

GREP

AEAL

Workshop on Phon

Barcelona April 20-21 2009



From CHILDES to Talkbank and PhonBank

Eliseo Diez-Itza

Universidad de Oviedo, Grupo LOGIN

Based on:

From CHILDES to TalkBank

Brian MacWhinney

Carnegie Mellon University

In: Research on Child Language Acquisition

Almgren, M. Barreña, A. Ezeizaberrena, M.,
Idiazabal, I., and MacWhinney, B. (Eds.) (2001)

Cascadilla: Somerville, MA. Pp. 17-34

Origins of Scientific Research on Child Language

- Data collection in diaries 1870
Observational method
Technology: pencil
- Language Testing: large samples 1930
Correlational method
Technology: printing
- Testing psycholinguistic hypothesis 1960
Experimental method
Technology: Tachistoscope

New Technology: Tape Recorders NAGRA

1960

Nagra-brand tape recorders were the *de-facto* standard sound recording systems for motion picture and single-camera television production from the 1960s until the 1990s. (Wikpd)



1961-1963: Adam, Eve and Sarah follow-up study

- Roger Brown: A First Language (1973)

Back to the Observational method

First system for sharing child language transcript data:
mimeographed copies sent by mail

collaborative annotation

Why data-sharing is important

- Increasing the size and reliability of the empirical basis
- Opening science to the community, practitioners, and students
- Opening science to collaborative commentary
- Creating transparency across disciplines

Brian MacWhinney

1981 PC's

Unlimited sharing dream became real

CHILDES Concord Massachusetts 1984



1990 INTERNET

Dream of dreams: instantaneous sharing

And...New Technologies...

- Sonic CHAT
- Bullets
- QuickTime Movies
- Sound editor by wave
- Movie editor by dragging
- Fast mode editing
- Web streaming of audio and video

1999 The Talkbank Project

Looking in the cupboard for tapes

Beyond Lack of Coding Standards, Indeterminacy and Tedium

A direct solution. There is now an effective way of dealing with the threeheaded monster of indeterminacy, tedium, and lack of standards in transcription.

The solution is to use programs that link transcripts and codes directly to the original audio or video data. The idea here is extremely simple. It involves an “end run” around the core problems in transcription.



Key Features of TalkBank

- Multimodal digitized data
- Internet access
- Defense of confidentiality
- Codon: transcription, coding, viewing, and analysis
- XML standard for underlying representation
- Alliance of databases from many fields

Brian MacWhinney

If we have sound...

What about Phonology?

C H I L D E S The PhonBank Project

“It is great to see so much progress in the area of Phonology, which was earlier one of the more difficult areas in child language in terms of lack of precise data.”

Brian MacWhinney