i.e. citation form) – i.e. BIRTHDAY3 and not BIRTHDAY3b, but having the phonological and morphological variants visible in SignBank and tagged as such can help annotators see which lemma a particular token belongs to.
- Manual negative incorporation is glossed using a negative suffix: KNOW-NOT not DON’T-KNOW. These signs will have a separate entry listed in BSL SignBank. If a sign signals negation using a non-manual headshake and with no modification to the sign, there is no need for a negative suffix. Instead, this information will be provided by the translations and will be annotated on the non-manual tiers in future.
- If a sign is repeated then each instance is annotated separately, as noted in the example below (courtesy of Johnston 2014). Note this is only done if the repetition is of the entire sign rather than for repetition of movement within a single sign for phonological or morphological purposes which are not included in the ID gloss tiers.
  - ID-gloss tiers: BOY SHOUT WOLF WOLF WOLF  
Free trans tier: *The boy cried “wolf, wolf, wolf”.*
- Most compounds are found with distinct ID-glosses in BSL SignBank, e.g., the compound combining MOTHER and FATHER is a standard compound PARENTS. If a pairing of signs cannot be found in BSL SignBank as a compound, the two signs are tentatively annotated as one sign with two ID-glosses but separated with a caret (e.g. FS:G-GRAPHIC^ART for 'graphic designer'). Each of these tokens will be returned to later and will either be given its own ID-gloss if it is found to be in widespread use or treated as two signs to be annotated and glossed separately. In some cases, these tokens may be best regarded as a collocation. (A collocation could be a potential compound if the overall meaning is not predictable from the two signs that are paired. A collocation is unlikely to be a compound if it is possible to insert another sign between the two signs.)

#### 4.4 Numbers

- If a signer uses a number to refer to anything, it is glossed using words and not digits (e.g. FOUR, FOUR2, etc *not* 4). All unique number signs are listed in BSL SignBank.
- If a number is incorporated into a sign, the number is added to the end of the ID gloss after a hyphen, again using words and not digits: e.g., AGE-FOURTEEN *not* AGE-14. If the ID gloss is a lexical variant with a number, the incorporated number should likewise be spelled out – e.g. RANKING2-THREE. Only add ONE as a suffix if the citation form does not use the 1-handshape (e.g. AGE-ONE, but POUND *not* POUND-ONE).
- If there is a number sequence, this is glossed as one unit in ELAN with carats (^) to separate the numbers. This is generally used when there is a sequence of signs that is difficult to segment further (e.g., suspected compounds or number sequences - NINETEEN^EIGHT^NINE *not* NINETEEN-EIGHTY-NINE(NINETEEN^EIGHT^NINE)).

#### 4.5 Sign Names

*NOTE: We acknowledge that specification of first and/or last names is not consistent in these guidelines. Additionally, it is likely that other aspects of annotation of sign names has not been fully consistent with these guidelines in the second release. For future releases we will aim to make these more consistent.*

- Signs name are entered with the prefix *SN:* followed by the proper name. The sign name for a person called *Peter* would be written as follows: *SN:PETER*, unless the sign name is identical in form to a lexical sign (see below).
- If the sign name is identical in form to a lexical sign, the ID gloss for the relevant sign may be identified after the name in brackets: e.g., *SN:MISS-JENKINS(HAIR-BUN)*, *SN:WEMBLEY(STADIUM)* or *SN:OSAMA-BIN-LADEN(BEARD)*.
- If the sign name uses a lexical sign that is in BSL SignBank but the annotator is unable to determine the name of the referent in this instance then the gloss *UNKNOWN* is used (e.g., *SN:UNKNOWN(WOLF)*).
- If the sign name is based on fingerspelling, the form is entered in brackets after the name and follows the conventions for fingerspelling as outlined below: e.g., *SN:PETER(FS:PETER(PR))*, *(SN:PETER(FS:P-PETER))*, or *SN:ALEX-FERGUSON(FS:A-ALEX^FS:F-FERGUSON)*.
- If the sign name represents a sequence of both fingerspelling and a lexical sign, the whole sequence is entered as one sign name. The fingerspelled element and lexical element are included in brackets separated by a caret (e.g., *SN:JOHN-KING(FS:J-JOHN^KING)*).
- It can be difficult to determine when a fingerspelled sequence is in fact a sign name. Generally, we assume that fingerspelled sequences that use the initial letter of the name or fingerspelled sequences that are reduced so that they appear like lexical signs are sign names that have some degree of conventionalisation. Therefore, fully fingerspelled sequences (e.g., *FS:BARRY* where each letter is articulated clearly) are typically entered as fingerspelled sequences and not sign names (i.e., they do not have the prefix *SN* attached to them).
- Sign names are often for people but may be for e.g. places, organisations, etc. Some signs names however have been judged to be institutionalised and consistently used (some more than others) across the British Deaf community and thus are included in BSL SignBank as lexical signs (e.g., *LONDON*, *BRISTOL*, *SEE-HEAR*, *DEAFINTELY-THEATRE*).
- If the annotator cannot determine what the signed sequence represents, the glosses *INDECIPHERABLE* or *unknown* may be used (e.g., *SN:INDECIPHERABLE(FS:INDECIPHERABLE(H))*, *SN:UNKNOWN(UNKNOWN)*).

