

# Notes to the 2<sup>nd</sup> release of BSL Corpus annotations

July 2015

The second release of the BSL Corpus annotations is one of the outcomes of the Digging Into Signs project, a joint effort of University College London (PI Kearsy Cormier) and Radboud University (PI Onno Crasborn) – for details of the project, see Crasborn, Bank & Cormier (2015). This release reflects the work accomplished within this project for the BSL Corpus. The current document serves to describe the published data and indicate the changes with respect to the previous release, and accompanies the annotations and annotation guidelines v.2.1 available on CAVA (<http://www.bslcorpusproject.org/cava>) as of July 2015.

Note that although eaf filenames in CAVA contain a range of codes for demographic information (Region, Participant number, Gender, Age, Ethnicity, Deaf/Hearing family, Task - e.g. BM07F35WHC), the tables in this document provide codes for only the Region, Participant number and Task as shorthand (e.g. BM7c).

## 1 BSL Corpus ELAN eaf files which contain annotations of manual activity

Table 1 shows the files which have been parsed and which have been annotated for RH-ID gloss tiers and LH-ID gloss tiers. In the Birmingham, London, Bristol, and Manchester files, the first approximately 500 tokens of segmentable manual activity from conversation have been glossed while in the Belfast and Glasgow files the first approximately 100 tokens of segmentable manual activity from narrative have been glossed, using the guidelines in sections 4.1-4-14 of annotation guidelines v.2.1, Cormier et al. (2015). Except for single vs double glossing (see section 3 below), all the other changes noted here apply to the files shown in Table 1.

Birmingham	London	Bristol	Manchester	Belfast	Glasgow
BM2c	L3c	BL1c	MC1c	BF1n	GW1n
BM3n-c	L4c	BL2c	MC2c	BF2n	GW2n
BM4c	L5c	BL3c	MC3c	BF3n	GW3n
BM5c	L6c	BL4c	MC6c	BF4n	GW4n
BM6c	L8c	BL5c	MC7c	BF5n	GW5n
BM7c	L9c	BL6c	MC9c	BF6n	GW6n
BM8c	L10c	BL7c	MC10c	BF7n	GW7n
BM9c	L12c	BL8c	MC11c	BF8n	GW8n
BM10c	L14c	BL9c	MC12c	BF9n	GW9n
BM11c	L15c	BL10c	MC13c	BF10n	GW10n
BM12c	L18c	BL11c	MC15c	BF11n	GW11n
BM13c	L19c	BL12c	MC16c	BF12n	GW12n
BM14c	L20c	BL13c	MC17c	BF13n	GW13n
BM15c	L21c	BL15c	MC18c	BF14n	GW14n
BM16c	L22c	BL16c	MC19c	BF15n	GW15n
BM17c	L23c	BL17c	MC20c	BF16n	GW16n
BM18c	L24c	BL21c	MC21c	BF17n	GW17n
BM19c	L25c	BL22c	MC22c	BF18n	GW18n
BM20c	L26c	BL23c	MC23c	BF19n	GW19n
BM21c	L27c	BL24c	MC24c	BF20n	GW20n
BM22c	L29c	BL28c	MC25c	BF21n	GW21n
BM23c	L33c	BL29c	MC26c	BF22n	GW22n
BM26n-c	L35c	BL30c	MC27c	BF23n	GW23n
BM28c	L36c	BL31c	MC28c	BF24n	GW24n
BM29c		BL32c	MC29c	BF25n	GW25n

Table 1. BSL Corpus annotation files which have been ID glossed (approx first 500 or 100 signs), from conversation or narrative, respectively.

## **2 Buoy (see section 4.7 in annotation guidelines v.2.1)**

In the first release, buoys were not annotated very consistently. This has been improved in the second release – list buoys, pointer buoys, fragment buoys and theme buoys have been checked and are now consistent with the v.2.1 guidelines. Also, list buoys have now consistently been given an ID-gloss suffix to indicate the incorporated number sign – e.g. LBUOY-FOUR02.

