# TalkBank and CLARIN

Brian MacWhinney
CMU - Psychology

send to clarin@clarin.eu by
Oct 21 — 15+10 minutes

# Basic Questions

- How did language emerge in the species?
- How does it change?
- How is it learned?
- How is it processed?
- What are the results of damage and variation?

# Areas

| | | | | |
|---|---|---|---|---|
| **Children:** | CHILDES | PhonBank | Narrative | Bilingual |
| **Clinical:** | AphasiaBank | FluencyBank | Dementia | TBIBank |
| **Adult:** | CABank | TutorBank | GestureBank | ClassBank |
| **Multilingualism:** | BilingBank | SLABank | Online Tutors | DOVE |

# Funded Projects

| | CHILDES | TalkBank | AphasiaBank | PhonBank | FluencyBank | LangBank | HomeBank |
|---|---|---|---|---|---|---|---|
| Age of Project | 28 | 12 | 8 | 6 | 0.2 | 1 | 1 |
| Words (millions) | 59 | 47 | 1.8 | 0.8 | 0.5 | 2 | audio |
| Linked Media (TB) | 2.8 | 1.1 | 0.4 | 0.7 | 0.3 | - | 3.5 |
| # Languages | 41 | 22 | 6 | 18 | 4 | 2 | 2 |
| # Publications | 7000 | 320 | 256 | 480 | - | 4 | 5 |
| # Users | 2950 | 930 | 390 | 182 | 25 | - | 22 |
| # Web Hits (millions) | 4.1 | 1.3 | 0.3 | 0.1 | - | - | |

# 41 languages (including Cantonese)

*MIC: 睇 吓5 Sophie 畫 咗 啲 咩 圖畫 .

%mor: v|tai2=look_at asp|haa5=tentative n:prop|Sophie v|waak6=draw asp|zo2=perfective cl|di1=some wh|me1=what n|+n|tou4+n|waa2=drawing .

*CHI: 呢 啲 我哋 彈 琴 架 .

%mor: sfp|ne1=how_about cl|di1=some pro|ngo5-PL=I v|daan6=bounce n|kam4=piano sfp|gaa3 .

*SIS: 你 鍾意 Alicia 定係 呀 呀 Lulu 定係 Sophie 定係 Timmy ?

%mor: pro|nei5=you v|+v|zung1+n|ji3=like n:prop|Alicia conn|ding6hai6=or sfp|aa3 sfp|aa3 n:prop|Lulu conn|ding6hai6=or n:prop|Sophie conn|ding6hai6=or n:prop|Timmy ?

# TalkBank Principles

- Community Driven
- Open access to data, media, derived corpora, and programs
- Standard format — CHAT, CHAT-XML, CHAT-CA
- CLAN programs running on CHAT format
- Transcripts linked to media
- Interoperable with other resources: R, Elan, Praat, SALT, Annis, CONLL, SpeechKitchen/Kaldi for ASR, LENA
- CHAT/PHON incorporates Praat

# CLARIN Principles in TalkBank

- CLARIN-B center
  - InCommon login through Shibbolet
  - OAI-PMH server for OLAC, VLO
  - DOI through HandleServer, EZ-Cite
- CLARIN-K center
  - focus on analysis of spoken language
  - tutorials for CLAN, video tutorials
  - help desk, 5 Google Groups discussion boards
- SamtaleBank as a CLARIN illustration

# Let's take a look on the web

- childes.talkbank.org
- talkbank.org
- homebank.talkbank.org
- sla.talkbank.org
- talkbank.org/access/SamtaleBank
- childes.talkbank.org/browser — Alicia at 3;3
- downloadable materials

# Major Methods

1. Corpus Analysis

2. Profiling

3. Microanalysis

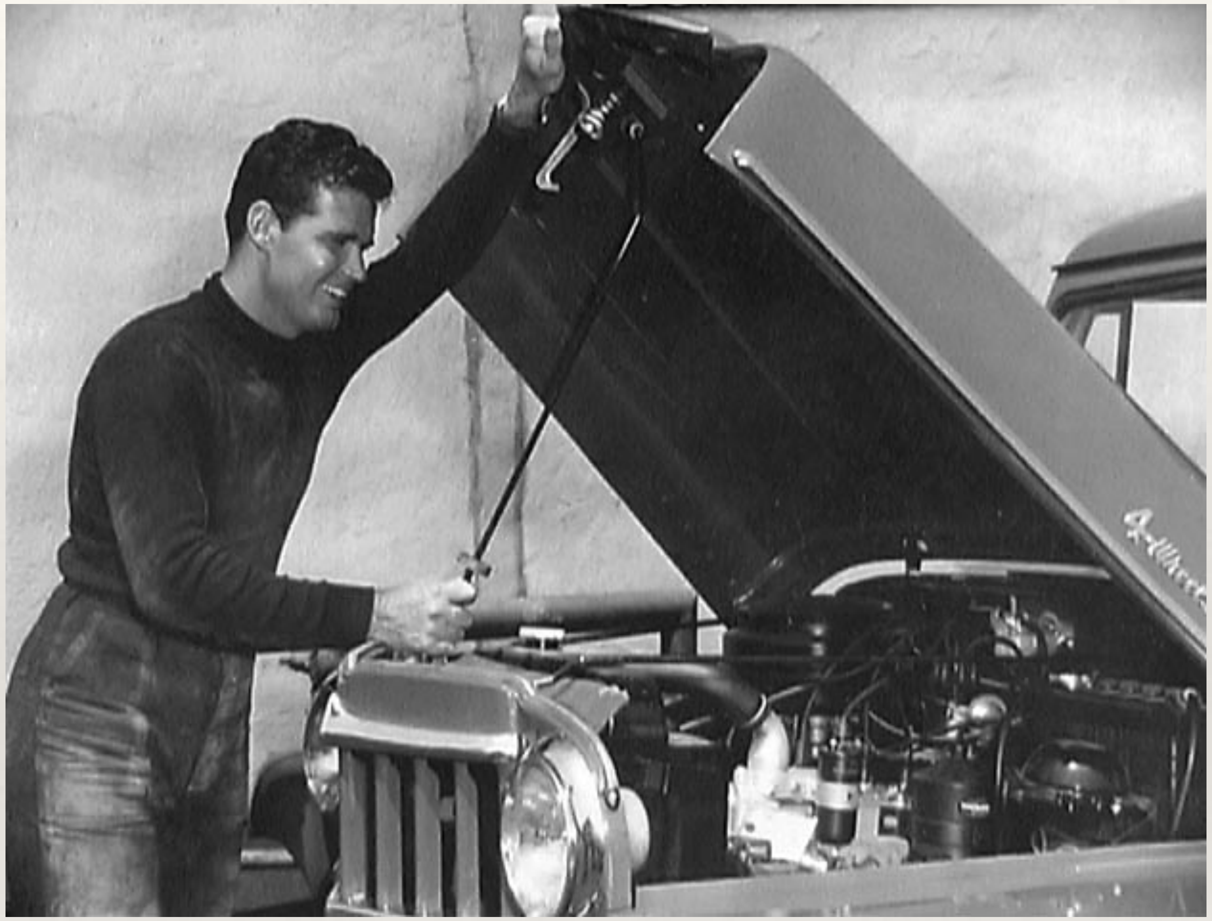
# 1. Corpus Analysis

- FREQ - Frequency analysis
  - wild cards
  - word files (morality words, LIWC, medical)
- KWAL - Key word and line
  - matches highlighted
- COMBO - Regular expression matching
- Hits can be triple-clicked to go back to transcript and play

# LENA2CHAT

- 24 hour/day recordings in the home
- Much like Deb Roy's database and the "water" example, but open
- Huge ITS files reduced automatically to manageable CHAT files
- Check out http://homebank.talkbank.org

# Looking under the Hood



# MOR, POST, GRASP

- 41 languages, but only 11 have MOR/POST
- Cantonese, Danish, Dutch, English, French, Italian, Hebrew, Japanese, German, Mandarin, Spanish
- GRASP for English, German, Hebrew, Spanish, Mandarin

# MOR

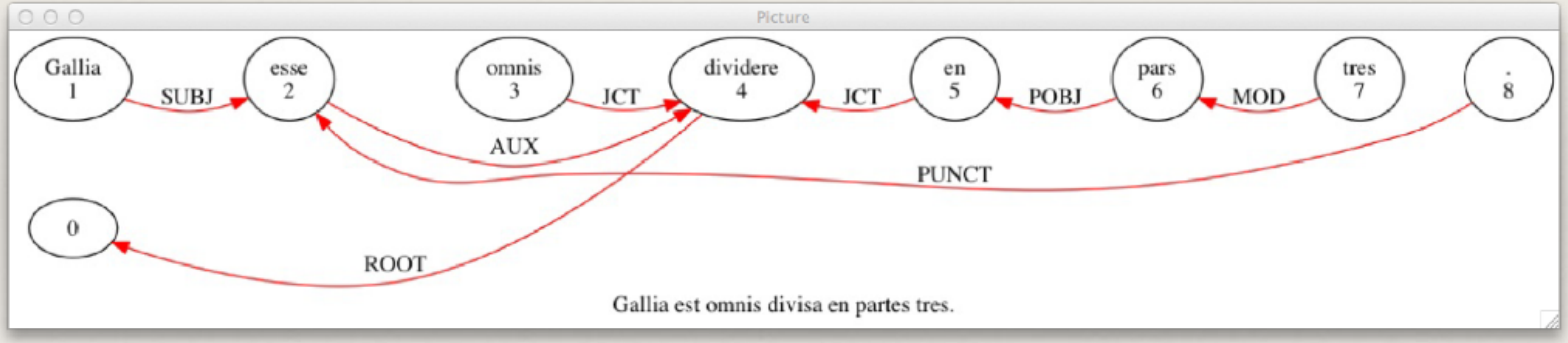
- More declarative than FST
- Part-of-speech tuned to spoken language
- Easy to use once there is a grammar
- Hard to build the grammar (A-rules, C-rules)
- 98% accuracy for English
- POSTMORTEM rules for German declension

# Bilingual MOR

- *CHL:  +" [- spa] <yo no la> [/] yo no la desmentí porque. [+ break]
- *CHL:  what's my word against hers &ladadada .
- *CHL:  +" [- spa] todos estamos con un calor and@s working@s .
- All words are tagged implicity; can be made explicit.
- Coding system makes code-switching junctures evident.
- Run English MOR, excluding [- spa], then Spanish MOR including [- spa]

# Dependency Graphs

Web service runs by triple-clicking on %gra line



Gallia est omnis divisa en partes tres.