#### **4.6 Signs from foreign sign languages / other sign systems**

- Occasionally, signers use a sign that appears to be borrowed from a foreign sign language. These signs are also assigned an ID gloss and occur in BSL SignBank, along with doubtful lexeme tag (i.e., to indicate it is not a sign that belongs to BSL's core lexicon) and a Note to indicate that it is a possible borrowing. A similar approach is taken with signs that are borrowings from other signed systems (e.g., *Paget-Gorman Sign System*).

#### **4.7 Buoys**

Buoys are configurations on the non-dominant hand while the dominant hand continues to sign. There are several types of buoys that can be expected in signed discourse. The table below details some of the buoys that have been observed and annotated to date – these are based on Liddell (2003).

ID Gloss	Definition
LBUOY	List buoy
PBUOY	Pointer buoy
FBUOY	Fragment buoy
TBUOY	Theme buoy

#### 4.7.1 List buoys

- When producing a list buoy, a certain number of fingers are held stretched out on the non-dominant hand, each one referring to an entity or idea, that are all somehow related, often sequentially. For example, an index finger can be held up to indicate the first of a series of items. The list buoy is annotated as LBUOY in each instance, followed by the ID gloss of the relevant number sign e.g. LBUOY-FOUR02. Note that list buoys can be lexicalised – e.g. OTHERS in SignBank.
- If the signer points to a list buoy with the dominant hand, the point is annotated as PT:LBUOY.

#### 4.7.2 Pointer buoys

- With a pointer buoy, the non-dominant hand points to a location associated with an important element in the discourse while the dominant hand continues signing. This is annotated as PBUOY.

#### 4.7.3 Fragment buoys

- With a fragment buoy, the non-dominant hand is held from a preceding sign, it is intended, and it carries some meaning. Typically this meaning is conveyed by the signer pointing to or looking at or directing attention to the fragment buoy in some way. This differs from perseveration whereby the non-dominant hand is held from a preceding sign but may or may not be intended to carry any meaning. This is annotated as FBUOY.
- If the signer points to a fragment buoy with the dominant hand, the point is annotated as PT:FBUOY.

#### 4.7.4 Theme buoys

- With a theme buoy, the non-dominant hand (normally in the form of a vertical extended index finger) represents an important theme in the discourse while the dominant hand continues signing. This is annotated as TBUOY.

### 4.8 Conventions for non-core native and non-native signs

This section includes pointing signs, gestures, classifier signs, and fingerspelling.

#### 4.8.1 Pointing Signs

- All pointing signs are prefixed with ‘PT’ so that they can be retrieved quickly across multiple files via the search function within ELAN. Each token is then further categorised according to its function. Below is a table detailing the different functions of pointing signs identified thus far in the BSL Corpus and the glossing conventions associated with each type. Note that not all of these are in BSL SignBank – only those which also appear in Brien 1992 (and, in some cases, their variants).

ID Gloss	Function
PT:PRO1SG	1 <sup>st</sup> person singular
PT:PRO2SG	2 <sup>nd</sup> person singular
PT:PRO3SG	3 <sup>rd</sup> person singular
PT:PRO1PL	1 <sup>st</sup> person plural
PT:PRO2PL	2 <sup>nd</sup> person plural



PT:PRO3PL	3 <sup>rd</sup> person plural
PT:DET	Determiner (singular)
PT:DETPL	Determiner (plural)
PT:LOC	Locative (singular)
PT:LOCPL	Locative (plural)
PT:POSS1SG	1 <sup>st</sup> person singular possessive
PT:POSS2SG	2 <sup>nd</sup> person singular possessive
PT:POSS3SG	3 <sup>rd</sup> person singular possessive
PT:POSS1PL	1 <sup>st</sup> person plural possessive
PT:POSS2PL	2 <sup>nd</sup> person plural possessive
PT:POSS3PL	3 <sup>rd</sup> person plural possessive
PT:BODY	Point to a body part
PT:LBUOY	Point to a list buoy
PT:FBUOY	Point to a fragment buoy
PT:BUOY	Point to a buoy (unclear if list or fragment)
PT:	Ambiguous point

