CHILDES Corpora

• **CHILd Language Data Exchange System**

• Started in 1984 at the Department of Psychology, Carnegie Mellon University

• Collection of 130+ corpora, 20 languages, can have transcribed text, audio, and/or video

• Downloadable CLAN (Computerized Language ANalysis) Program to analyse files in the corpora

• Fully available on: [http://childes.psy.cmu.edu/](http://childes.psy.cmu.edu/)
Transcription Format

• All transcripts available in .cha (chat) format, some available also in (.xml)

• Metadata:
  • Recording date, Language code, Participants (eg. Child, Mother, Investigator), Role
  • @ID: eng|Braunwald|CHI| || |Target_Child| | - CHI is the target child

• Annotation:

  %mor: cojokay.
  *FAT: what's bear doing ? •
  %mor: pro:wh|what~v:aux|be&3S|n|bear|vido-PROG ?
  *CHI: writing a letter , letter •
  %mor: v|write-PROG dt|a nt|letter n|letter .
  *FAT: who is he writing to ? •
  %mor: pro:wh|who|v:aux|be&3S pro|he v|write-PROG prep|to ?
  *CHI: friends . •
  %mor: n|friend-PL
  *FAT: why ? •
  %mor: adv:wh|why ?

  FAT=Father, he is saying “what's bear doing?”
  CHI=Children, saying “writing a letter, letter”

  %mor=morphological tier, list parts of speech
  “n” is NOUN, “PL” is plural, so “friends” is a plural noun

• Also coded: action, phonology, intonation, facial gesture, etc...
Sudden Development of Syntax

• How does a child’s language network develop over time?

• Building network from corpus:
  
  a. CHI: [Telephone go right here]
     CHI: xxx [need it] [my need it]
     CHI: xxx (...)
     CHI: [Put in here]

  b. Telephone
     VP
     go
     right there

  c. My VP
     need it

• Apply on Peters’ corpora in the CHILDES database
  
  • Transcripts of conversations between father and child at various stages of development
Sudden Development of Syntax Cont’d

• Remarkable transition in the structure of the network:
  • Sparse tree-like structure to dense, inter-connected complex network
  • Sudden phase-transition
    • Percolation theory
  • Existing network-growth models unable to account for sudden spurt

25 months old 26 months old 28 months old

Obtained from Corominas-Murta, 2012
References


• Corpora fully available on: http://childes.psy.cmu.edu/
### Example on Peters’ Corpus:

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Peters’ Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Title</strong></td>
<td>Peters’ Corpus</td>
</tr>
<tr>
<td><strong>Resource Type</strong></td>
<td>Speech</td>
</tr>
<tr>
<td><strong>Format/MIME Type</strong></td>
<td>text(CHAT)/audio(mp3)</td>
</tr>
<tr>
<td><strong>Size/Duration</strong></td>
<td>1.7 MB for transcript only</td>
</tr>
<tr>
<td><strong>Access Medium</strong></td>
<td>Online</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>These are transcripts of audiotapes made by Dr. Bob Wilson of himself and his son Seth, who has a severe visual impairment. It was created between February and November 1982 during Seth’s infancy.</td>
</tr>
<tr>
<td><strong>Version</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Media Type</strong></td>
<td>Transcribed Speech and Audio</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>English</td>
</tr>
<tr>
<td><strong>Resource Creation</strong></td>
<td>Bob Wilson, Ann Peters</td>
</tr>
<tr>
<td><strong>Distributor</strong></td>
<td>CHILDES</td>
</tr>
<tr>
<td><strong>Rights Holder</strong></td>
<td>Bob Wilson, Ann Peters</td>
</tr>
</tbody>
</table>