# 2. Profiling - EVAL/KIDEVAL

- This all depends on MOR and GRASP
- Comparison database with s.d. scores
- IPSyn, DSS
- MLU, MLT
- TTR, vocD, MATTR
- Brown’s 14 morphemes
- TIMEDUR

# EVAL

MLU, TTR

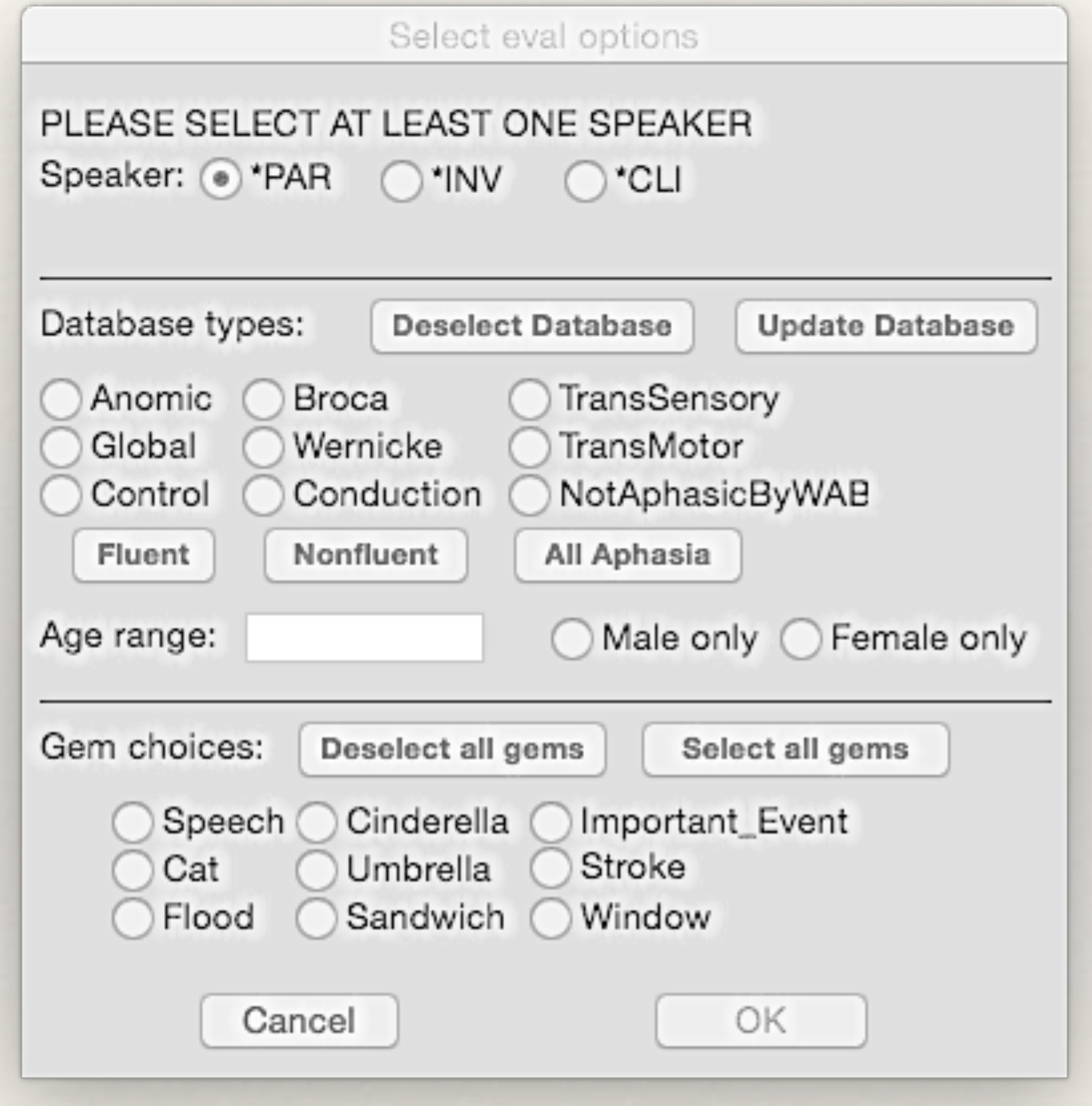
Verbs/Utt

% errors

% N, V, Aux, Adv, Conj, Pro

% PAST, PASTP, PL

Retracing, repetition



Select eval options

PLEASE SELECT AT LEAST ONE SPEAKER

Speaker: ● *PAR ○ *INV ○ *CLI

Database types: Deselect Database | Update Database

○ Anomic ○ Broca ○ TransSensory
○ Global ○ Wernicke ○ TransMotor
○ Control ○ Conduction ○ NotAphasicByWAB

Fluent | Nonfluent | All Aphasia

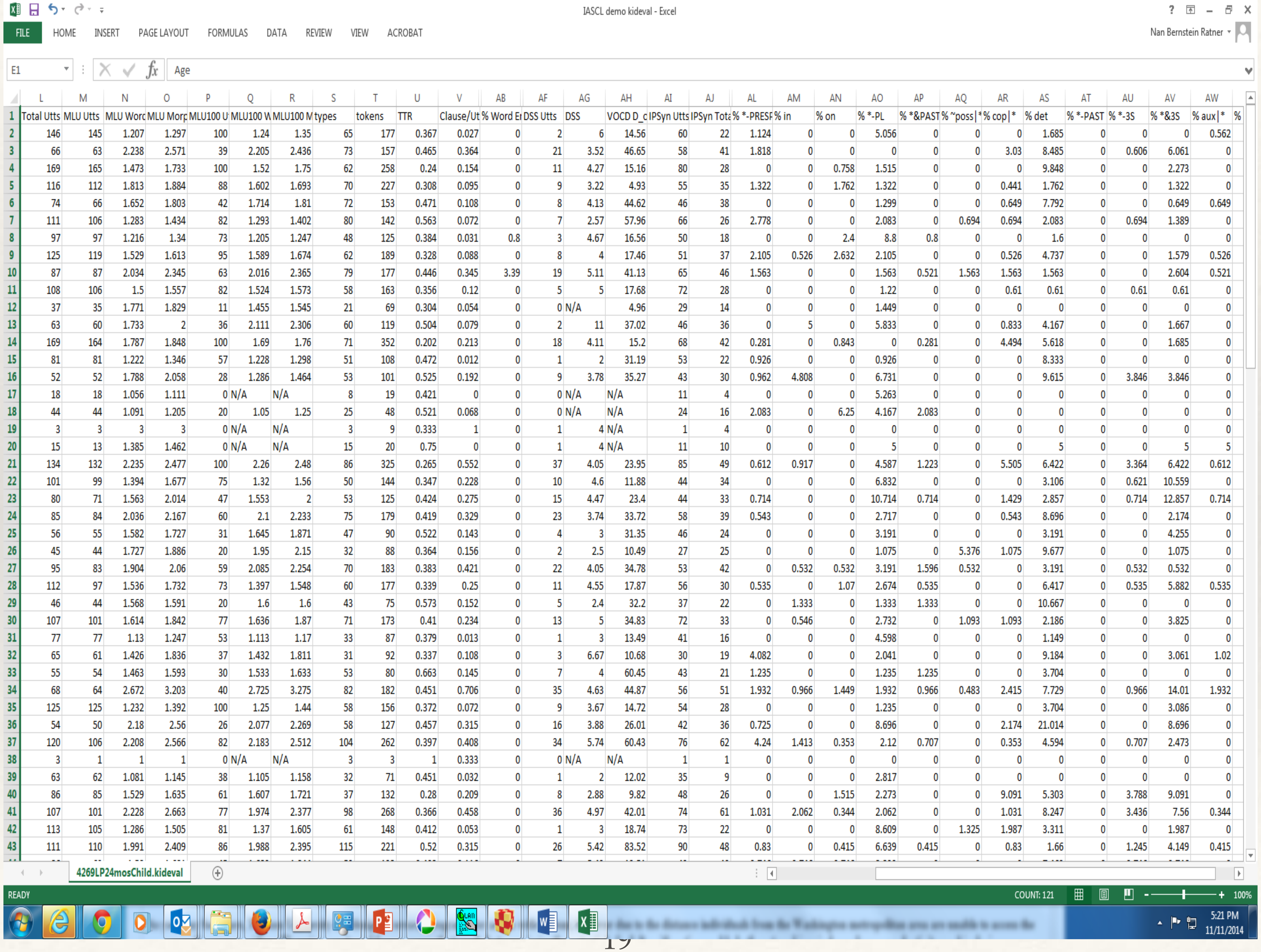
Age range: ○ Male only ○ Female only

Gem choices: Deselect all gems | Select all gems

○ Speech ○ Cinderella ○ Important_Event
○ Cat ○ Umbrella ○ Stroke
○ Flood ○ Sandwich ○ Window