- First person pronouns are typically unambiguous (as are second person pronouns), at least in terms of the referent they are pointing to (not necessarily the referent intended, depending on e.g. presence of constructed action). However, third person pronouns and locative points (i.e., adverbials) are often difficult to distinguish - there is some discussion in the literature about whether strictly pronominal functions for pointing can be distinguished from other uses of pointing in sign languages or indeed in gesture (Cormier, Schembri, & Woll, 2013; Johnston, 2013a, 2013b).
- Third person pronouns are typically identified as a point to the peripheral signing space, away from the conversational partner, and serving as a referent for another person in the discourse. Locative points are identified as a point to a location (e.g., associated with a place-name mentioned in the discourse) or the location at which a topic being discussed was situated. Ambiguous points may be ambiguous between a referent and the location of that referent (e.g., a person or the location where that person is standing), making it difficult to determine whether they are primarily locative or pronominal pointing signs.
- If two possible functions exist for a given pointing sign and it is difficult to decide between the two, then both possibilities are entered separated by a forward slash (e.g., PT:LOC/PT:PRO3SG). More than two possibilities may also be entered each separated by a forward slash.
- If there are more than two possible functions for a given pointing sign, then just the prefix is entered 'PT:'.
- Locative points (singular or plural) can function as adverbials. They may point to a location in space or they may be used to refer to a specific point in time (e.g., points outwards from the signer may refer to the future, downwards points may refer to the present, and points directed behind may refer to the past).
- Pointing signs functioning as determiners are identified by their syntactic position adjacent to nominal signs and by prosody. If the point (whether it occurred before or after a noun) could be grouped with a noun as a single cohesive prosodic unit, then it is classed as a determiner.
  - Demonstratives (e.g., English equivalent of 'that' or 'this') are glossed as either PT:PRO3SG or PT:DET depending on its syntactic position. For example, the signed equivalent of 'that house' would be glossed as PT:DET and the signed equivalent of 'where do you want **this**?' would be glossed as PT:PRO3SG.
  - The same holds for plural demonstratives – either PT:PRO3PL or PT:DETPL depending on syntactic position.

NOTE: It is likely there is some inconsistency in annotating locative points when they function as determiners – e.g., HOUSE PT:LOC vs. HOUSE PT:DET.

- Points to the body are also prefixed with ‘PT’ followed by BODY. We also provide additional information as to which body part the pointing sign is directed towards (e.g., PT:BODY-LEG). Body parts annotated this way include: PT:BODY-EAR, PT:BODY-NOSE, PT:BODY-PHILTRUM, PT:BODY-KNEE, PT:BODY-LOWER-BACK, PT:BODY-ANKLE, PT:BODY-FOOT, PT:BODY-NECK, PT:BODY-THROAT, PT:BODY-EYE, PT:BODY-HEAD, PT:BODY-BACK – this list can be added to if necessary. Note the only PT:BODY signs that occur in BSL SignBank are those which are listed in Brien 1992. Note that some body parts have lexical signs that do not involve pointing; these are in BSL SignBank- e.g. ARM, LEG, SHOULDER.
- If a pointing sign points to the same location several times (e.g., is repeated for emphasis), this is annotated with one gloss. That is, the full annotation includes the sequence of pointing signs made to the same location.
- If there are multiple points to the different locations close to each other and one right after the other, and each successive point has a smaller movement than the previous, annotate this as a plural pointing sign annotated with one gloss (e.g. PT:PRO3PL). If there are multiple points to the different locations one right after the other (close to each other or not), and each has about equal emphasis, annotate this as separate tokens of singular pointing glosses (e.g. PT:PRO3SG PT:PRO3SG PT:PRO3SG).
- Plural points are determined based on function (i.e. meaning) rather than form. This is so we can determine whether there is a specific form associated with plural marking (e.g., an arc movement) at a later date.

NOTE: It is possible that annotation of plurality has not always been based on function over form.

- Plural points may also incorporate number signs. If a number sign has been incorporated into a plural point, then the ID gloss for the number it represents is suffixed to the end of the gloss after a hyphen (e.g., PT:PRO1PL-TWO, PT:PRO2PL-THREE2, PT:PRO3PL-FOUR3).

#### 4.8.2 Classifier/depicting signs

- Depicting signs are annotated with the prefix *DS* and additional letters specifying the type of classifier sign, followed by handshape code (see Appendix B), followed by movement type. This class of signs is often referred to as classifier signs. Here we follow Johnston (2014) in referring to them as depicting signs (cf. also Liddell 2003 and Cormier et al. 2012). Note that there are special annotation guidelines for type-like depicting signs - i.e. those which are thought to occur very frequently – see below for those.

