The metadata trick in the Corpus NGT

Inge Zwitserlood, Onno Crasborn

Department of Linguistics / Centre for Language Studies (CLS)
Radboud University Nijmegen

Project funded by the Netherlands Organisation for Scientific Research, 380.70-008

Corpus size

• Participants: 92
• Total duration: 71.5 hours
• Total number of clips: 2375
• 4-6 view angles of each clip (body, face, top)

Good metadata are crucial

• Main aim of the corpus NGT:
  
  use the data for various (linguistic) research purposes

• Different uses of the data require different types of metadata

So metadata should be as detailed as possible

Good metadata are crucial

• Another aim of the corpus:
  
  share the data with others
  
  (researchers, interpreters, teachers, the Deaf public, …)

• Participants want protection of their privacy

So metadata should be restricted as much as possible
Choices we made

• Provide only crucial information about participants for (linguistic) researchers:
  – Age
  – Sex
  – Region (school)
  – Language (skills)
  – Deafness status (Deaf)
  – Age of NGT exposure
  – Handedness

Choices we made

Content of clips is more specific
Choices we made

• Provide only crucial information about participants for others:
  – Signer codes
  – Age range (in decades)
  – Region (school)
We have more information about the participants

- Deafness status of parents, siblings, grandparents
- Sign language use by parents, siblings, grandparents
- Type of education
- School means of communication (Dutch, TC, bilingual)
- Different residences
- Member of Deaf (sports, etc.) clubs
- Deafness status and language use of partner
- Estimation of own language skills

www.corpusngt.nl

Onno Crasborn: o.crasborn@let.ru.nl
Inge Zwitserlood: i.zwitserlood@let.ru.nl