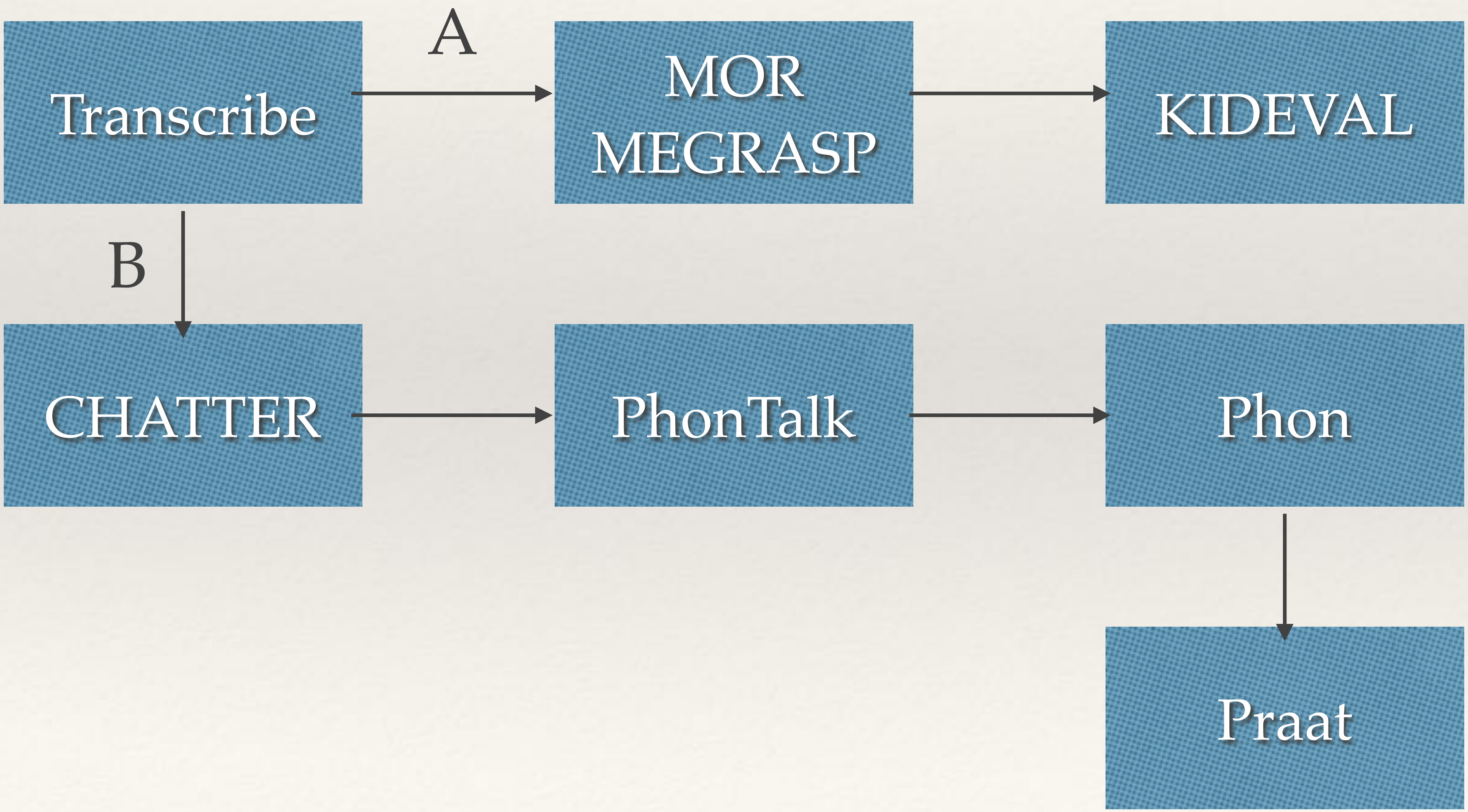
Cancel | OK

# Sample Output



| | L | M | N | O | P | Q | R | S | T | U | V | AB | AF | AG | AH | AI | AJ | AL | AM | AN | AO | AP | AQ | AR | AS | AT | AU | AV | AW | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Total Utts | MLU Utts | MLU Worc | MLU Morp | MLU100 U | MLU100 W | MLU100 M | types | tokens | TTR | Clause/Ut | % Word Er | DSS Utts | DSS | VOCD D_c | IPSyn Utts | IPSyn Tota | % *-PRESF | % in | % on | % *-PL | % *&PAST | % ~poss\|* | % cop\|* | % det | % *-PAST | % *-3S | % *&3S | % aux\|* | % |
| 2 | 146 | 145 | 1.207 | 1.297 | 100 | 1.24 | 1.35 | 65 | 177 | 0.367 | 0.027 | 0 | 2 | 6 | 14.56 | 60 | 22 | 1.124 | 0 | 0 | 5.056 | 0 | 0 | 0 | 1.685 | 0 | 0 | 0 | 0.562 | |
| 3 | 66 | 63 | 2.238 | 2.571 | 39 | 2.205 | 2.436 | 73 | 157 | 0.465 | 0.364 | 0 | 21 | 3.52 | 46.65 | 58 | 41 | 1.818 | 0 | 0 | 0 | 0 | 0 | 3.03 | 8.485 | 0 | 0.606 | 6.061 | 0 | |
| 4 | 169 | 165 | 1.473 | 1.733 | 100 | 1.52 | 1.75 | 62 | 258 | 0.24 | 0.154 | 0 | 11 | 4.27 | 15.16 | 80 | 28 | 0 | 0 | 0.758 | 1.515 | 0 | 0 | 0 | 9.848 | 0 | 0 | 2.273 | 0 | |
| 5 | 116 | 112 | 1.813 | 1.884 | 88 | 1.602 | 1.693 | 70 | 227 | 0.308 | 0.095 | 0 | 9 | 3.22 | 4.93 | 55 | 35 | 1.322 | 0 | 1.762 | 1.322 | 0 | 0 | 0.441 | 1.762 | 0 | 0 | 1.322 | 0 | |
| 6 | 74 | 66 | 1.652 | 1.803 | 42 | 1.714 | 1.81 | 72 | 153 | 0.471 | 0.108 | 0 | 8 | 4.13 | 44.62 | 46 | 38 | 0 | 0 | 0 | 1.299 | 0 | 0 | 0.649 | 7.792 | 0 | 0 | 0.649 | 0.649 | |
| 7 | 111 | 106 | 1.283 | 1.434 | 82 | 1.293 | 1.402 | 80 | 142 | 0.563 | 0.072 | 0 | 7 | 2.57 | 57.96 | 66 | 26 | 2.778 | 0 | 0 | 2.083 | 0 | 0.694 | 0.694 | 2.083 | 0 | 0.694 | 1.389 | 0 | |
| 8 | 97 | 97 | 1.216 | 1.34 | 73 | 1.205 | 1.247 | 48 | 125 | 0.384 | 0.031 | 0.8 | 3 | 4.67 | 16.56 | 50 | 18 | 0 | 0 | 2.4 | 8.8 | 0.8 | 0 | 0 | 1.6 | 0 | 0 | 0 | 0 | |
| 9 | 125 | 119 | 1.529 | 1.613 | 95 | 1.589 | 1.674 | 62 | 189 | 0.328 | 0.088 | 0 | 8 | 4 | 17.46 | 51 | 37 | 2.105 | 0.526 | 2.632 | 2.105 | 0 | 0 | 0.526 | 4.737 | 0 | 0 | 1.579 | 0.526 | |
| 10 | 87 | 87 | 2.034 | 2.345 | 63 | 2.016 | 2.365 | 79 | 177 | 0.446 | 0.345 | 3.39 | 19 | 5.11 | 41.13 | 65 | 46 | 1.563 | 0 | 0 | 1.563 | 0.521 | 1.563 | 1.563 | 1.563 | 0 | 0 | 2.604 | 0.521 | |
| 11 | 108 | 106 | 1.5 | 1.557 | 82 | 1.524 | 1.573 | 58 | 163 | 0.356 | 0.12 | 0 | 5 | 5 | 17.68 | 72 | 28 | 0 | 0 | 0 | 1.22 | 0 | 0 | 0.61 | 0.61 | 0 | 0.61 | 0.61 | 0 | |
| 12 | 37 | 35 | 1.771 | 1.829 | 11 | 1.455 | 1.545 | 21 | 69 | 0.304 | 0.054 | 0 | 0 | N/A | 4.96 | 29 | 14 | 0 | 0 | 0 | 1.449 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 13 | 63 | 60 | 1.733 | 2 | 36 | 2.111 | 2.306 | 60 | 119 | 0.504 | 0.079 | 0 | 2 | 11 | 37.02 | 46 | 36 | 0 | 5 | 0 | 5.833 | 0 | 0 | 0.833 | 4.167 | 0 | 0 | 1.667 | 0 | |
| 14 | 169 | 164 | 1.787 | 1.848 | 100 | 1.69 | 1.76 | 71 | 352 | 0.202 | 0.213 | 0 | 18 | 4.11 | 15.2 | 68 | 42 | 0.281 | 0 | 0.843 | 0 | 0.281 | 0 | 4.494 | 5.618 | 0 | 0 | 1.685 | 0 | |
| 15 | 81 | 81 | 1.222 | 1.346 | 57 | 1.228 | 1.298 | 51 | 108 | 0.472 | 0.012 | 0 | 1 | 2 | 31.19 | 53 | 22 | 0.926 | 0 | 0 | 0.926 | 0 | 0 | 0 | 8.333 | 0 | 0 | 0 | 0 | |
| 16 | 52 | 52 | 1.788 | 2.058 | 28 | 1.286 | 1.464 | 53 | 101 | 0.525 | 0.192 | 0 | 9 | 3.78 | 35.27 | 43 | 30 | 0.962 | 4.808 | 0 | 6.731 | 0 | 0 | 0 | 9.615 | 0 | 3.846 | 3.846 | 0 | |
| 17 | 18 | 18 | 1.056 | 1.111 | 0 | N/A | N/A | 8 | 19 | 0.421 | 0 | 0 | 0 | N/A | N/A | 11 | 4 | 0 | 0 | 0 | 5.263 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 18 | 44 | 44 | 1.091 | 1.205 | 20 | 1.05 | 1.25 | 25 | 48 | 0.521 | 0.068 | 0 | 0 | N/A | N/A | 24 | 16 | 2.083 | 0 | 6.25 | 4.167 | 2.083 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 19 | 3 | 3 | 3 | 3 | 0 | N/A | N/A | 3 | 9 | 0.333 | 1 | 0 | 1 | 4 | N/A | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 20 | 15 | 13 | 1.385 | 1.462 | 0 | N/A | N/A | 15 | 20 | 0.75 | 0 | 0 | 1 | 4 | N/A | 11 | 10 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 5 | 0 | 0 | 5 | 5 | |
| 21 | 134 | 132 | 2.235 | 2.477 | 100 | 2.26 | 2.48 | 86 | 325 | 0.265 | 0.552 | 0 | 37 | 4.05 | 23.95 | 85 | 49 | 0.612 | 0.917 | 0 | 4.587 | 1.223 | 0 | 5.505 | 6.422 | 0 | 3.364 | 6.422 | 0.612 | |