Function	Name	Explanation
DSEW	Depicting sign: entity (whole)	Depicts whole entities that can move freely in space
DSEP	Depicting sign: entity (part)	Depicts parts of entities (e.g. body parts) that are tied to the body
DSH	Depicting sign: Handling	Depicts the handling of an object
DSS	Depicting sign: showing size and shape	Depicts the size and shape of entities, most often with a tracing movement but also sometimes with a hold

Movement type	Explanation
MOVE	Path movement of a referent through space, i.e. from one location to another location. This may be independent movement of a referent (e.g. a cat moving to and fro), or a manipulated movement (e.g. a case and a birdcage being picked up).


PIVOT	Change of position of a referent, expressed by a change in hand orientation. Although here, too, there is an intended, meaningful movement, it does not indicate a spatial trajectory but rather a movement around a pivot point e.g. via wrist nodding or forearm rotation.
AT	Localisation of a referent, i.e. when the hand makes a (short) movement towards a particular location in sign space
BE	<i>BE</i> is used in cases where a signer uses a depicting sign but there is no intended meaningful movement, localization, or change of position visible. Note: Only use this when the construction is <i>not</i> a fragment buoy (FBUOY).








- Examples include DSEW(1)-AT, DSEP(2)-PIVOT, DSH(FIST)-MOVE, DSS(1). Movement type is not specified for DSS.
- In case a depicting sign concerns both a movement through space and a change in the referent's orientation (e.g. expressing that a referent falls down from a window), this is annotated as MOVE (not as PIVOT, or MOVE-PIVOT).
- Depictions of handling which appear to be the same as one would actually handle an object may be ambiguous between handling (DSH) and constructed action (G:CA). These are annotated as other ambiguity is annotated, with both options with a slash in between – e.g. DSH(FIST)-MOVE/G:CA:push-pram. Depictions of handling which appear different from how one would handle an object including those with evidence of modification are annotated simply as handling – e.g. for plural modification with handling such as DSH(OPEN)-MOVE depicting handing out of pagers.
- Depicting sign sub-type categorisation is usually made easier by looking at the immediate linguistic environment or context-of-utterance rather than simply at the form of the sign alone. For example, in the following two strings the same form on the dominant hand is given handling status in one but a whole entity status in the other, as a result of considering the type of sign that immediately precedes each instance (pronominal in the first, verbal in the second):

RH-IDgloss	PRO1SG	DSH(CYL)-MOVE
LH-IDgloss		DSEW(FLAT)-BE
RH-IDgloss	HAVE	DSEW(CYL)-MOVE
LH-IDgloss		DSEW(FLAT)-BE

Following on from this, it will be evident that even though many classifier signs use both hands, often only a single entity or action is depicted. However, each hand usually carries its own semantic load in that depiction, so the annotator may categorise each hand differently, e.g., the dominant as DSEW-MOVE and the subordinate as DSEW-BE as above.

For type-like depicting signs – i.e. those that occur most commonly – the following specific conventions are used (following Johnston (2014)). These are the same as for other depicting signs except that a specific set of handshape codes are used in which orientation is encoded as well, and information about the animacy of the referent is included at the end (after a colon). An example of a type-like depicting sign annotation is DSEW(1-VERT)-MOVE:HUMAN.

Type-like depicting signs	Explanation
DSEW(1-VERT) 	Something tall-ish and thin-ish. This is the basic form of the annotation for a depicting sign using the upright index handshape which is oriented and located/moved in the signing space. It can face in any direction. Use this if the thing that is located/moving is human, animal, or inanimate (real/imagined, concrete/abstract, literal/metaphorical). The palm side is assumed to be the front of the entity and the finger the head, if relevant.

DSEW(1-HORI) 	Something longish and thin-ish. This is the basic form of the annotation for a depicting sign using the horizontal index handshape which is oriented and located/moved in the signing space. It can face in any direction. Use this if the thing that is located (lying down) or moving is human, animal, or inanimate (real/imagined, concrete/abstract, literal/metaphorical). The palm side is assumed to be the front of the entity and the finger the head, if relevant.
DSEW(2-DOWN) 	Something two-legged and standing. This is the basic form of the annotation for a depicting sign using the two handshape with the fingertips pointing downwards which is located/moved with a certain orientation in the signing space. It can face in any direction. Use this if the thing that is located/moving is human or animal. The knuckle side is assumed to be the front of the person/animal, and the fingertips the feet.
DSEW(2-HORI) 	Something two-legged and reclining. This is the basic form of the annotation for a depicting sign using the two handshape with the fingertips pointing horizontally which is placed with a certain orientation in the signing space. It can face in any direction and be placed in any location. Use this if the thing that is located/moving is human or animal. The palm side is assumed to be the front or stomach side of the entity, and the fingertips the feet.
DSEW(BENT2-HORI) 	Something two-legged and reclining. This is the basic form of the annotation for a depicting sign using the bent two handshape with the fingertips pointing downwards which is located/moved with a certain orientation in the signing space. It can face in any direction and be placed in any location. Use this if the thing that is located/moving is human or animal. The knuckle side is assumed to be the front of the entity, and the fingertips the feet.
DSEW(5-HORI) 	Multiple/many things. This is the basic form of the annotation for a depicting sign using the horizontal five handshape with the palm facing downwards (supine) and the fingertips pointing horizontally which is located and moved in the signing space. It can face and be move in any direction. Use this if the thing that moves is many humans, animals or entities (real or imaginary, concrete or abstract). The fingertips face the direction of real or metaphorical movement, and may wiggle.
DSEW(FLAT-LATERAL) 	Something vehicle-like. This is the basic form of the annotation for a depicting sign using the flat handshape with the palm facing sideways (laterally) and the fingertips pointing horizontally which is located/moved in the signing space. It can face in any direction and be placed in any location. Use this if the thing that is located/moving is a vehicle. The fingertips are assumed to be the front of the vehicle and the little finger edge of the hand the underside.
DSEW(FLAT-HORI) 	Something vehicle-like. This is the basic form of the annotation for a depicting sign using the horizontal flat handshape with the palm facing downwards (supine) and the fingertips pointing horizontally which is located/moved in the signing space. It can face in any direction and be placed in any location. Use this if the thing that is located/moving is a vehicle. The fingertips are assumed to be the front of the vehicle and the palm side the underside of the vehicle.

