



From open access signed language corpora to deaf digital libraries

The Sign Linguistics Corpora Network



Onno Crasborn

Centre for Language Studies, Radboud University Nijmegen

Corpus construction: option 1

Data are newly recorded for the corpus to create a well-balanced representative sample of the language.

Advantages: quick & clean

Disadvantage: it's still elicited to some extent

Corpus construction: option 2

Sign videos are harvested from various online sources. Selection of useful data is done during the research process, not at the data collection stage.

Advantage: no need for elicitation (just as for text corpora): the amount of online video is increasing quickly.

Disadvantage: a lot of variation without proper metadata; L1 and L2 signing will be mixed. Potential copyright issues.

Deaf digital libraries

The two types of data can be combined. Detailed metadata descriptions allow for a wide range of search possibilities and the creation of different portals for different user groups.

Data are not only accessible to researchers, but also to the language community itself (including L2 learners).

In terms of research, the same signed language corpora may be of interest not only for linguistics, but also for deaf studies, pedagogy, and other fields.

The SLCN workshops aim to explore evolutions such as the one sketched above, by looking at what researchers currently do and what people are missing or dreaming about.



www.signlanguagecorpora.org

The Sign Linguistics Corpora Network

2008-2011

- A three-year networking grant of the Netherlands Organisation for Scientific Research (NWO)
- Four workshops and a wiki
- Several partners co-organising the workshops:
 - Graham Turner (Edinburgh)
 - Jens Heßmann (Magdeburg)
 - Eleni Efthimiou (Athens)
 - Johanna Mesch (Stockholm)
 - Adam Schembri (London)
 - Thomas Hanke (Hamburg)
 - Ernst Thoutenhoofd (Groningen)
- EU grant application in the near future

Workshop 1: data collection

Existing and ongoing corpus projects use the same approach: dialogues are elicited, containing narratives and discussions.

July 2009, London

Workshop 3: annotation

There is very little if any standardisation in any annotation domain. Publishing well-documented practices is crucial.

June 2010, Stockholm

Workshop 2: metadata

The sign language profile for the IMDI metadata standard is still valid and needs only little modification.

November 2009, Nijmegen

Workshop 4: exploitation

You?

Everybody is welcome. A call for posters will be distributed soon.

December 3+4, 2010, Berlin