| 22 | 101 | 99 | 1.394 | 1.677 | 75 | 1.32 | 1.56 | 50 | 144 | 0.347 | 0.228 | 0 | 10 | 4.6 | 11.88 | 44 | 34 | 0 | 0 | 0 | 6.832 | 0 | 0 | 0 | 3.106 | 0 | 0.621 | 10.559 | 0 | |
| 23 | 80 | 71 | 1.563 | 2.014 | 47 | 1.553 | 2 | 53 | 125 | 0.424 | 0.275 | 0 | 15 | 4.47 | 23.4 | 44 | 33 | 0.714 | 0 | 0 | 10.714 | 0.714 | 0 | 1.429 | 2.857 | 0 | 0.714 | 12.857 | 0.714 | |
| 24 | 85 | 84 | 2.036 | 2.167 | 60 | 2.1 | 2.233 | 75 | 179 | 0.419 | 0.329 | 0 | 23 | 3.74 | 33.72 | 58 | 39 | 0.543 | 0 | 0 | 2.717 | 0 | 0 | 0.543 | 8.696 | 0 | 0 | 2.174 | 0 | |
| 25 | 56 | 55 | 1.582 | 1.727 | 31 | 1.645 | 1.871 | 47 | 90 | 0.522 | 0.143 | 0 | 4 | 3 | 31.35 | 46 | 24 | 0 | 0 | 0 | 3.191 | 0 | 0 | 0 | 3.191 | 0 | 0 | 4.255 | 0 | |
| 26 | 45 | 44 | 1.727 | 1.886 | 20 | 1.95 | 2.15 | 32 | 88 | 0.364 | 0.156 | 0 | 2 | 2.5 | 10.49 | 27 | 25 | 0 | 0 | 0 | 1.075 | 0 | 5.376 | 1.075 | 9.677 | 0 | 0 | 1.075 | 0 | |
| 27 | 95 | 83 | 1.904 | 2.06 | 59 | 2.085 | 2.254 | 70 | 183 | 0.383 | 0.421 | 0 | 22 | 4.05 | 34.78 | 53 | 42 | 0 | 0.532 | 0.532 | 3.191 | 1.596 | 0.532 | 0 | 3.191 | 0 | 0.532 | 0.532 | 0 | |
| 28 | 112 | 97 | 1.536 | 1.732 | 73 | 1.397 | 1.548 | 60 | 177 | 0.339 | 0.25 | 0 | 11 | 4.55 | 17.87 | 56 | 30 | 0.535 | 0 | 1.07 | 2.674 | 0.535 | 0 | 0 | 6.417 | 0 | 0.535 | 5.882 | 0.535 | |
| 29 | 46 | 44 | 1.568 | 1.591 | 20 | 1.6 | 1.6 | 43 | 75 | 0.573 | 0.152 | 0 | 5 | 2.4 | 32.2 | 37 | 22 | 0 | 1.333 | 0 | 1.333 | 1.333 | 0 | 0 | 10.667 | 0 | 0 | 0 | 0 | |
| 30 | 107 | 101 | 1.614 | 1.842 | 77 | 1.636 | 1.87 | 71 | 173 | 0.41 | 0.234 | 0 | 13 | 5 | 34.83 | 72 | 33 | 0 | 0.546 | 0 | 2.732 | 0 | 1.093 | 1.093 | 2.186 | 0 | 0 | 3.825 | 0 | |
| 31 | 77 | 77 | 1.13 | 1.247 | 53 | 1.113 | 1.17 | 33 | 87 | 0.379 | 0.013 | 0 | 1 | 3 | 13.49 | 41 | 16 | 0 | 0 | 0 | 4.598 | 0 | 0 | 0 | 1.149 | 0 | 0 | 0 | 0 | |
| 32 | 65 | 61 | 1.426 | 1.836 | 37 | 1.432 | 1.811 | 31 | 92 | 0.337 | 0.108 | 0 | 3 | 6.67 | 10.68 | 30 | 19 | 4.082 | 0 | 0 | 2.041 | 0 | 0 | 0 | 9.184 | 0 | 0 | 3.061 | 1.02 | |
| 33 | 55 | 54 | 1.463 | 1.593 | 30 | 1.533 | 1.633 | 53 | 80 | 0.663 | 0.145 | 0 | 7 | 4 | 60.45 | 43 | 21 | 1.235 | 0 | 0 | 1.235 | 1.235 | 0 | 0 | 3.704 | 0 | 0 | 0 | 0 | |
| 34 | 68 | 64 | 2.672 | 3.203 | 40 | 2.725 | 3.275 | 82 | 182 | 0.451 | 0.706 | 0 | 35 | 4.63 | 44.87 | 56 | 51 | 1.932 | 0.966 | 1.449 | 1.932 | 0.966 | 0.483 | 2.415 | 7.729 | 0 | 0.966 | 14.01 | 1.932 | |
| 35 | 125 | 125 | 1.232 | 1.392 | 100 | 1.25 | 1.44 | 58 | 156 | 0.372 | 0.072 | 0 | 9 | 3.67 | 14.72 | 54 | 28 | 0 | 0 | 0 | 1.235 | 0 | 0 | 0 | 3.704 | 0 | 0 | 3.086 | 0 | |
| 36 | 54 | 50 | 2.18 | 2.56 | 26 | 2.077 | 2.269 | 58 | 127 | 0.457 | 0.315 | 0 | 16 | 3.88 | 26.01 | 42 | 36 | 0.725 | 0 | 0 | 8.696 | 0 | 0 | 2.174 | 21.014 | 0 | 0 | 8.696 | 0 | |
| 37 | 120 | 106 | 2.208 | 2.566 | 82 | 2.183 | 2.512 | 104 | 262 | 0.397 | 0.408 | 0 | 34 | 5.74 | 60.43 | 76 | 62 | 4.24 | 1.413 | 0.353 | 2.12 | 0.707 | 0 | 0.353 | 4.594 | 0 | 0.707 | 2.473 | 0 | |
| 38 | 3 | 1 | 1 | 1 | 0 | N/A | N/A | 3 | 3 | 1 | 0.333 | 0 | 0 | N/A | N/A | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 39 | 63 | 62 | 1.081 | 1.145 | 38 | 1.105 | 1.158 | 32 | 71 | 0.451 | 0.032 | 0 | 1 | 2 | 12.02 | 35 | 9 | 0 | 0 | 0 | 2.817 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 40 | 86 | 85 | 1.529 | 1.635 | 61 | 1.607 | 1.721 | 37 | 132 | 0.28 | 0.209 | 0 | 8 | 2.88 | 9.82 | 48 | 26 | 0 | 0 | 1.515 | 2.273 | 0 | 0 | 9.091 | 5.303 | 0 | 3.788 | 9.091 | 0 | |
| 41 | 107 | 101 | 2.228 | 2.663 | 77 | 1.974 | 2.377 | 98 | 268 | 0.366 | 0.458 | 0 | 36 | 4.97 | 42.01 | 74 | 61 | 1.031 | 2.062 | 0.344 | 2.062 | 0 | 0 | 1.031 | 8.247 | 0 | 3.436 | 7.56 | 0.344 | |
| 42 | 113 | 105 | 1.286 | 1.505 | 81 | 1.37 | 1.605 | 61 | 148 | 0.412 | 0.053 | 0 | 1 | 3 | 18.74 | 73 | 22 | 0 | 0 | 0 | 8.609 | 0 | 1.325 | 1.987 | 3.311 | 0 | 0 | 1.987 | 0 | |
| 43 | 111 | 110 | 1.991 | 2.409 | 86 | 1.988 | 2.395 | 115 | 221 | 0.52 | 0.315 | 0 | 26 | 5.42 | 83.52 | 90 | 48 | 0.83 | 0 | 0.415 | 6.639 | 0.415 | 0 | 0.83 | 1.66 | 0 | 1.245 | 4.149 | 0.415 | |