Given the possible combinations with proposed movement categories noted above, examples of full annotations of type-like depicting signs include the following (note that for most of the following, 1-VERT could be replaced by 1-HORI, 2-DOWN, 2-HORI, or BENT2-HORI):

- DSEW(1-VERT)-MOVE:HUMAN
- DSEW(1-VERT)-PIVOT:HUMAN (unlikely)
- DSEW(1-VERT)-AT:HUMAN
- DSEW(1-VERT)-BE:HUMAN
- DSEW(1-VERT)-MOVE:ANIMAL
- DSEW(1-VERT)-PIVOT:ANIMAL (unlikely)
- DSEW(1-VERT)-AT:ANIMAL
- DSEW(1-VERT)-BE:ANIMAL
- DSEW(1-VERT)-MOVE:ENTITY
- DSEW(1-VERT)-PIVOT:ENTITY (unlikely)

- DSEW(1-VERT)-AT:ENTITY
- DSEW(1-VERT)-BE:ENTITY
- DSEW(5-HORI)-MOVE:MANY-HUMANS
- DSEW(5-HORI)-MOVE:MANY-ANIMALS
- DSEW(5-HORI)-MOVE:MANY-ENTITIES
- DSEW(FLAT-LATERAL)-MOVE:VEHICLE
- DSEW(FLAT-LATERAL)-PIVOT:VEHICLE (unlikely)
- DSEW(FLAT-LATERAL)-AT:VEHICLE
- DSEW(FLAT-LATERAL)-BE:VEHICLE
- DSEW(FLAT-HORI)-MOVE:VEHICLE
- DSEW(FLAT-HORI)-PIVOT:VEHICLE (unlikely)
- DSEW(FLAT-HORI)-AT:VEHICLE
- DSEW(FLAT-HORI)-BE:VEHICLE

For any depicting sign that does not fall into one of the type-like categories noted above, use the handshape categories in the Appendix.

Sometimes it is difficult to decide whether a sign should be glossed as a depicting sign or a lexical sign (i.e. a sign with a fixed form and a specialized meaning). This is especially challenging when the sign expresses or implies a movement, for example in DANCE or MEET. Tips for distinguishing a depicting sign from a lexical sign are:

- Depicting sign: The signer intentionally conveys information about movement through or in space of a referent or shape or size of an object
- Depicting sign: The signer looks at his/her hands while articulating the sign;
- Depicting sign: The signer uses a specific body posture or facial expression;
- Lexical sign: The signer does not look at his/her hands while articulating a lexical sign
- Lexical sign: The signer mouths an English keyword while articulating the sign.

Note that these are only tips. As such, they will not always provide sufficient means for the distinction.

#### 4.8.3 Gesture

- All gesture annotations begin with ‘G:’ followed by a brief description of its meaning (not form – i.e. G:HOW-STUPID-OF-ME and not G:HIT-PALM-ON-FOREHEAD).
- The gesture with upturned hands (also known as the ‘palm-up gesture’) is annotated as G:WELL. This is the second most frequent token in the BSL Corpus to date.
- Some emblems are lexicalised and glossed as lexical signs without the gesture prefix (e.g., GOOD) although this is not always the case (e.g., G:FUCK-OFF). Whether with the gesture prefix or not, emblems have been added and are being added to BSL SignBank. As with G:WELL, the form associated with G:FUCK-OFF is consistently recognised using the same gloss (also provided in BSL SignBank). This is in recognition of the fact that they appear to have consistent form/meaning mappings even though their lexical status is unclear.
- Tokens of constructed action are also recognised as instances of gestural activity. Such instances are marked with the prefix ‘G:CA:’. As with classifier signs and other types of gesture, this prefix is followed by a brief description of the token’s meaning (e.g., G:CA:HOLD-HANDS-UP-IN-FRIGHT). CA tokens are not lexicalised and thus are not included in BSL SignBank.

