

British Sign Language Corpus Project:

sociolinguistic variation, language change, language contact and lexical frequency in BSL

PI: Adam Schembri

CIs: Kearsy Cormier, Margaret Deuchar, Frances Elton, Donall O Baoill, Rachel Sutton-Spence, Graham H. Turner & Bencie Woll



What is a 'corpus'?

- According to McEnery & Wilson (2001):
 - A representative language sample
 - Finite size
 - Machine-readable form
 - A standard reference

Why a corpus of BSL? (1)

- "It is worth stressing that the amount of research that has been undertaken and is being undertaken with respect to BSL is extremely limited...One of the major difficulties facing sign language linguistics is that many of us are working in 'applied' contexts, before theories and descriptions have been fully elaborated: indeed in some areas, work has not even begun" (Mary Brennan, 1994).

Why a corpus of BSL? (2)

- An on-line, open-access corpus of annotated BSL digital video data will become a shared, peer-reviewable resource and standard reference for BSL researchers and teachers
- Full participant consent and metadata (background data about participants etc) will be included
- Will provide an evidence base for BSL research and teaching, helping to answer many questions and begin to ask new ones

What's a corpus good for?

- Phonology/phonetics
- Lexical studies
- Grammar
- Semantics
- Pragmatics
- Sociolinguistics
- Stylistics
- Language teaching
- Historical linguistics
- Psycholinguistics
- Social psychology
- Cultural studies

Using a BSL corpus: e.g., phonology

- Will provide largest ever sample of BSL data
- Data less likely to be subject to production monitoring by signers
- Large scale quantitative analyses will be possible (e.g., frequency of specific handshapes)
- Possible to explore the interrelationship between phonological and other linguistic levels

Using a BSL corpus: e.g., psycholinguistics

- Annotated corpora can provide more concrete and reliable information about lexical frequency
- Garhnam et al.'s (1981) corpus-based study on the occurrence of speech errors in conversational English
- It may be possible to classify and count frequencies of different errors types, as well as estimate overall frequency of 'slips of the hand' in BSL

Using a BSL corpus: e.g., sign language teaching

- Research into BSL will feed into the development of teaching resources for BSL teaching and the training of BSL tutors
- Open-access on-line BSL collection with a mix of signers from different regional backgrounds will be an invaluable resource to be consulted by both tutors and students

Aim of the BSL Corpus Project

- To create a BSL resource that will make corpus-based approaches to the study of BSL structure and use possible

Specific studies:

- (1) Linguistic and sociolinguistic variation and change in
 - (a) phonology (e.g., handshape, location, movement, non-manual features, weak drop etc)
 - (b) vocabulary (50 target lexical items)
 - (c) grammar of BSL (e.g., pro drop, pronoun copy, aspect marking, agreement verbs, manual vs. non-manual negation etc)

Specific studies:

- (2) Sociolinguistic variation and language contact in BSL (e.g., fingerspelling, mouthing, grammatical borrowing)
- (3) Lexical frequency based on the annotation of 100,000 signs

Project team

- 1 principal investigator/project director (2008-10) & 7 co-investigators
- Advisory Group: 9 members of Deaf community
- 1 RA (2008): data collection coordinator
- 1 postdoc RA (2009-10): annotation and analysis for projects on sociolinguistic variation and change
- 8 RAs: deaf community fieldworkers (150-200 hrs each post in 2008)
- 1 RA (0.6, 2009-10): provide written English translation of data
- 1 RA (0.6, 2009-10): annotate 100,000 signs
- 1 PhD studentship: language contact
- 1 PhD studentship: sociolinguistic variation and change
- 1 project technician (0.5, 2008-10)

Methodology

- Film 30 deaf native and near-native signers (BSL exposure by 7 years of age) in 8 regions across the UK:
 - England (London, Bristol, Birmingham, Manchester, Newcastle)
 - Wales: Cardiff
 - Scotland: Glasgow
 - Northern Ireland: Belfast
- Sample balanced for age, gender, language background, social class and possibly ethnicity



Recruitment & data collection

- Deaf community fieldworkers (cf. 'contact people') will recruit 240 (30 x 8 regions) participants that match project criteria
- Filming over a 10 day period
- No hearing people present
- Pairs of signers matched for region, age and social class
- All exposed resident in their region for 10 years at least
- Participants of the same or different genders
- 1 or 2 high definition video-camera(s) focused on each participant

Materials

- Phase 1: 240 signers for 2 hours?
- 30-45 minutes free conversation
 - 20-30 minute interview on deaf issues and language attitudes
 - Lexical elicitation task
- Phase 2: 100 native signers for 2 hours?
- Elicited narratives
 - Personal anecdote
 - Elicitation tasks to be decided (e.g., eliciting various key aspects of BSL grammar)

Annotation

- Annotation using ELAN
 - Tagging for specific variables
 - ID glosses for 100,000 signs
 - Written English translation
 - Coding for social and linguistic factors
- Annotation of the remainder to be shared, on-going task by certified annotators?



Open access archive

- 300+ hours of BSL digital video data
- Accompanying written English translation
- Accompanying metadata (data about the data)
- ELAN annotation files to be made publicly accessible via ANNEX (new software which would mean you would not need ELAN to see the files)?

BSL Corpus Project: Acknowledgements

- Thanks to the following researchers whose work influenced our research design:
 - Trevor Johnston (Australia)
 - Ceil Lucas (USA)
 - David McKee & Graeme Kennedy (New Zealand)
- Thanks to my project co-investigators