## **3 Pointing signs (see section 4.8.1 in annotation guidelines v.2.1)**

All tokens of PT:BUOY have been checked to ensure that this label is only used for points to buoys of an ambiguous type (e.g. unclear if list or fragment buoy) and not used for pointer buoys (which are annotated as PBUOY). Annotation conventions for points to the body (PT:BODY) and their relationship to signs for body parts in BSL SignBank have been clarified in the annotation guidelines. Plural pointing signs that incorporate a number sign have been given an ID-gloss suffix to indicate the incorporated number sign – e.g. PT:PRO1PL-THREE02).

## **4 Classifier/depicting signs (see section 4.8.2 in annotation guidelines v.2.1)**

The system for annotating classifier signs has changed, incorporating elements from Johnston (2014) and from Crasborn, Bank, Zwitserlood et al. (2015). The convention from the first release (prefix CL then type of classifier – location, movement, handling, size/shape – followed by description of meaning) has been replaced by slightly different types (prefix DS for depicting sign then type of depicting sign – whole entity, part entity, handling, size/shape), followed by handshape, followed by movement type. We have also introduced Johnston's (2014) type-like depicting signs.

## **5 Annotations of two-handed signs on RH-ID gloss tiers and LH-ID gloss tiers (see section 4 in annotation guidelines v.2.1)**

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 (what we are calling here for shorthand “single glossing”). 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). We are currently moving towards annotating two-handed signs on both the RH and LH tiers, with the start and endpoint determined by the dominant hand, what we refer to here for shorthand as “double glossing”. 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. Additionally, for two-handed signs or for dependent combinations (such as buoys and their counterpart on the dominant hand), for the files noted as “double glossed”, the start and end of the sign follows the dominant hand. Table 2 shows which files have had “double glossing applied and which have had “single glossing” applied.

Birmingham	London	Bristol	Manchester	Belfast	Glasgow
BM2c (s)	L3c (d)	BL1c (s)	MC1c (d)	BF1n (d)	GW1n (d)
BM3n-c (s)	L4c (d)	BL2c (s)	MC2c (d)	BF2n (d)	GW2n (d)
BM4c (s)	L5c (d)	BL3c (s)	MC3c (d)	BF3n (d)	GW3n (d)
BM5c (s)	L6c (d)	BL4c (s)	MC6c (d)	BF4n (d)	GW4n (d)
BM6c (s)	L8c (d)	BL5c (s)	MC7c (d)	BF5n (d)	GW5n (d)
BM7c (s)	L9c (d)	BL6c (s)	MC9c (d)	BF6n (d)	GW6n (d)
BM8c (s)	L10c (d)	BL7c (s)	MC10c (d)	BF7n (d)	GW7n (d)
BM9c (s)	L12c (d)	BL8c (s)	MC11c (d)	BF8n (d)	GW8n (d)
BM10c (s)	L14c (d)	BL9c (s)	MC12c (d)	BF9n (d)	GW9n (d)
BM11c (s)	L15c (d)	BL10c (s)	MC13c (d)	BF10n (d)	GW10n (d)
BM12c (s)	L18c (d)	BL11c (s)	MC15c (d)	BF11n (d)	GW11n (d)
BM13c (s)	L19c (d)	BL12c (s)	MC16c (d)	BF12n (d)	GW12n (d)
BM14c (s)	L20c (d)	BL13c (s)	MC17c (d)	BF13n (d)	GW13n (d)
BM15c (s)	L21c (d)	BL15c (s)	MC18c (d)	BF14n (d)	GW14n (d)
BM16c (s)	L22c (s)	BL16c (s)	MC19c (d)	BF15n (d)	GW15n (d)
BM17c (s)	L23c (s)	BL17c (s)	MC20c (d)	BF16n (d)	GW16n (d)
BM18c (s)	L24c (s)	BL21c (s)	MC21c (d)	BF17n (d)	GW17n (d)
BM19c (s)	L25c (s)	BL22c (s)	MC22c (d)	BF18n (d)	GW18n (d)
BM20c (s)	L26c (s)	BL23c (s)	MC23c (d)	BF19n (d)	GW19n (d)
BM21c (s)	L27c (s)	BL24c (s)	MC24c (d)	BF20n (d)	GW20n (d)
BM22c (s)	L29c (s)	BL28c (s)	MC25c (d)	BF21n (d)	GW21n (d)
BM23c (s)	L33c (s)	BL29c (s)	MC26c (d)	BF22n (d)	GW22n (d)
BM26n-c (s)	L35c (s)	BL30c (s)	MC27c (d)	BF23n (d)	GW23n (d)
BM28c (s)	L36c (s)	BL31c (s)	MC28c (d)	BF24n (d)	GW24n (d)
BM29c (s)		BL32c (s)	MC29c (d)	BF25n (d)	GW25n (d)

Table 2. BSL Corpus annotation files which have been “single” (s) versus “double” (d) glossed.