#### 4.8.4 Fingerspelling

Fingerspelled forms in BSL represent a sequence of hand configurations that have a one-to-one correspondence with the letters of the English alphabet. Fingerspelled forms often violate phonological constraints associated

with core native signs and are said to belong to what is known as the ‘non-native lexicon’ (Brentari & Padden, 2001). Below are the conventions used for fingerspelled sequences whilst annotating the BSL Corpus.

- Fingerspelling is annotated with the fingerspelled word prefixed with *FS* for ‘fingerspelling’ followed by a colon and then the word spelled, e.g., *FS:WORD*.
- If not all the letters of a word are spelled, but it is clear what word the signer is attempting to fingerspell, the full spelling of the intended word is entered (not the misspelling – e.g. *FS:WORD(WRD)* and not *FS:WRD*).
- If not all the letters of a word are spelled, and it is *not* clear what word the signer is attempting to fingerspell, the gloss *INDECIPHERABLE* is used followed by the actual letters produced in brackets (e.g., *FS:INDECIPHERABLE(GTH)* or *FS:B-INDECIPHERABLE*).
- If the fingerspelling is for multiple words, a new annotation per word is begun even if it is one continuous act of fingerspelling (e.g., *FS:MISS FS:JENKINS* and not *FS:MISSJENKINS*).
- If the form is a single fingerspelled letter (or single fingerspelled letter repeated), the letter and the word it stands for are included in the annotation. In this case, the fingerspelled letter precedes the word it represents (e.g., *FS:F-FORTUNE*, *FS:C-CONTRIBUTION*). These sequences are also known as single manual letter signs (SMLS signs). Even if the single manual letter is repeated, only one English letter is included in the annotation (e.g., *FS:F-FORTUNE* not *FS:FF-FORTUNE*).
- Some SMLS or fingerspelled sequences are actually lexical signs in their own right. These include *MOTHER*, *YEAR*, *YELLOW*, *DAUGHTER* and *CLUB*. Although they are based on fingerspelling, these signs do not have the prefix ‘*FS:*’ because we do not use this prefix when the sign in question is a fully lexical sign. These signs are in BSL SignBank.
- The distinction between fingerspelled forms and lexicalised fingerspelled signs is often difficult to maintain given that many fingerspelled forms can appear partly nativised (i.e., may be in the process of becoming a fully lexical sign). For example, the sign *SATURDAY3* is a fingerspelled loan (based on *S-A-T*) that is considered partly nativised for it does not follow constraints imposed on fully lexical signs - e.g., it violates the selected fingers constraint (Brentari, 1998). To categorise these tokens in a principled way, we use guidelines based on Cormier et al. (2008). Fingerspelled loans (with remnants of 2 or more letters) are accepted as lexical signs if there is evidence of phonological restructuring to make them more native-like. Additionally, independent agreement from native signers can also be sought to be certain that this form is consistently used for that meaning (e.g., *SATURDAY3* was accorded lexical status following these criteria).

#### 4.9 Errors, mistakes, uncertainties, false-starts, unknown signs

- If a sign appears to be a lexical sign but is not known by the annotator, this is glossed as *ADD-TO-SIGNBANK(UNKNOWN)*.
- If two possibilities exist for a single token and it is difficult to decide between the two, then both ID glosses are entered with a forward slash between each token (e.g., *LOOK/THINK*). More than two ID glosses are entered, each separated by a forward slash, if more than two possibilities exist (e.g., *LOOK/THINK/SEE*). This convention is also used to demonstrate ambiguity between sign types - for example, when one is uncertain if a given token is a lexical sign or a sequence of constructed action (e.g., *TEA/G:CA:DRINK-TEA*).
- If the identity of a sign is uncertain, but a possibility exists, then the ID gloss is entered prefixed by a ‘?’ (e.g., *?HOME*). This may be used because the token looks like it could be a phonological variant of this lemma or it may be a separate (new) lemma.
- If a sign cannot be identified (e.g., it is poorly articulated or has not been completed), it is impossible to say with any certainty what the sign is, and it is likely that nobody would be able to tell what the sign is, then the sign is glossed as *INDECIPHERABLE*. *Note: There are some possible inconsistencies in*

*annotation of INDECIPHERABLE versus ADD-TO-SIGNBANK(UNKNOWN) that we hope to fix in later releases.*

- If a sign is not finished but is easily identified, the corresponding ID gloss is entered followed by FALSE-START in brackets (e.g., DOG(FALSE-START)). This convention is also extended to signs that are indecipherable because the signer did not finish what they were going to sign (INDECIPHERABLE(FALSE-START)).

