

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/>

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:  co|okay .  
*FAT:  what's bear doing ? ▪  
%mor:  pro:wh|what~v:aux|be&3S n|bear v|do-PROG ?  
*CHI:  writing a letter , letter . ▪  
%mor:  v|write-PROG det|a n|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**

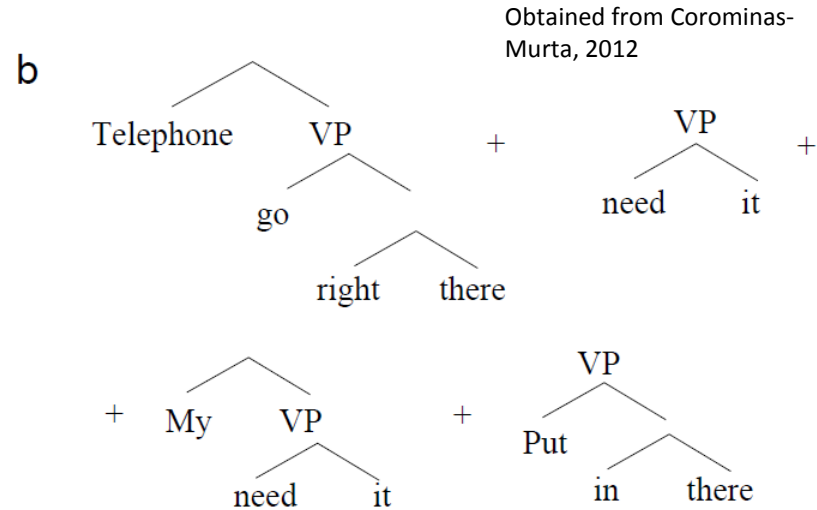
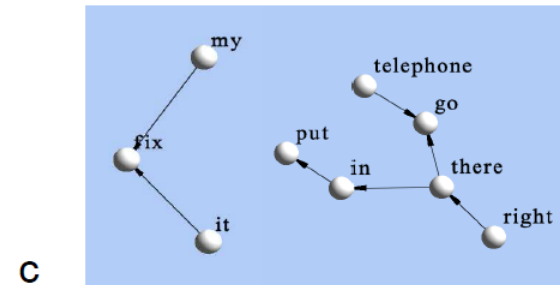
Obtained from Uta Lam, 2007

- 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:

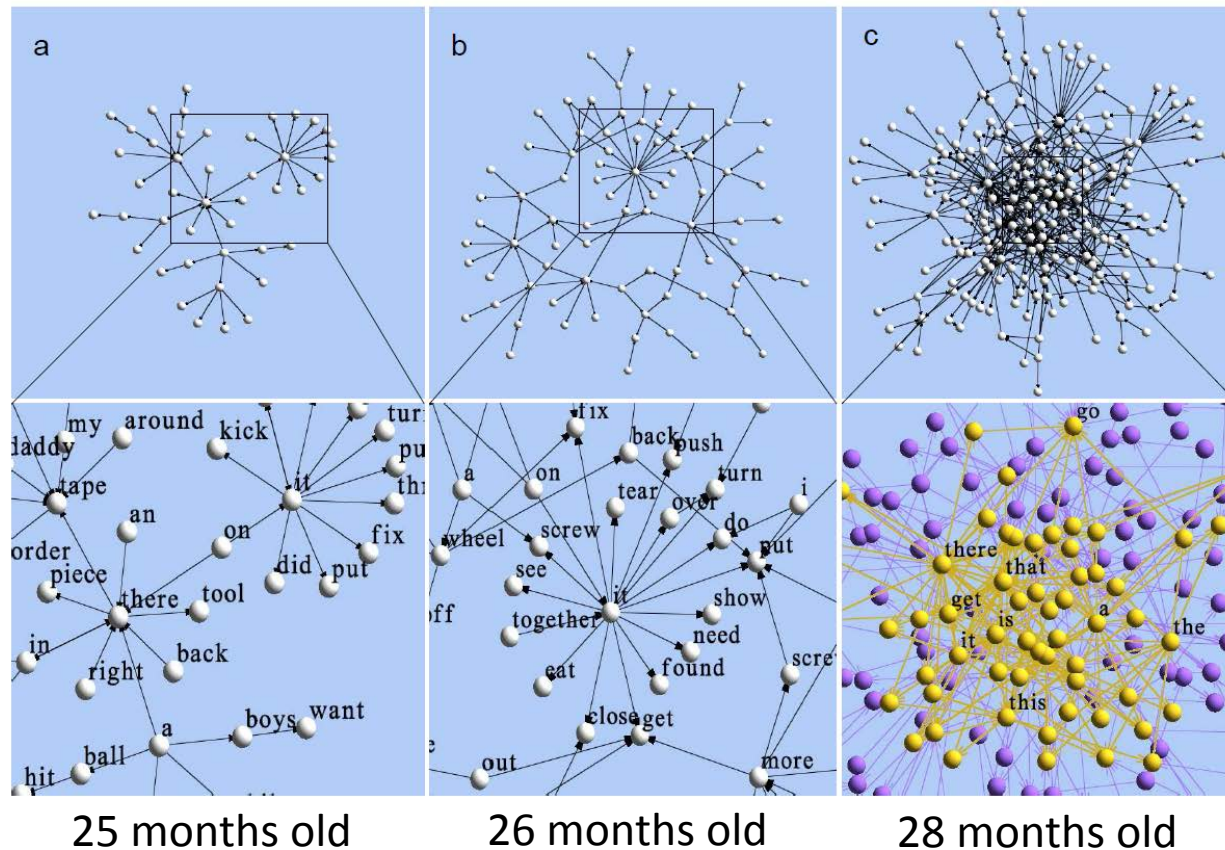
a CHI: [Telephone go right here]
CHI: xxx [need it] [my need it]
CHI: xxx (...)
CHI: [Put in here]



- 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:



Obtained from Corominas-Murta, 2012

- 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

References

- Corominas-Murtra, B.; Valverde, S.; Solé, R.V. The ontogeny of scale-free syntax networks. Evidence for a phase transition in early language acquisition. *Adv Complex Syst* 2009, 12, 371–392.
- MacWhinney, B. *The CHILDES Project: Tools for Analyzing Talk*, 3rd ed.; Lawrence Erlbaum Associates: Mahwah, NJ, 2000.
- Corpora fully available on: <http://childes.psy.cmu.edu/>

ISLRN Metadata

- Example on Peters' Corpus:

Title	Peters' Corpus
Full Title	Peters' Corpus
Resource Type	Speech
Source/URL	http://childes.psy.cmu.edu/browser/index.php?url=Eng-NA-MOR/Peters/text(CHAT)/audio(mp3)
Format/MIME Type	text(CHAT)/audio(mp3)
Size/Duration	1.7 MB for transcript only
Access Medium	Online
Description	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.
Version	None
Media Type	Transcribed Speech and Audio
Language	English
Resource Creation	Bob Wilson, Ann Peters
Distributor	CHILDES
Rights Holder	Bob Wilson, Ann Peters