Best Practices for Building a Bi-modal Bi-Lingual Bi-National Corpus of Child Language

Diane Lillo-Martin
University of Connecticut and Haskins Laboratories
Deborah Chen Pichler
Gallaudet University
Ronice Müller de Quadros
Universidade Federal de Santa Catarina

Workshop on Sign Language Corpora: Linguistic Issues
24 July 2009
London, UK

INTRODUCTION

- Building a bi-bi-bi corpus of child language
  - We are interested in the simultaneous development of a sign language and a spoken language
    - How does development in one language affect the other?
    - How is this cross-linguistic influence different for bimodal language pairs?
  - We are building corpora of naturalistic spontaneous production samples from children in the US and Brazil

---

INTRODUCTION

- Similarities to studies of sign acquisition by Deaf children
  - Baker and Woll (2009) offered some best practices for development of sign language acquisition corpora
  - Lillo-Martin and Chen Pichler (2008) overviewed the development of our own corpora on the acquisition of ASL by Deaf children

---

INTRODUCTION

- Additional concerns for a bi-bi-bi project
  - Need to collect and represent data from two languages
  - Need to standardize procedures across labs and across countries
  - Need to understand special sociolinguistic situation of bimodal bilinguals
  - Need to understand target languages as used by bimodal bilingual adults

---

1. COLLECTING THE DATA

- General considerations
  Use what we know about building a corpus of data from
  - Spoken language
  - Sign language
  - Children
- This gives us guidance in areas such as participant factors, data collection factors, etc.

Special Considerations

- Participant selection
  - Kodas: At least one parent Deaf; both fluent signers
  - CI's: Exposure to SL by the age of 2; implanted by the age of 2
  - At least 10 hours per week exposure to each language

Special Considerations

- Who does the data collection?
  - The families are generally involved, to different degrees
  - Information for families on creating a good sample

Special Considerations

- Aiming for samples of each language
  - We use different interlocutors on alternating sessions – Deaf signers for the SL sessions; Hearing for the OL
  - Petitto et al. (2001) argued that young bilingual children are sensitive to the language used by their interlocutor
  - Genesee and colleagues have made similar observations for unimodal bilinguals
- However, code-blending is still very common

Code Blending

- The use of sign language and spoken language at the same time
  - NOT THE SAME AS SIM COM
  - Similar to code-switching as used by unimodal bilinguals in some ways
- Code blending is very common in adult and child bimodal bilinguals
  - (van den Bogaerde 2000; van den Bogaerde & Baker 2005; Bishop & Hicks 2008; Emmorey et al. 2008; Pyers & Emmorey 2008)

Code Blending

- Why does code blending occur?
  - Children may see that Deaf parents do have access to aspects of spoken language – may not fully appreciate the need to sign without blending
  - Even for adults, it may be difficult to switch between languages/modes. Emmorey argues that both languages are "on" – and there is greater ability to leave both on (more difficulty to suppress one) for bimodals.

Particular Issues for Bibibi

- Questions on how to define code-blending
  - Mouthing / voice
    - Must an utterance have full voice on to be a blend?

    (Van den Bogaerde & Baker vs. Petitto)
  - Timing
    - Must the spoken and signed utterance match in timing?

    (Emmorey et al. 2008; Pyers & Emmorey 2009)

2. TRANSCRIPTION

- General Principles
  - (Johnston 2001, Johnston & Schembri 2007; Miller 2001; Pizzuto & Pietrandrea 2001)
  - A machine-readable record of language samples
    - Not necessarily sufficient for reader to reproduce in exactly the same way
    - Important that the records can be searched to find all occurrences of phenomena of interest
  - Multiple annotation parses focusing on different phenomena
  - Documentation of data behind analysis decisions
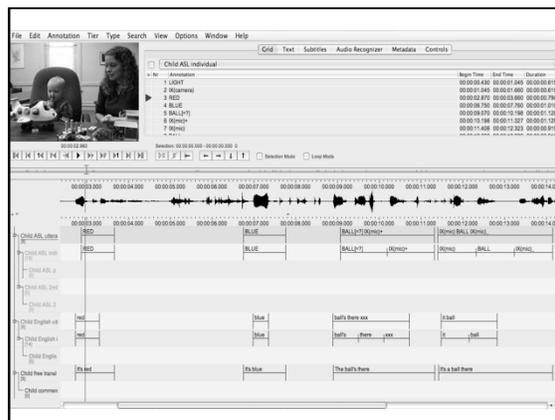
Details of our Approach