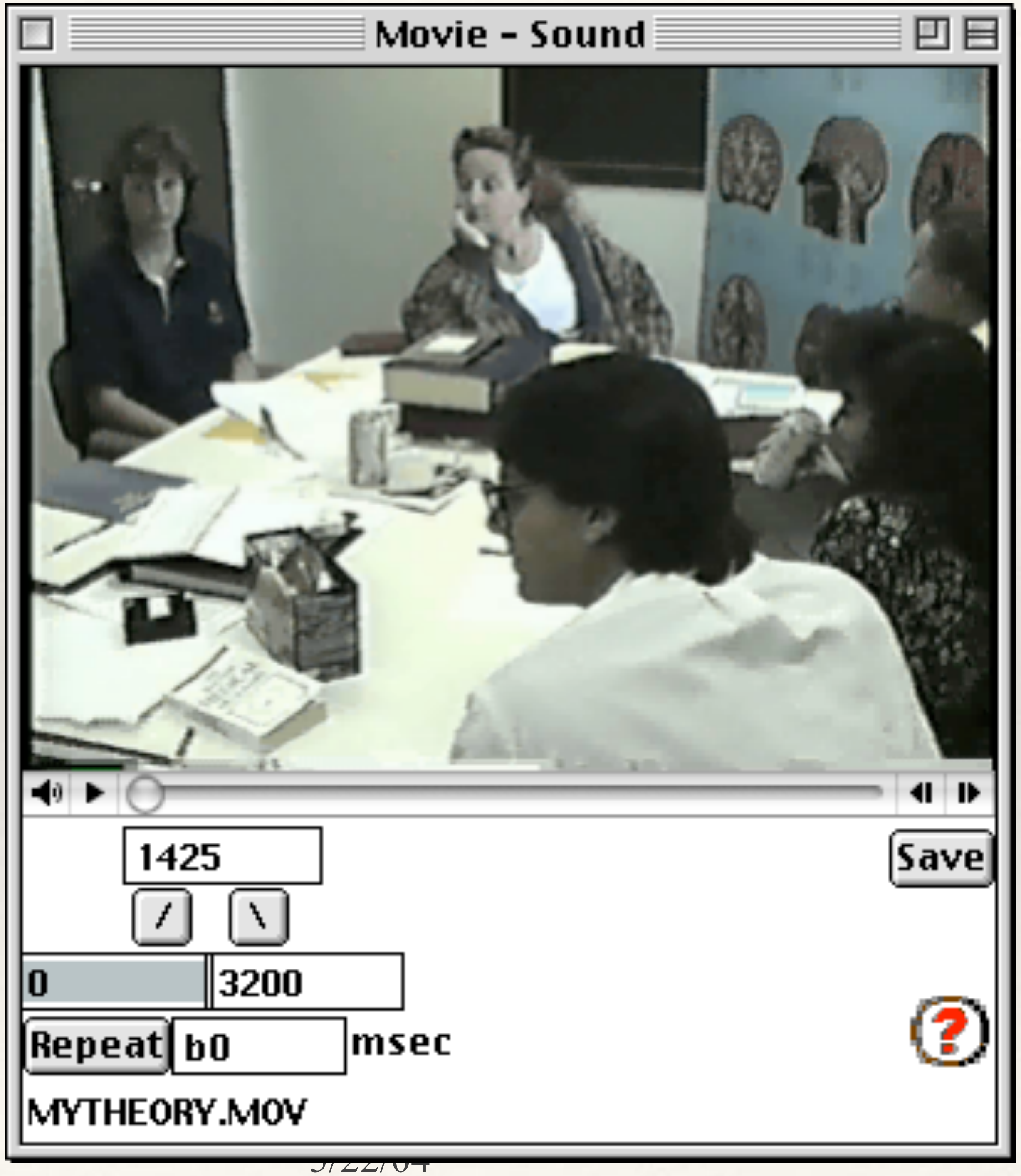
# Analysis Pathways

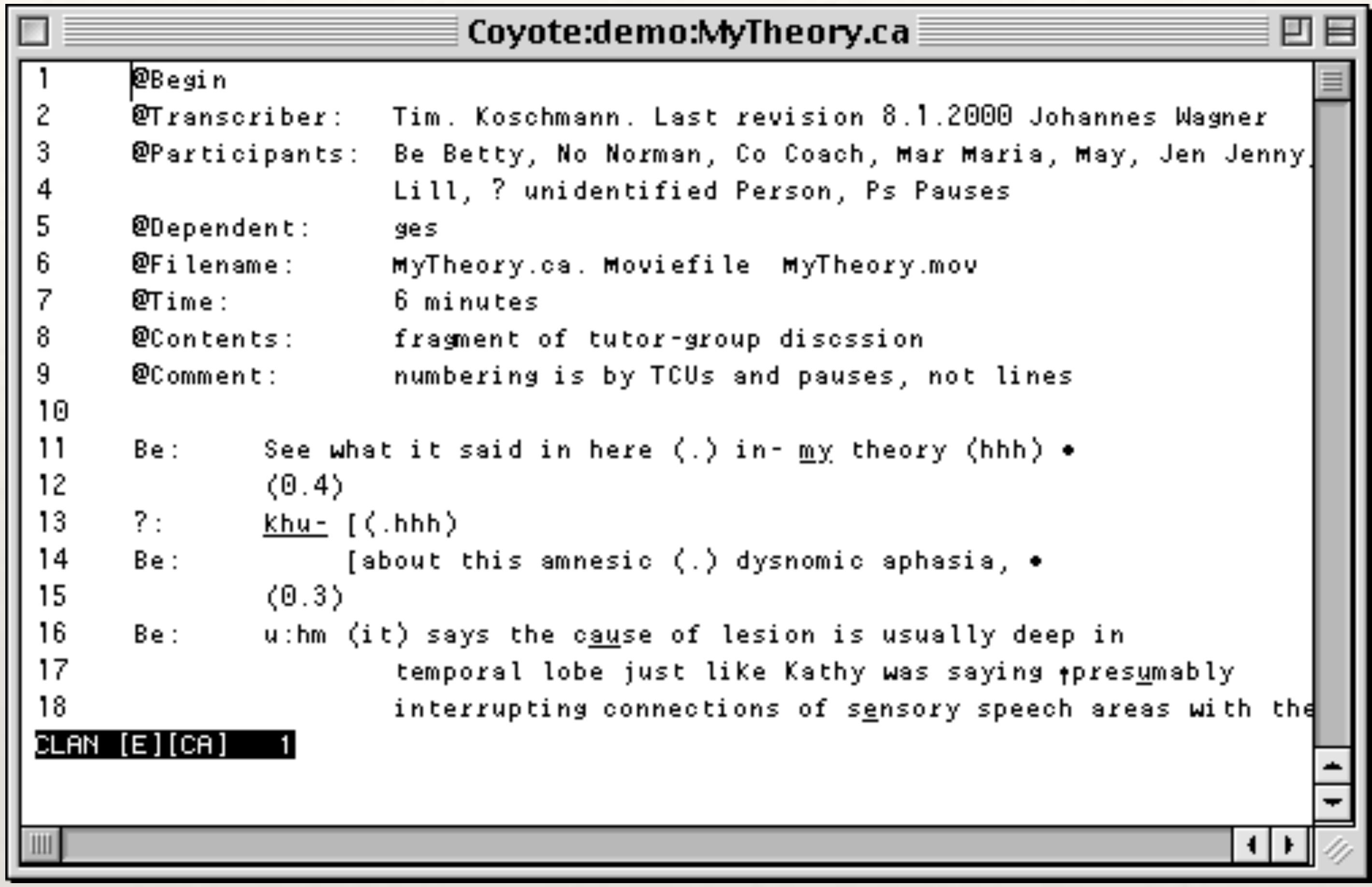


# Error Analysis

- [*p] phonological p:w, p:n, p:m
- [* s] semantic  s:r, s:ur. s:uk, s:per
- [* n] neologism n:k, n:uk, n:k:s, n:uk:s
- [* d] dysfluency
- [* m] morphology m:a:0es  etc.
- [* f] formal lexical
- [+ gram] [+ jar] [+ es] [+ per] [+ cir]

# 3. Microanalysis (CA and Gesture)

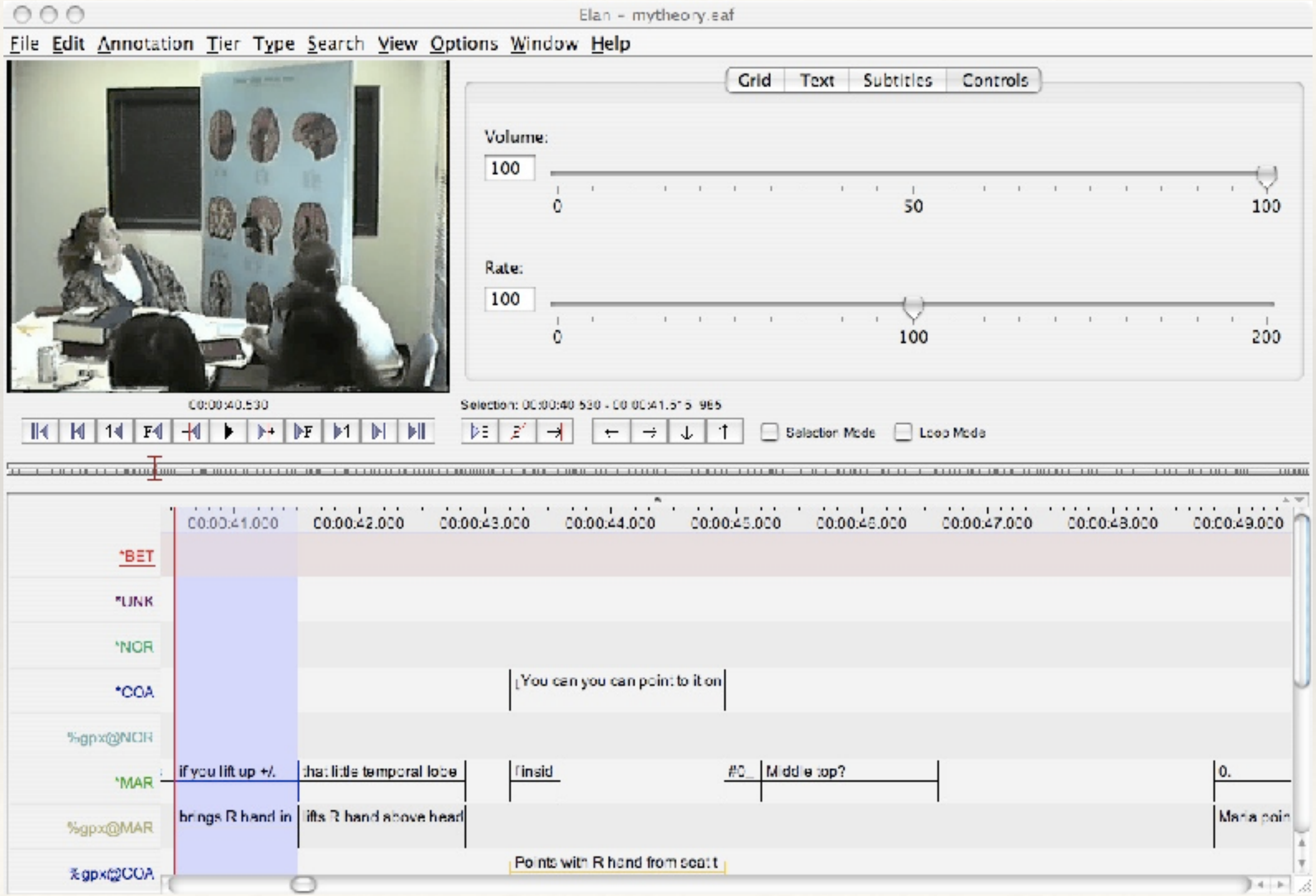



Coyote:demo:MyTheory.ca

```
1       @Begin
2       @Transcriber:   Tim. Koschmann. Last revision 8.1.2000 Johannes Wagner
3       @Participants:  Be Betty, No Norman, Co Coach, Mar Maria, May, Jen Jenny,
4                       Lill, ? unidentified Person, Ps Pauses
5       @Dependent:     ges
6       @Filename:      MyTheory.ca. Moviefile  MyTheory.mov
7       @Time:          6 minutes
8       @Contents:      fragment of tutor-group discssion
9       @Comment:       numbering is by TCUs and pauses, not lines
10
11      Be:     See what it said in here (.) in- my theory (hhh) •
12              (0.4)
13      ?:      khu- [(.hhh)
14      Be:          [about this amnesic (.) dysnomic aphasia, •
15              (0.3)
16      Be:     u:hm (it) says the cause of lesion is usually deep in
17                      temporal lobe just like Kathy was saying ↑presumably
18                      interrupting connections of sensory speech areas with the
CLAN [E][CA]    1
```

