# Advancing the Utility of the Transcript: A Computer-Enhanced Methodology

Methods XII, Moncton, New Brunswick

August 2<sup>nd</sup>, 2005

# Tyler Kendall <tsk3@duke.edu>

Duke University & NC State University

## Project Overview

- Building researcher tool(s) to store, analyze, interact with, and present linguistic data
- Attempting to treat natural speech data like any other type of data.
- Focused initially (version 1) on applications in discourse analysis (transcript-centric) but expanding to be a complete digital audio library and analysis suite (version 2).

## Transcript Features

- User control of the format (e.g., vertical ⇔ columnar ⇔ paragraph ⇔ tablature (Figure 1))
- User control over the amount and kind of data displayed (e.g., translations for multilingual transcripts, syntactic information, nonlinguistic information, time information, pause information...)
- Direct link to Praat phonetic software
  - Utterance-level coding (instrumentally-determined utterances uploadable from Praat TextGrid files)
  - Active link to Praat allows for on-the-fly phonetic analysis and audio clip extraction (Figure 2)

# Toward a Complete Digital Archive

- Through a partnership with the NC State University Libraries
- Working toward a complete digital archive of the North Carolina Language and Life Project's sociolinguistic interviews (> 1,500 tapes)
- Users can access, download, and extract from the audio, listen to the audio online, and can annotate sound files (Figure 3)
- Searchable (e.g., by research site, by demographic information, by transcript text, by annotations on the audio files, etc.)
- Queryable (can generate summary statistics from transcript/annotation information)

#### Selected References

- Edwards, Jane (2001), The Transcription of Discourse, in *Handbook of Discourse Analysis*, eds. Deborah Tannen, Deborah Schiffrin, and Heidi Hamilton, Oxford and Malden, Ma: Blackwell: 321-348.
- Edwards, Jane and Martin Lampert, eds. (1993), Talking Data: Transcription and Coding in Discourse Research, Hillsdale, NJ: Lawrence Erlbaum Assoc, Inc.
- Poplack, Shana (1980), Sometimes I'll Start a Sentence in Spanish y Termino Español: Toward a Typology of Code-Switching, *Linguistics*, 18: 581-618.
- Ochs, Elinor (1979), Transcription as Theory, in *Developmental Pragmatics*, eds. Elinor Ochs and Bambi Schieffelin, New York: Academic Press: 43-72.

## Figures

(1) e.g., vertical format ⇔ columnar format

Towners by columnar format

Towners by columnar format

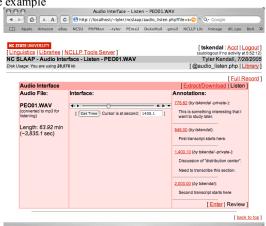
Towners by columnar format

Towners by columnar format

(1) for the columnar format forma



(2) pitch analysis example Analyze line(s) from transcript ▲ A A C http://localhost/~tyler/ncslaap/analyze.php?trans=prvlC ^ ^ Q- Google Apple Amazon eBay NCSU PHPMan -tyler PEmail DukeMail gmail NCLLP Lib Vonage dli\_cpu BoA [tskendal: Acct | Logout] (autologout if no activity at 5:35:38) Tyler Kendall, 7/28/2005 Linguistics II ibraries LNCLLP Tools Server NC SLAAP - Analyze line(s) from transcript [Analyze: ○ Formants | ● Pitch | ○ Intensity | ∪se ImageMap ☑ ] [Show Audio Player ☑ ] [Praat Settings: default | | Pitch Floor: 75 | Pitch Ceilling: 600 | Change Settings | Save Settings ] Data: time = 1029.504 s : pitch = 142.82256551108074 Hz Line Start Snkr Text <162> [1028.849] PEO: It was like a dead smell [1030.684] I duration: 1.835 s Laccess PitchTier file) f back to top 1 (3) listen & annotate example Audio Interface - Listen - PEO01.WAV ▲ ► 🏠 A A C 🕞 http://localhost/~tyler/ncslaap/audio\_listen.php?file=si۞ ^ Q~ Go



<sup>\*</sup> I would like to gratefully acknowledge the financial and infrastructural support of the NC State University Libraries for this project. I would also like to thank Walt Wolfram, Erik Thomas, Amanda French, Phillip Carter, Christine Mallinson, and Drew Grimes for the use of their data and/or for their help and feedback on various stages of this project.