- ELAN

http://www.lat-mpi.eu/tools/elan/

Crasborn & Sloetjes (2008)

- It's the cat's pajamas!

http://www.jennilesart.com/angelcats/images/
cats_pajamas_web.jpg

Details of our Approach

- Transcription conventions
  – Where possible, we follow the CHILDES conventions established for child language data in transcribing both speech and sign (though we do not use BTS)
  http://childes.psy.cmu.edu/manuals/chat.pdf
  – When the CHILDES conventions conflict with our sign-specific goals, we create new conventions to be followed for transcribing both sign and speech
  – It is important to keep the sign and speech transcriptions comparable

Details of our Approach

- ID glossing
  – It is our goal to use ID glosses in our transcripts – for reasons clearly justified by Johnston and colleagues
  – Complete use of ID glosses requires a lexicon and means for adding to it

Details of our Approach

- ID glossing
  – Brazil:
    - We started with the on-line Dicionário da Língua Brasileira de Sinais www.acessobrasil.org.br/libras
    - We created a special folder on the Brazilian group's server for additions to this list
      – Make a clip of the sign from tapes or a new recording
      – Use commonly accepted ID gloss if available
      – The team discusses and agrees on all new glosses

Details of our Approach

- ID glossing
  – Brazil:
    - Because of the limitations of the Dicionário, we are building our own glossary
      – Computerized database
      – Will be searchable by handshape and/or location
      – Will be made available to users

Details of our Approach

- ID glossing
  - US: We are working with colleagues to develop an ID-gloss database for ASL. Meanwhile, we use many common sign labels, striving for consistency. Our conventions list gives guiding principles for cases which are commonly variable.

Details of our Approach

- Cross-site comparability
  - US / Brazil
    - Same selection criteria, approach to data collection
    - Same ELAN template
    - Same general transcription principles
    - Working together regularly
  - Two US labs
    - Joint lab meetings
    - Shared documents through Google Docs
    - Common server for files

Held signs	Add [_] (underscore) to end of gloss	MOTHER[_]
Pause within utterance	Represent pauses with a single hatch mark (#)	IX(self) CHOOSE # RED
Interruption	Add [//] to end of last sign before interruption	WANT[//]
Trailing off	Add [...] to end of last sign before trailing off	WANT[...]
Gesture	Gloss with g followed by concise meaning in parentheses	g(angry face)
Mouthing	Gloss with m followed by word mouthed	m(okay)
Sign is not clear (but transcriber is fairly confident of meaning)	Add [?] to end of gloss in question; add description in phonological tier if necessary	WANT APPLE[?] PLEASE
Sign is not clear (transcriber offers alternative gloss)	Type best guess first, followed by [=?ALTERNATIVE]	WANT APPLE[=?ONION]
Sign is not clear (transcriber does not know meaning but can describe form)	*Each* unclear sign in an utterance is glossed as YYY (there may be more than one). Add description in phonological tier of each YYY gloss.	WANT YYY PLEASE
Sign is not clear (transcriber does not know meaning and cannot describe form)	*Each* unclear sign in an utterance is glossed as XXX (there may be more than one)	WANT XXX PLEASE
Sound effects	Sounds such as cries, laughs, and whistles are indicated with &=sound	&=cries &=laughs
	Sounds imitating another person, animal, or machine are indicated with &=imit:sound	&=imit:baby &=imit:plane