3/22/04

# CHAT2ELAN



Volume: 100

Rate: 100

00:00:40.530

Selection: 00:00:40.530 - 00:00:41.575 965

*COA: [You can you can point to it on

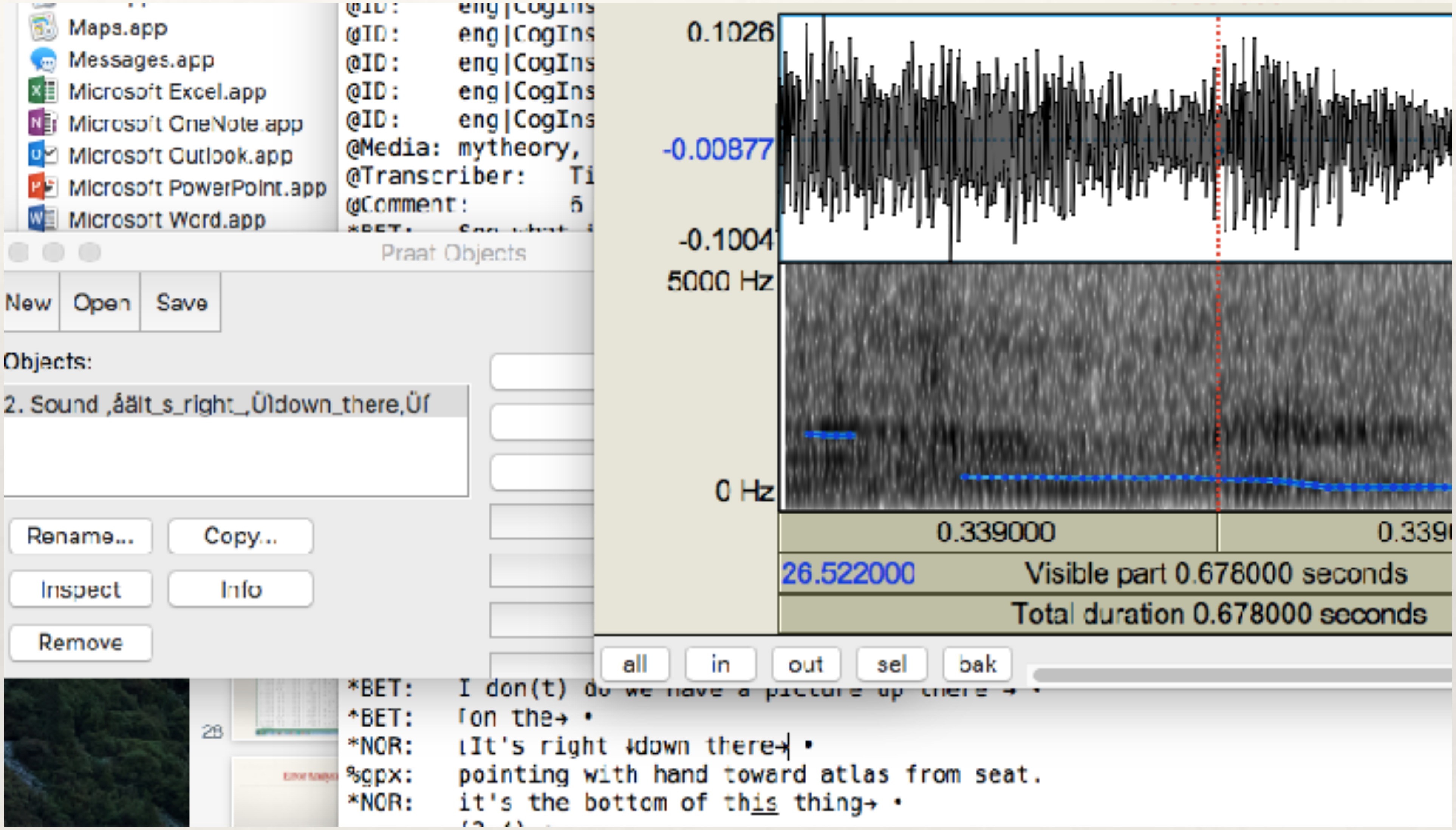
*MAR: if you lift up +/. | that little temporal lobe | [insid | #0_ | Middle top? | 0.

%gpx@MAR: brings R hand in | lifts R hand above head | Maria poin

%gpx@COA: Points with R hand from seat t

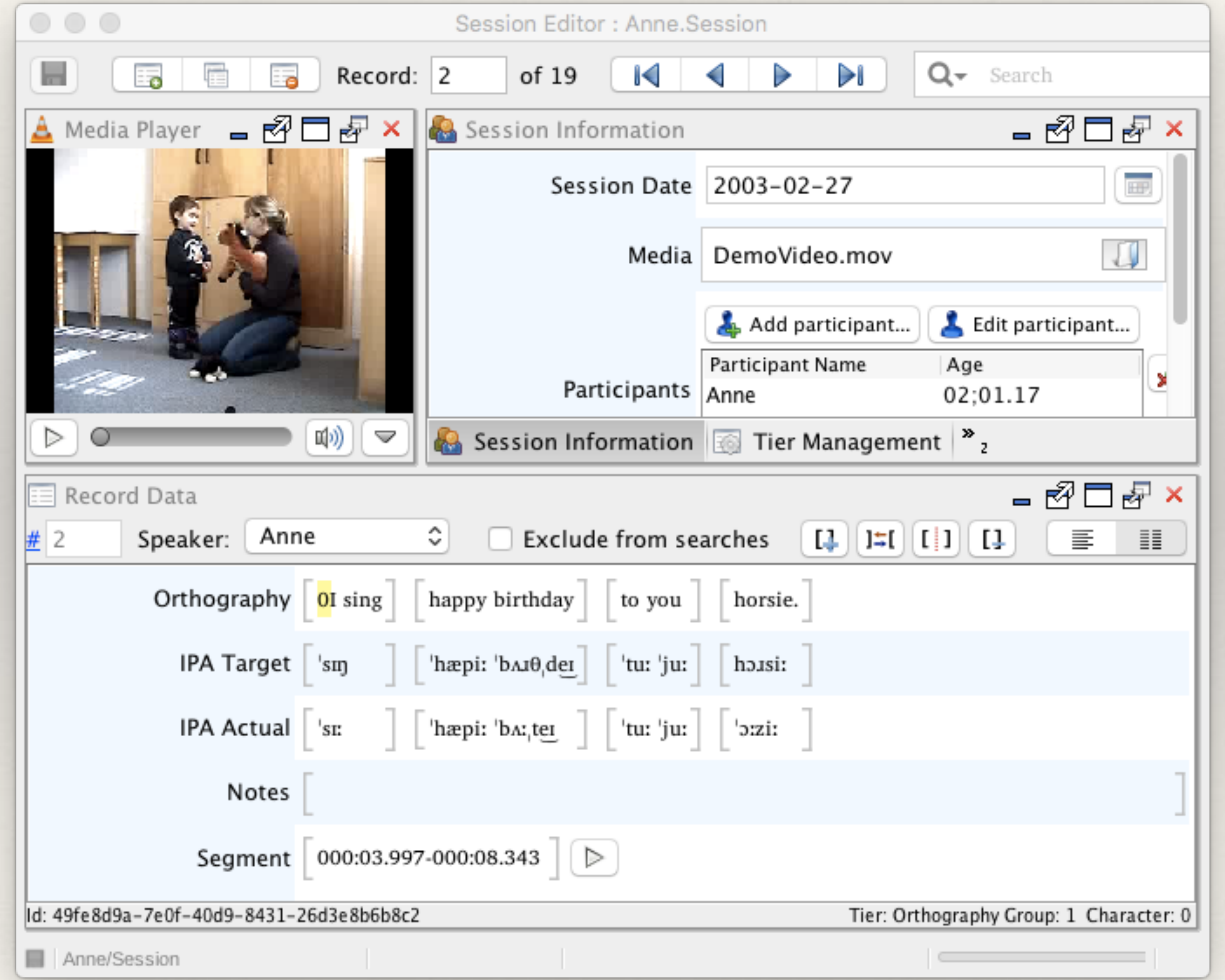
# CHAT2PRAAT - sociophonetics

- Highlight utterance bullet
- Send to sound analyzer
- Extracts audio from video
- In Praat, draw a picture



```
@ID:	eng|CogIns
@ID:	eng|CogIns
@ID:	eng|CogIns
@ID:	eng|CogIns
@ID:	eng|CogIns
@Media:	mytheory,
@Transcriber:	Ti
@Comment:	6
*BET:	I don(t) do we have a picture up there →  •
*BET:	⌈on the→ •
*NOR:	⌊It's right ↓down there→ •
%gpx:	pointing with hand toward atlas from seat.
*NOR:	it's the bottom of this thing→ •
```