## 5 Free Translation

The Free Translation tier provides a full, readable translation of the BSL into written English (unlike the ID glossing which is not readable). Although the tier that we used is called Free Translation, this is to distinguish it from the Literal Translation tier which we expect to use later on and which is likely to be so literal that it may be at times unreadable as English. The Free Translation here is actually quite close to the source BSL. We avoid using a formal register or a style that is very informal – the aim is plain, simple English.

### 5.1 Segmentation

The primary method of segmenting is to allow for one annotation per English sentence. When the translation needs a full stop, we start a new annotation. As a general rule:

- We avoid semicolons, commas, dashes and hyphens to join phrases in preference to simple sentences.
- We avoid multiple conjunctions (and, but, so) in a single annotation.

Additionally, when segmenting, a few frames in between each annotation are used so that annotations do not touch each other (to assist with exporting).

Each annotation contains all and only the signs that are relevant to the translation in question. Where there are periods of no signing, there are no annotations. The starting point roughly corresponds to the beginning of the first sign of the translated utterance (i.e., roughly when the hands start to move to form this sign) and end at the last sign of the translated utterance (i.e., when the hands start to move to the next utterance or start to move towards the signer's lap).

Where possible, full English sentences are used. However, this is not always possible, especially in the conversation data. False starts are not included in the translation.

### 5.2 Referring to signs or fingerspelling

When signers talk about particular signs or particular fingerspelled words, these are translated based on what may be expected in a voiceover – e.g. *'they would teach us the sign for 'tree' first followed by the fingerspelling'* or *'my sign for 'mother' looks like this'*. When talking about multiple signs for the same concept, we pick out a distinguishing feature across the signs – e.g. *'Some people sign 'mother' at the side of the forehead but I sign it on the palm'*.




## 6 Appendix A: Work-in-progress basic template in ELAN as of March 2017

Below is a table showing the full set of tiers of the basic template applied to all ELAN files when initially created, for work-in-progress. The ordering of tiers follows the hierarchy of tiers within the basic template. The tiers listed in yellow denote parent tiers. Labels that have been indented represent child tiers. Rows in a lighter shade are child tiers that are, in turn, also parent tiers. Some of these tiers have been annotated for some files; these will be made available online at a later date when they have been more systematically annotated and checked. Additional tiers not listed here are study-specific.









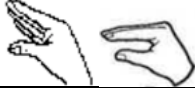


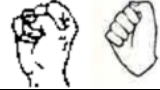



Tier	In use?	CV available?	Definition (unless self-explanatory)
------	---------	---------------	--------------------------------------





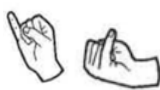



<b>RH-IDgloss</b>	<b>YES</b>		<b>This tier represent all manual material articulated on the right hand</b>
RH-Handshape	YES	YES	
RH-Location	NO		
RH-Movement	NO		
RH-OtherPhon	NO		
RH-Orientation	NO		
RH-Grammatical Category	YES	YES	
<b>LH-IDgloss</b>	<b>YES</b>		<b>This tier represent all manual material articulated on the right hand</b>
LH-Handshape	NO	YES	
LH-Location	NO		
LH-Movement	NO		
LH-OtherPhon	NO		
LH-Orientation	NO		
LH-Grammatical Category	YES	YES	
<b>Clause</b>	<b>YES</b>		<b>Potential clause-like unit</b>
<b>RH-Argument</b>	<b>YES</b>	<b>YES</b>	<b>Arguments produced on the right hand</b>
RH-Animacy	YES	YES	
RH-Coreference	YES	YES	
RH-Person	YES	YES	
RH-Number	YES	YES	
<b>LH-Argument</b>	<b>YES</b>	<b>YES</b>	<b>Arguments produced on the left hand</b>
LH-Animacy	YES	YES	
LH-Coreference	YES	YES	
LH-Person	YES	YES	
LH-Number	YES	YES	
<b>Actor non-present</b>	<b>YES</b>	<b>YES</b>	
Actor-Animacy	YES	YES	
Actor-Coreference	YES	YES	
Actor-Person	YES	YES	
Actor-Number	YES	YES	
<b>Undergoer non-present</b>	<b>YES</b>	<b>YES</b>	
Undergoer-Animacy	YES	YES	
Undergoer-Coreference	YES	YES	
Undergoer-Person	YES	YES	
Undergoer-Number	YES	YES	
Free Translation	YES		
Comments	YES		
Utterance	YES		

## 7 Appendix B: Handshape codes for depicting signs

Proposed BSL code	Proposed BSL forms	Proposed BSL possible meaning (examples)
1		long thin objects (e.g. pen, person)
2		dual objects (e.g. legs, arms)
SPOON		straight, narrow objects (e.g. path, animal ears)