## Notes to the 1<sup>st</sup> release of BSL Corpus annotations – October 2014

These release notes accompany the annotations that were previously available on CAVA (<http://www.bslcorpusproject.org/cava>) as of October 2014.

### BSL Corpus ELAN eaf files which contain annotations of manual activity

Table 3 shows the files which were parsed and annotated for RH-ID gloss tiers and LH-ID gloss tiers. In all of these files, the first 500 tokens of segmentable manual activity were glossed using the guidelines in section 4 of annotation guidelines v.1, Cormier & Fenlon (October 2014).

Birmingham	London	Bristol	Manchester
BM2c	L3c	BL1c	MC1c
BM3n-c	L4c	BL2c	MC2c
BM4c	L5c	BL3c	MC3c
BM5c	L6c	BL4c	MC6c
BM6c	L8c	BL5c	MC7c
BM7c	L9c	BL6c	MC9c
BM8c	L10c	BL7c	MC10c
BM9c	L12c	BL8c	MC11c
BM10c	L14c	BL9c	MC12c
BM11c	L15c	BL10c	MC13c
BM12c	L18c	BL11c	MC15c
BM13c	L19c	BL12c	MC16c
BM14c	L20c	BL13c	MC17c
BM15c	L21c	BL15c	MC18c
BM16c	L22c	BL16c	MC19c
BM17c	L23c	BL17c	MC20c
BM18c	L24c	BL21c	MC21c
BM19c	L25c	BL22c	MC22c
BM20c	L26c	BL23c	MC23c
BM21c	L27c	BL24c	MC24c
BM22c	L29c	BL28c	MC25c
BM23c	L33c	BL29c	MC26c
BM26n-c	L35c	BL30c	MC27c
BM28c	L36c	BL31c	MC28c
BM29c		BL32c	MC29c

Table 3. BSL Corpus annotation files which have been ID glossed (approx first 500 signs), from conversation.

## 6 References

- Cormier, K., & Fenlon, J. (2014). BSL Corpus Annotation Guidelines, v. 1: Deafness, Cognition and Language Research Centre, University College London. [http://www.bsllcorpusproject.org/wp-content/uploads/BSLCorpusAnnotationGuidelines\\_23October2014.pdf](http://www.bsllcorpusproject.org/wp-content/uploads/BSLCorpusAnnotationGuidelines_23October2014.pdf)
- Cormier, K., Fenlon, J., Gulamani, S., & Smith, S. (2015). BSL Corpus Annotation Conventions, v. 2.1: Deafness, Cognition and Language Research Centre, University College London. [http://www.bsllcorpusproject.org/wp-content/uploads/BSLCorpus\\_AnnotationConventions\\_v2.1\\_July2015.pdf](http://www.bsllcorpusproject.org/wp-content/uploads/BSLCorpus_AnnotationConventions_v2.1_July2015.pdf)
- Crasborn, O., Bank, R., & Cormier, K. (2015). Digging into Signs: Towards a gloss annotation standard for sign language corpora. Nijmegen: Radboud University. [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., De Meijer, A., & Sáfár, A. (2015). Annotation conventions for the Corpus NGT. Version 3, June 2015. Radboud Universiteit Nijmegen.
- Johnston, T. (2014). *Auslan Corpus Annotation Guidelines*, Sydney, Australia: Macquarie University. [http://media.auslan.org.au/attachments/Johnston\\_AuslanCorpusAnnotationGuidelines\\_14June2014.pdf](http://media.auslan.org.au/attachments/Johnston_AuslanCorpusAnnotationGuidelines_14June2014.pdf)