3. ANALYSIS

- Bilingual issues
  - How does development of one language influence the other?
    - Detailed analysis of development of particular components of each language
    - Examination of blended structures

Multicyclicity in Early ASL
Previous reports

- Three Deaf ASL signers 0;9 - 1;5 show strong preference for multicyclicity (Meier et al. 2008)
  - Multicyclic signs represent the majority of spontaneously produced signs (over 75%)
  - Both multicyclic and monocyclic targets tend to be produced with reduplication
    - Multicyclic targets correctly reduplicated (81-93% of time)
   - Monocyclic targets also reduplicated (50-80% of time)

Multicyclicity in Early ASL
Application to koda development

- Does bimodality affect this tendency towards multicyclicity in koda ASL production?
  - Syllable structure susceptible to crosslinguistic transfer (Paradis 2001)
  - 1 koda child (Ben) at 1;10 and 1;11

Searching for Cyclicity Errors
Multicyclicity in early koda ASL

Comparing Cyclicity Patterns
Monolingual vs koda ASL
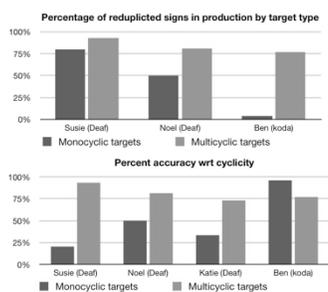
LOOKING FORWARD

- Continue to build and expand our corpus
- Address theoretical questions regarding bi-modal bi-lingual development and comparison with unimodal development in speech or sign only
- Eager to discuss with other teams:
  – new challenges and issues for corpus work
  – Extensions to other areas (e.g. adult coda language use; co-speech gesture)

ACKNOWLEDGMENTS

REFERENCES

- Baker & Woll (2009) *Sign Language Acquisition*. John Benjamins.
- Bishop & Hicks (2009) *Hearing, Mother-father Deaf: Hearing People in Deaf Families*. Gallaudet University Press.
- Crasborn & Sloetjes (2008) Enhanced ELAN functionality for sign language corpora. *Proceedings of LREC 2008.*
- Emmorey et al. (2008) Bimodal bilingualism. *Bilingualism: Language and Cognition*, 11(1), 43-61.
- Johnston (2001) The lexical database of Auslan (Australian Sign Language). *Sign Language and Linguistics 4(1/2),* 145-169.
- Johnston & Schembri (2007) Testing language description through language documentation, archiving and corpus creation: the case of indicating verbs in the Auslan Archive Corpus. *Proceedings of Conference on Language Documentation and Linguistic Theory. London: SOAS. pp. 145-154.*
- Lillo-Martin & Chen Pichler (2008). Development of sign language acquisition corpora. *Proceedings of LREC 2008.*
- Meier et al (2008) The form of children's early signs: Iconic or motoric determinants? *Language Learning and Development*, 4(1), 63–98.
- Miller, C. (2001). Some reflections on the need for a common sign notation. *Sign Language and Linguistics,* 4(1/2), 11-28.
- Paradis (2001) Do bilingual two-year-olds have separate phonological systems? *International Journal of Bilingualism*, 5, 19-38.
- Petitto et al (2001) Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language*, 28, 453-496.
- Pizzuto & Pietrandrea (2001) The notation of signed texts: Open questions for further research. *Sign Language & Linguistics 4(1/2),* 29-45.
- Pyers & Emmorey (2008) The face of bimodal bilingualism: ASL grammatical facial expressions are produced when bilinguals speak to English monolinguals. *Psychological Science*, 19, 531-535.
- Van den Bogaerde (2000) Input and interaction in Deaf families. *Sign Language & Linguistics, 3*(1), 143–151.
- Van den Bogaerde & Baker (2005) Code-mixing in mother-child interaction in deaf families. *Sign Language & Linguistics,* 8(1/2), 151-174.