# CHAT2PHON



Session Editor : Anne.Session

Record: 2 of 19

Session Date: 2003-02-27

Media: DemoVideo.mov

Participants:

| Participant Name | Age |
|---|---|
| Anne | 02;01.17 |

Record Data

# 2 Speaker: Anne

□ Exclude from searches

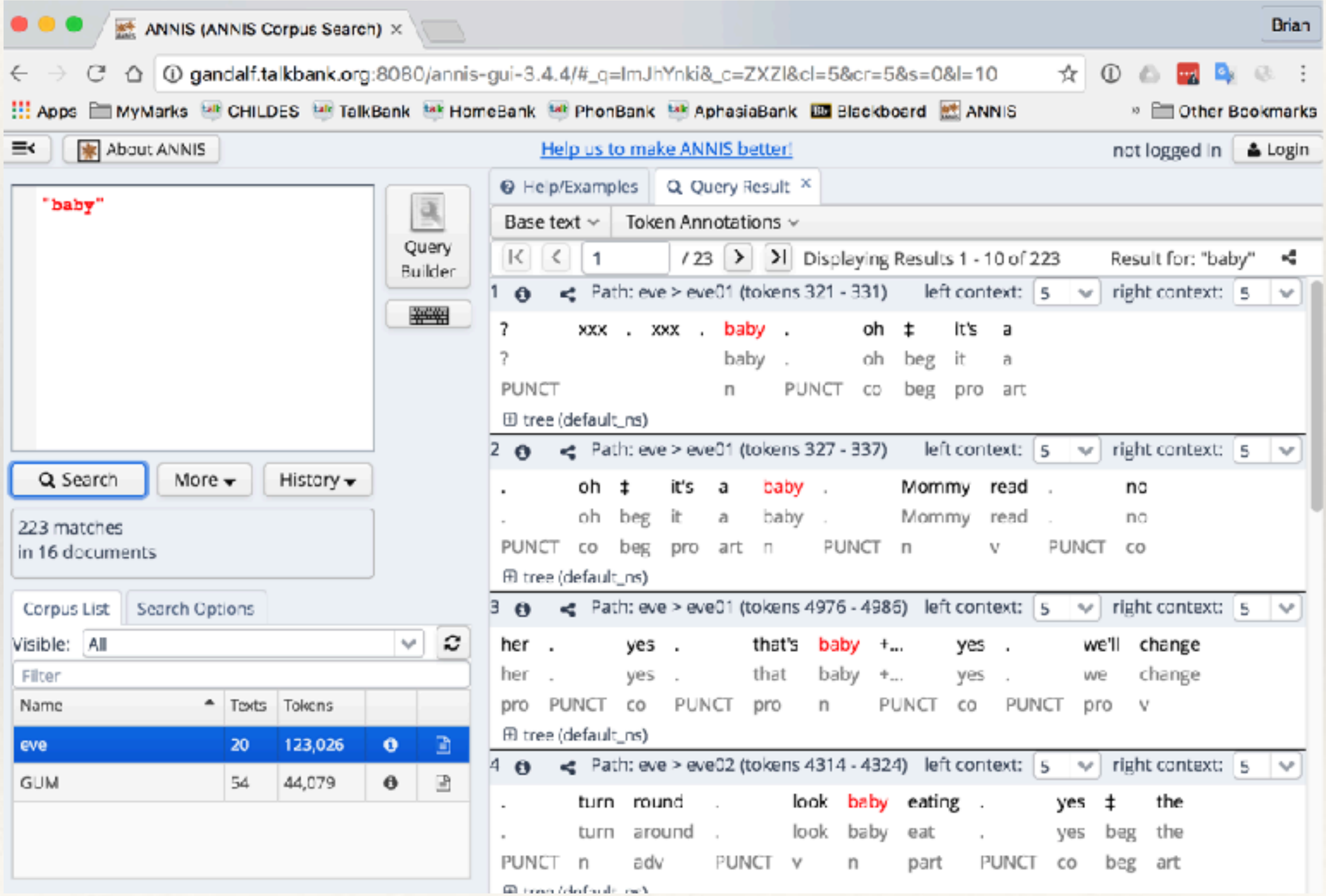
| | | | | |
|---|---|---|---|---|
| Orthography | 0I sing | happy birthday | to you | horsie. |
| IPA Target | ˈsɪŋ | ˈhæpiː ˈbʌɹθˌde͜ɪ | ˈtuː ˈjuː | hɔɹsiː |
| IPA Actual | ˈsɪː | ˈhæpiː ˈbʌːˌte͜ɪ | ˈtuː ˈjuː | ˈɔːziː |
| Notes | | | | |
| Segment | 000:03.997-000:08.343 | | | |

Id: 49fe8d9a-7e0f-40d9-8431-26d3e8b6b8c2

Tier: Orthography Group: 1 Character: 0

Anne/Session

# CHAT in ANNIS



ANNIS (ANNIS Corpus Search)

gandalf.talkbank.org:8080/annis-gui-3.4.4/#_q=ImJhYnki&_c=ZXZI&cl=5&cr=5&s=0&l=10

"baby"

223 matches
in 16 documents

Corpus List | Search Options

Visible: All

Filter

| Name | Texts | Tokens |
|---|---|---|
| eve | 20 | 123,026 |
| GUM | 54 | 44,079 |

1 / 23 Displaying Results 1 - 10 of 223 Result for: "baby"

1 Path: eve > eve01 (tokens 321 - 331) left context: 5 right context: 5

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ? | xxx | . | xxx | . | baby | . | oh | ‡ | It's | a |
| ? | | | | | baby | . | oh | beg | it | a |
| PUNCT | | | | | n | PUNCT | co | beg | pro | art |

tree (default_ns)

2 Path: eve > eve01 (tokens 327 - 337) left context: 5 right context: 5

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| . | oh | ‡ | it's | a | baby | . | Mommy | read | . | no |
| . | oh | beg | it | a | baby | . | Mommy | read | . | no |
| PUNCT | co | beg | pro | art | n | PUNCT | n | v | PUNCT | co |

tree (default_ns)

3 Path: eve > eve01 (tokens 4976 - 4986) left context: 5 right context: 5

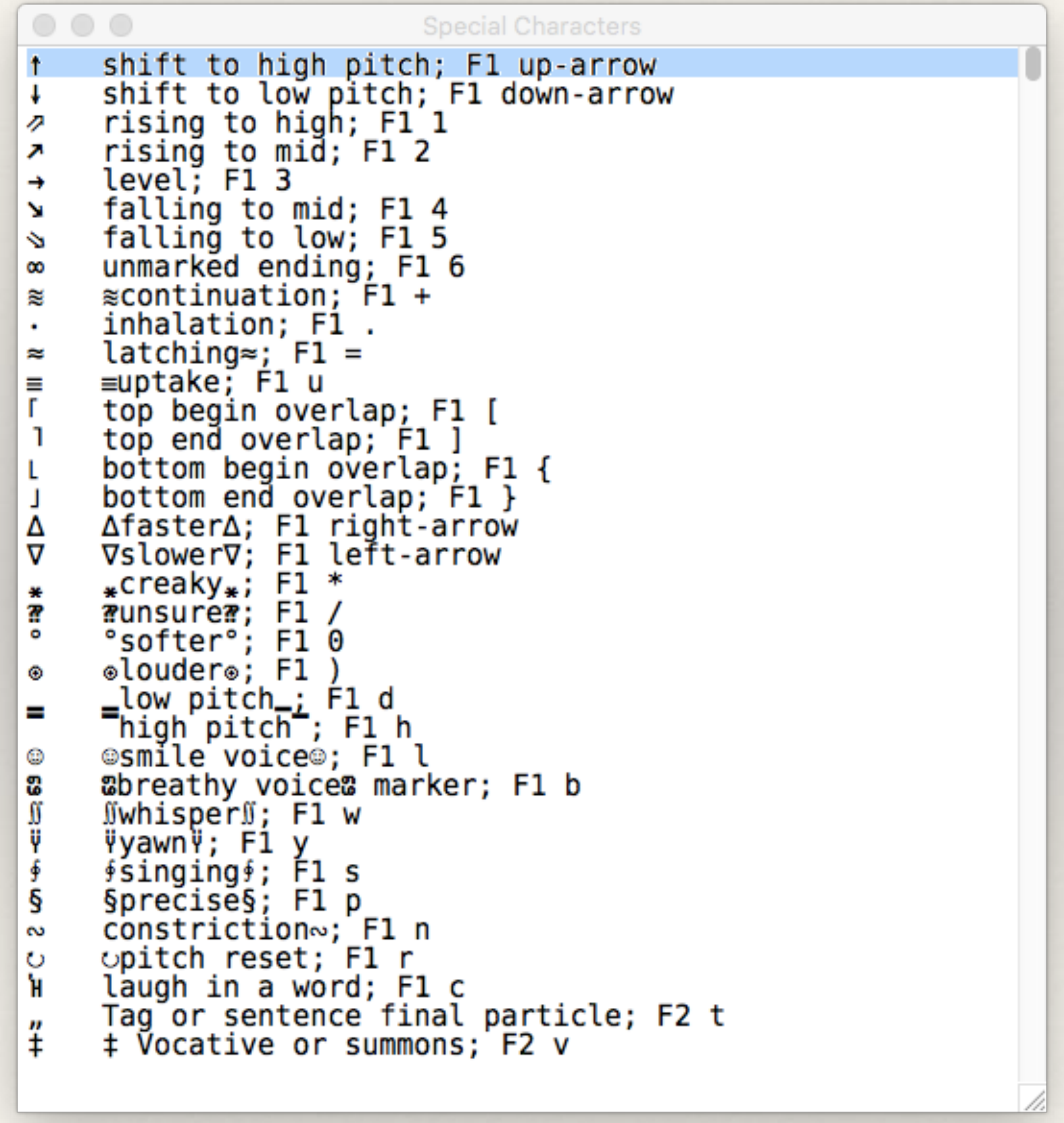
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| her | . | yes | . | that's | baby | +... | yes | . | we'll | change |
| her | . | yes | . | that | baby | +... | yes | . | we | change |
| pro | PUNCT | co | PUNCT | pro | n | PUNCT | co | PUNCT | pro | v |

tree (default_ns)

4 Path: eve > eve02 (tokens 4314 - 4324) left context: 5 right context: 5

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| . | turn | round | . | look | baby | eating | . | yes | ‡ | the |
| . | turn | around | . | look | baby | eat | . | yes | beg | the |
| PUNCT | n | adv | PUNCT | v | n | part | PUNCT | co | beg | art |