3		long thin objects (e.g. pen, person)
4		long thin objects (e.g. pen, person)
5		many
FLAT		flat objects (e.g. paper), handling of flat or large objects (e.g. pile of towels, box)
O		(handling of) cylindrical or spherical objects (e.g. ball, pole)
CYL		(handling of) cylindrical objects (e.g. pole)
SPHERE		round, bulky or shapeless objects (e.g. ball, pile, town), handling of spherical objects (e.g. ball, apple)
CLOSED		handling of thin flattish objects (e.g. sheet of paper)
OPEN		handling of thick flattish objects (e.g. book)
SMALL_CLOSED		flat round objects (e.g. coin), handling of thin or very small objects (e.g. handkerchief, pin)
SMALL_OPEN		flat round objects (e.g. coin, saucer), handling of small or flattish objects (e.g. sugar cube)
FIST		handling of (thin) objects (e.g. pole)
GRIP		handling of (thin) objects (e.g. fishing rod)
Y		objects with two opposite extensions (e.g. airplane)
A-DOT		small compact object with short upright extension (e.g. vinegar, oil shaker)

GUN		objects with one long and one short extensions at right angles (e.g walking stick, set square)
HOOK		bent or curved objects (e.g sitting with legs apart, bent legs, hooked nose, hearing aid, handling cylindrical objects or objects whose use requires a squeezing action (e.g shotgun, hose, torch, toothpaste, filler)
RUDE		handling of objects with fingertip (e.g. contact lens)
WISH		long narrow objects intertwined (e.g rope, braids)
PINKY		small narrow objects
HORNS		objects with two vertical extensions (e.g television aerial, goalposts)
OPEN-8		touch delicately or swipe
ASL-VEH		vehicles

## 8 References

- Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.
- Brentari, D., & Padden, C. A. (2001). Native and foreign vocabulary in American Sign Language: A lexicon with multiple origins. In D. Brentari (Ed.), *Foreign vocabulary: A cross-linguistic investigation of word formation* (pp. 87-119). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cormier, K., & Fenlon, J. (2014). *BSL Corpus Annotation Guidelines*, v. 1 [http://www.bslcorpusproject.org/wp-content/uploads/BSLCorpusAnnotationGuidelines\\_23October2014.pdf](http://www.bslcorpusproject.org/wp-content/uploads/BSLCorpusAnnotationGuidelines_23October2014.pdf). Deafness, Cognition and Language Research Centre, University College London.
- Cormier, K., Schembri, A., & Tyrone, M. E. (2008). One hand or two? Nativisation of fingerspelling in ASL and BANZSL. *Sign Language and Linguistics*, 11(1), 3-44. doi: 10.1075/sl&l.11.1.03cor
- Cormier, K., Schembri, A., & Woll, B. (2013). Pronouns and pointing in sign languages. *Lingua*, 137, 230-247.
- Fenlon, J., Cormier, K., & Schembri, A. (under review). Building BSL SignBank: The lemma dilemma revisited.
- Fenlon, J., Schembri, A., & Cormier, K. (2014, 8-11 July 2014). *The role of gesture in directional verbs in British Sign Language: a corpus-based study*. Paper presented at the Sixth conference of the International Society for Gesture Studies, San Diego, CA.

- Fenlon, J., Schembri, A., Rentelis, R., Vinson, D., & Cormier, K. (2014). Using conversational data to determine lexical frequency in British Sign Language: The influence of text type. *Lingua*, 143, 187-202. doi: 10.1016/j.lingua.2014.02.003
- Johnston, T. (2013a). Functional and formational characteristics of pointing signs in a corpus of Auslan (Australian sign language): are the data sufficient to posit a grammatical class of 'pronouns' in Auslan? *Corpus Linguistics and Linguistic Theory*, 9(1), 109-159.
- Johnston, T. (2013b). Towards a comparative semiotics of pointing actions in signed and spoken languages. *Gesture*, 13(2), 109-142. doi: <http://dx.doi.org/10.1075/gest.13.2.01joh>
- Johnston, T. (2014). *Auslan Corpus Annotation Guidelines*, [http://media.auslan.org.au/attachments/Johnston\\_AuslanCorpusAnnotationGuidelines\\_14June2014.pdf](http://media.auslan.org.au/attachments/Johnston_AuslanCorpusAnnotationGuidelines_14June2014.pdf). Sydney, Australia: Macquarie University.
- Liddell, S. K. (2003). *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.
- Schembri, A., Fenlon, J., Rentelis, R., & Cormier, K. (2014). *British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008-2014 (Second Edition)*. London: University College London. <http://www.bslcorpusproject.org>.
- Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S., & Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation and Conservation*, 7, 136-154.