Multiple access levels and multiple interfaces of the British and Australian Sign Language Corpora

Kearsy Cormier, Jordan Fenlon, Ramas Rentelis, Adam Schembri & Trevor Johnston
University College London; LaTrobe University, Melbourne; Macquarie University, Sydney

“Building Sign Language Corpora in North America” workshop, Gallaudet University
May 21-22 2011

Overview

• Background to BSL and Auslan Corpus Projects
• Access: How, when, what and for whom?
  • Planning and monitoring open vs restricted access
  • Multiple interfaces and multiple levels of access of BSL and Auslan corpora
General aims of the BSL and Auslan Corpus Projects

- To create an on-line, open-access corpus of annotated BSL/Auslan digital video data that will become a shared, peer-reviewable resource and standard reference for BSL/Auslan researchers and teachers. Full participant metadata (background data about participants etc) will be included.

- BSL online by June 2011. Auslan by August 2011

Corpus content

**BSL**
- Personal experience narratives
- Free conversation
- Interview on Deaf issues and language attitudes
- Vocabulary elicitation task

**Auslan**
- Personal interview (identification & background info)
- Spontaneous narratives produced during free conversation
- Responses to face-to-face questionnaire on issues relating to the Deaf community
- Grammatical elicitation tasks
- Re-tells of prepared narratives based on previously read text
- Descriptions of a cartoon viewed on screen
Access and consent: in planning stages

- Signers might monitor their signing more carefully if they have consented for their data to be fully public (Schembri 2010)
- Participants from both projects consented to some of the activities being publicly available, but some of the activities would be restricted to researchers (upon signing a confidentiality agreement)
  - BSLCP restricted data: Conversation data
  - Auslan restricted data: Interview and questionnaire data


Monitoring “open access” data

- Unexpected issues may arise in viewing/annotating/translating the corpus which may require sensitive handling
- Interview data in BSLCP
  - One question: “What does ‘good’ or ‘correct’ BSL look like? Who uses the ‘best’ sign language in the Deaf community?”
  - Expected responses: characteristics of people with BSL skills that are seen as good – e.g. native signers, BSL teachers, etc.
  - Actual responses included names of individual people in Deaf community with “good” BSL, and also names of groups and individuals with “bad” BSL
  - BSLCP interview data is now restricted, along with conversation data
- Similarly Auslan conversation data is now restricted due to sensitive material
BSL corpus: Two interfaces

www.bslcorpusproject.org
• Aimed mainly at non-academics, including Deaf community
• Allows viewing of open access BSLCP data (personal narratives and vocabulary elicitation)
• Simple, easy to navigate, Deaf-friendly interface
• Browsable via region, activity/task, age group
• As of 2011, will consist of video files only

CAVA
• UCL human Communication Audio-Visual Archive
• Aimed mainly at academics (researchers, teachers)
• Allows viewing and downloading of all BSLCP data
• Restricted data access via user license (requires confidentiality agreement)
• Fully searchable via participant/session metadata
• As of 2011, will consist of video files only
• Annotations/translations to be added as they are completed and checked

http://www.bslcorpusproject.org/
Welcome! Choose below from two different ways in which you can look at the video data from the BSL Corpus.

BSL Corpus Project Video Data

Click here to watch videos of deaf people around the UK talking stories or showing their signs for different things. Recommended for the casual visitor.

Please choose a region

Choose a region by clicking on the map to the right. You will then be able to choose one region and age group to view the videos.
CAVA front page (BSLCP: sample)

Auslan corpus access: SOAS

- Hans Rausing Endangered Languages Documentation Project, School of Oriental and African Studies (SOAS), University of London
- Auslan deposit is still being curated at SOAS
- Basic language resources have been deposited
  - movies
  - metadata
  - IDglossing and translation of small percentage of the deposited material
- Once curated, movies and basic annotation (if available) can be located through metadata and viewed
  - region, age, gender, task/genre
Auslan corpus access: SOAS
http://elar.soas.ac.uk/

Stage | Resource available | Public | Researcher
--- | --- | --- | ---
1 | Movie only (available from August 2011) | ✓ | ✓
2 | Basic annotation file for each movie (ID glosses and translation, available as they are completed) | ✓ | ✓
3 | Detailed eaf annotation files for each movie (available upon application to corpus creator) | (√) | ---

Also depositor level access (full access to all deposited data)

Contact: k.cormier@ucl.ac.uk
Conclusions and recommendations

- Corpus curation and maintenance is long, evolving, cumulative process
  - Notion of “access” needs to be able to evolve, too
  - Changes in access may need to be made along the way
- Making different types of data available in stages allows for access as early as possible
  - More complex annotations available on request (to facilitate collaboration between end users and corpus creators)
- A dual interface design of a sign language corpus allows different ways of accessing and viewing video data for different types of users

Lexical databases, ID glossing & access

- An accessible CORPUS requires an accessible LEXICAL DATABASE.
- Accessibility of both corpus and lexical database to multiple research groups requires common ID gloss system
## Acknowledgements

### BSL Corpus Project
- UK Economic and Social Research Council grant RES-629-28-6001 (Schembri, Cormier, Deuchar, Elton, O’Booill, Sutton-Spence, Turner & Woll)
- BSLCP Deaf community fieldworkers & Deaf community Advisory Group members
- Kyra Pollitt and BSL translation team
- Sally Reynolds for data collection
- Mick Canavan and Rick Hurst (Eyegaze) for BSL Corpus Project website, Matt Mahon and Martin Moyle for CAVA website
- The 249 deaf signers across 8 regions of the UK who participated in the BSL Corpus Project data collection

### Auslan Archive & Corpus
- Hans Rausing Endangered Languages Documentation Project grant #MDP088
- Australian Research Council grants #LP09346973 (Schembri, Johnston), #DP0665254 (Johnston, de Beuzeville), #DP1094572 (Johnston)
- Co-researchers, research assistants or annotators who have contributed to current body of corpus annotations since 2003—Julia Allen, Donovan Cresdee, Karin Banna, Michael Gray, Dani Fired, Della Goswell, Ben Hatchard and Gerry Shearim—and to current postgraduate students who have also contributed to corpus annotations—Lindsay Ferrara, Gabrielle Hodge, and Michael Gray.
- The hundreds of deaf signers who contributed material to the Auslan archive and corpus.