Vad tecknas och hur?
Korpusar inom teckenspråksforskning och språkutbildning

What is signed and how?
Corpora in sign language research and language teaching

Dr Johanna Mesch
Department of Linguistics, Sign Language Section
Stockholm University
Seminar “How can we use a sign language corpus?”
December 12, 2013, Copenhagen

What are the different needs from a corpus?
- I am presenting about our corpora in sign language research and language teaching at the Department of Linguistics, Stockholm University

What is a sign language corpus?
- Machine-readable form
- Annotation tool ELAN, developed at the Max Planck Institute for Psycholinguistics in Nijmegen
- Authentic
- A representative language sample
- Representative - region, sex, age group, education
- Video files
- Annotations, e.g. glosses and translation
- Accessibility

Swedish Sign Language Corpus
- The Swedish Sign Language Corpus (2009-2011) consists of recordings of 42 informants, male and female, ranging from 20 to 82 years of age and from different regions in Sweden. Financed by Riksbankens Jubileumsfond.
- 25 hours of video, 299 files (totally 167 hours of seven camera views, 2093 files)
- 19% (58 files) of informal signed dialogue/narrative has been annotated with glosses and translated in Swedish so far (Nov 2013).
- 38% (113 files) only translated in Swedish.
**Text types: conversation and narrative**

- **Presentation/introduction**
  Each participant is asked to present him/herself (name, born, education, work title)

- **Conversation**
  Various issues, eg. Communication with family members, the use of sign language in schools and work places, deaf club, activities, events

- **Narrative**
  Some memorable events

- **Elicited narrative – picture book**
  One informant reads a picture story "Frog Where Are You?" in 15 min and retells to the other. Another informant reads a picture story "Snow men".

- **Elicited narrative – cartoon**
  Each informant watches three illustrated (different) cartoons and tells to each other.

- **Elicited narrative – movie**
  Each informant watches a 5-minute selected movie from “Plankan” and “Mr Bean” (2 of 42 “Pingu”). They retell it to each other.

- **Free conversation**
  About TV program, movie or other issues

---

**Recording**

- The video studio with five video cameras at the Department of Linguistics, Stockholm University
- Two-hour sessions: dialogue and elicitation
- Recordings 2 hours x 21 pairs
- Editing and coding of video material
- 7 video files per conversation issue to select for annotation work in ELAN

  S1_b (body)
  S1_f (face)
  S1_t (top incl two variants)
  S2_b (body)
  S2_f (face)
  S2_t (top incl two variants)
  S1_S2_p (pair)

---

**Upper cameras**

---

**Region, age, sex**

- Svealand 24
- Norrland 4
- Götaland 14
- Total: 50:23:30:00:40:45:50:45:60:70

- Women 20
- Men 22

---

**Education**

- Deaf school in Stockholm 9
- Special Class/mixed class 6
- Integrated 1
- Deaf school in Örebro 6
- Deaf school in Vänersborg 8
- Deaf school in Lund 5

---

**Not only recordings but also informed consent**

- SLLS Ethics Statement for sign language research
- Vetting the ethics of research involving humans
Annotations: glosses and translations

Potential uses of an annotated corpus

The corpus shall serve several purposes, for instance:
- Lexicography
- Language teaching (as elements in teaching)
- Language learning (as self-study material)
- Grammatical description (as data for SL research, expanding earlier studies)
  - Cf. introspection, small amount of material, elicitation
  - Studies with corpus data nowadays and future
- Lexical frequency
- Cultural analysis

Lexicography and corpus

Svenskt teckenpråkslexikon online 2008—
Svensk teckenpråkkorpus 2012

Token Variants Transcription Glosa

<table>
<thead>
<tr>
<th>Token</th>
<th>Variant</th>
<th>Transcription</th>
<th>Glosa</th>
</tr>
</thead>
<tbody>
<tr>
<td>arbeta</td>
<td>arbet</td>
<td>Rain</td>
<td>Rain</td>
</tr>
</tbody>
</table>
Using corpus data for teaching purposes

- The Sign Language Section has two kinds of courses in and about Swedish Sign Language: Swedish Sign Language (focusing on sign linguistics) and Swedish Sign Language for beginners.
- Our first three-year bachelor’s programme in sign language and interpreting started in September 2013.

Feedback under building the corpus

- Using corpus data for teaching while the corpus is still under construction means students can give feedback on:
  - how well suited the glosses are
  - search problems they encounter
  - what type of discourse is best suited for teaching
- It also gives teachers access to teaching materials

Corpus data for students

- The corpus data is available for students’ essays (BA & MA level).
  - Börstell 2011; Mårtensson 2012; Sjons 2013
- Using corpus material students will not have to spend a lot of time filming, getting permissions to use the material, annotate and gloss the material, etc.
- They will “only” have to analyze the chosen material in accordance with their research questions.

Corpus data for teachers/instructors

- Teachers can search the corpus for usage examples of the kind of structure that a particular lecture will present
- Ready made, glossed examples to show as a complement to printed materials
- Film clips given to students for assignments and as parts of exam questions

Corpus data for students in interpreting

- Introduction to ELAN
  - can later be used to analyze own interpreting
- Gloss and annotate signed discourse + describe in Word-document
  - different discourse types
  - also dialogue material to analyze backchannel signals
  - interpreted material to compare source and target message
- Students send eaf files to each other
- A teacher merges tiers for discussion
All students transcribed the same: HOW

All students transcribed differently:

move the spoon thoughtfully

Issue: why 'thoughtfully'?

OPEN - TIN - poly

PM - STIR IN SOUP

PICK - IN FOOD

PICK hand sideways semi-circular movement

PICK WITH FORK

Pinch hand sideways

Semi-circular movement

Pedagogical corpus

Planning a pedagogical corpus

- We also offer courses in Swedish Sign Language for beginners/students in interpreting
- Large need for adapted materials for these courses
- We are now producing materials especially suited for beginners, and are planning to include this material in a "pedagogical corpus".

Some research studies with corpus data

- Backchannel signals (SSLC)
- Mouth spreading (ECHO)
- Mouthings (SSLC)
- Reduplication (SSLC, ECHO and other materials)
- Non-dominant hand (SSLC)
- Turn taking (TSSL)

JA@b (YES-fingerspelling) as backchannel signal

Mesch, Nilsson & Wallin, submitted
DELIMIT

INLAND SHORELINE OCEAN EXTENT

DELIMIT

SSLC02_169

Mesch & Wallin, 2013

Data in SSL Corpus with 19 signers, 22 eaf-files, 80 DELIMIT tokens

Ongoing projects

Tactile sign language corpus

Corpus data of 7 deafblind signers. Now editing the materials.

Mesch & Wallin, 2013

Corpus of L2 learners

We are currently starting a corpus study of L2 learners of signed languages. 17 hearing Swedish students who are learning Swedish Sign Language as a second language. Recordings in September, December 2013, March 2014.

Co-investigators: Mesch & Schönström

References

Tack!
Frågor?
johanna.mesch@ling.su.se
www.ling.su.se/teckensprak