# CA Coding



```
↑    shift to high pitch; F1 up-arrow
↓    shift to low pitch; F1 down-arrow
⇗    rising to high; F1 1
↗    rising to mid; F1 2
→    level; F1 3
↘    falling to mid; F1 4
⇘    falling to low; F1 5
∞    unmarked ending; F1 6
≋    ≋continuation; F1 +
∙    inhalation; F1 .
≈    latching≈; F1 =
≡    ≡uptake; F1 u
⌈    top begin overlap; F1 [
⌉    top end overlap; F1 ]
⌊    bottom begin overlap; F1 {
⌋    bottom end overlap; F1 }
∆    ∆faster∆; F1 right-arrow
∇    ∇slower∇; F1 left-arrow
⁎    ⁎creaky⁎; F1 *
⁇    ⁇unsure⁇; F1 /
°    °softer°; F1 0
◉    ◉louder◉; F1 )
▁    ▁low pitch▁; F1 d
▔    ▔high pitch▔; F1 h
☺    ☺smile voice☺; F1 l
♋    ♋breathy voice♋ marker; F1 b
∬    ∬whisper∬; F1 w
Ϋ    ΫyawnΫ; F1 y
∮    ∮singing∮; F1 s
§    §precise§; F1 p
∾    constriction∾; F1 n
↻    ↻pitch reset; F1 r
Ἡ    laugh in a word; F1 c
„    Tag or sentence final particle; F2 t
‡    ‡ Vocative or summons; F2 v
```

# Gestural Detail

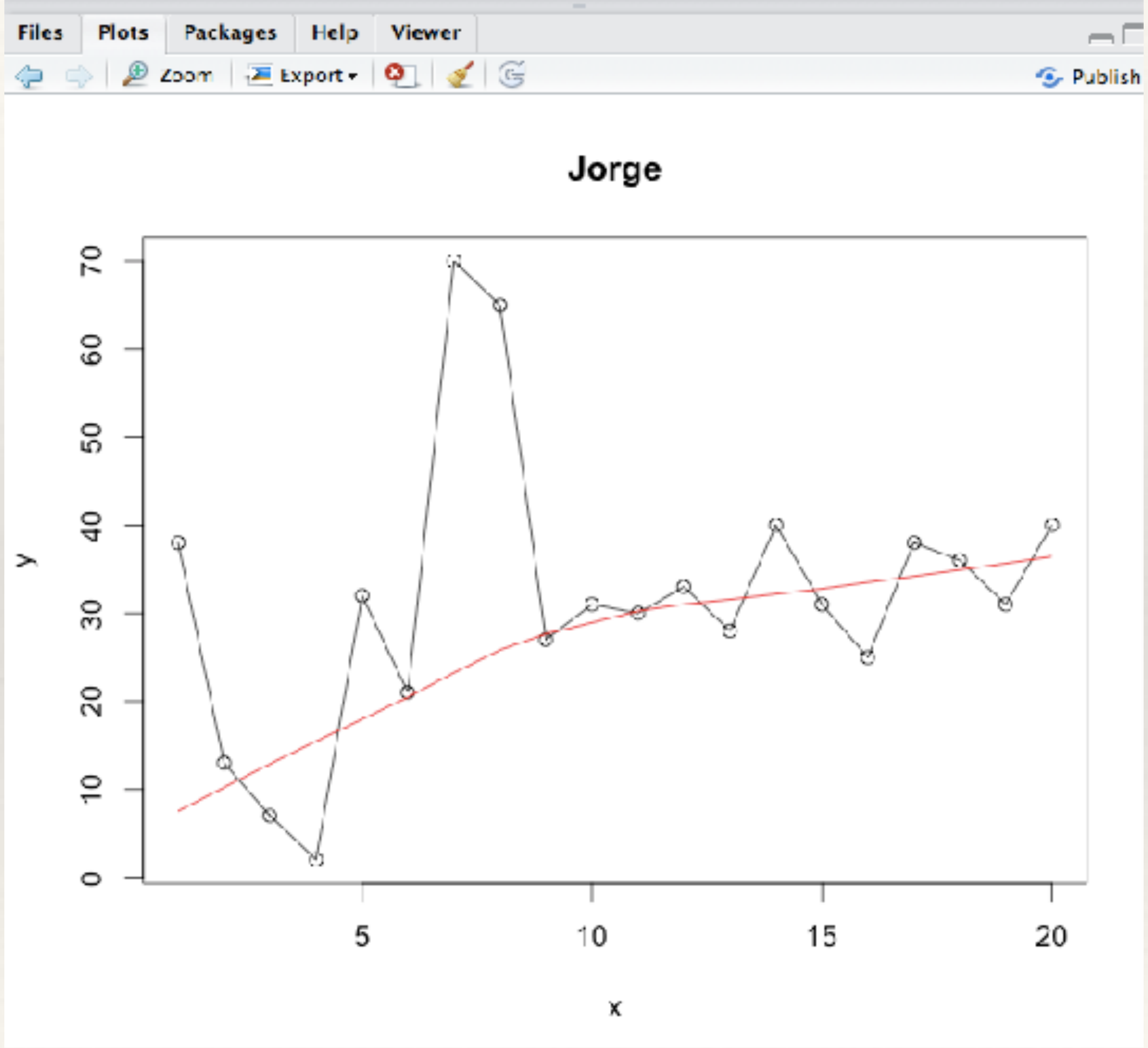
- Interaction / Sequence / Segment
- Each participant coded through sequence
  - Deedee 1a-1b-1c
  - Nina 1a-1b-1c
- Bullet links each segment back to transcript
- Coding: gaze direction, action, classification, meaning, language
- Rapport coding through gaze, smile, language

# Discourse Analysis

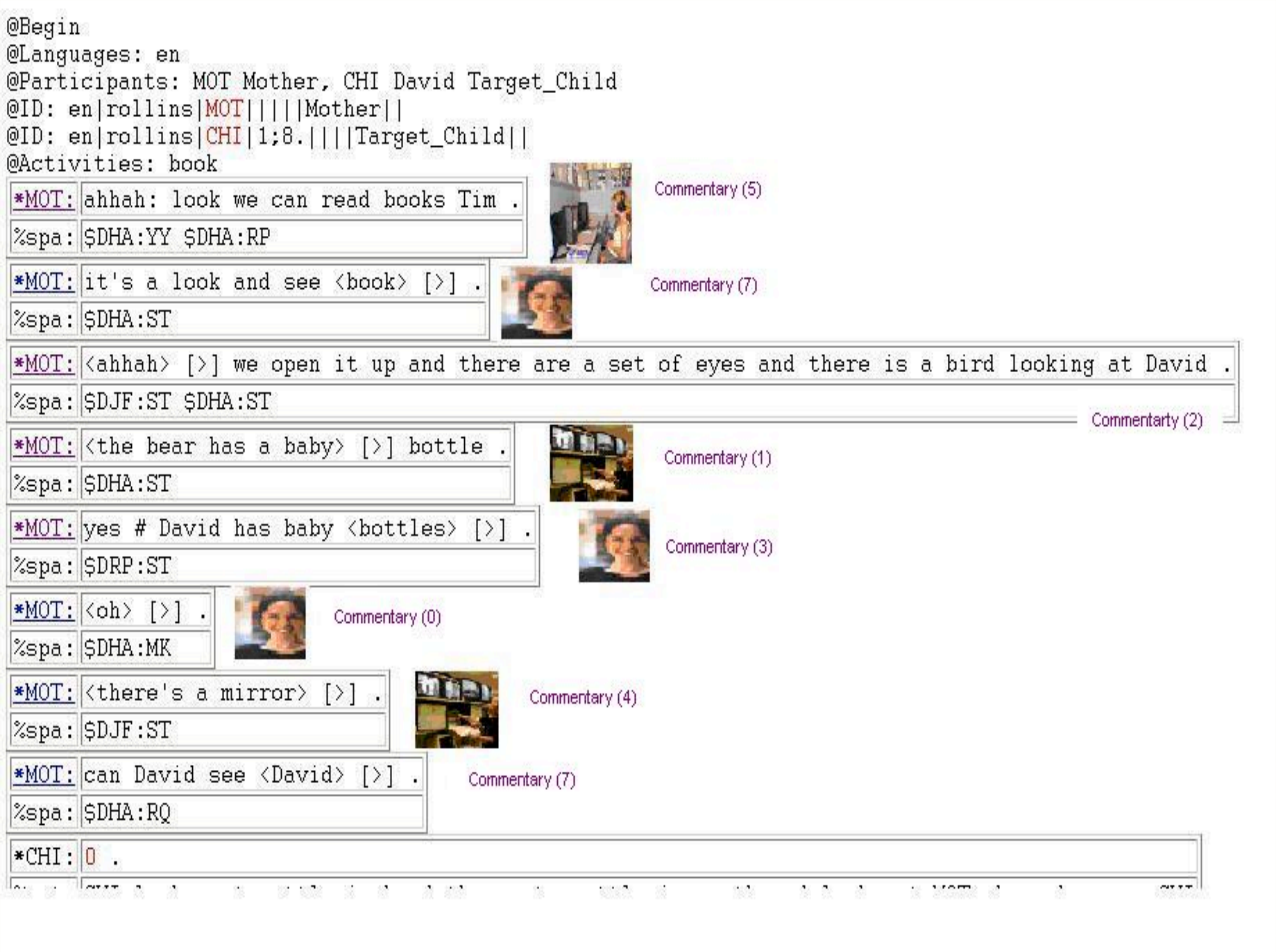
- CHAINS, KEYMAP, DIST
- CHIP
- PD (Propositional density)
- CI (Complexity index)
- SCRIPT + Speech Kitchen ASR

# Time Series - corpora to R

Alberto and Jorge — I no go.



# Collaborative Commentary



```
@Begin
@Languages: en
@Participants: MOT Mother, CHI David Target_Child
@ID: en|rollins|MOT||||Mother||
@ID: en|rollins|CHI|1;8.||||Target_Child||
@Activities: book
*MOT: ahhah: look we can read books Tim .
%spa: $DHA:YY $DHA:RP
*MOT: it's a look and see <book> [>] .
%spa: $DHA:ST
*MOT: <ahhah> [>] we open it up and there are a set of eyes and there is a bird looking at David .
%spa: $DJF:ST $DHA:ST
*MOT: <the bear has a baby> [>] bottle .
%spa: $DHA:ST
*MOT: yes # David has baby <bottles> [>] .
%spa: $DRP:ST
*MOT: <oh> [>] .
%spa: $DHA:MK
*MOT: <there's a mirror> [>] .
%spa: $DJF:ST
*MOT: can David see <David> [>] .
%spa: $DHA:RQ
*CHI: 0 .
```

# Messages for CLARIN

- Importance of open access
- Importance of uniform transcription format linked to analysis programs
- Importance of focus on specific research communities for:
  - corpus development
  - tool development
  - FUNDING

# Conclusions

- We need to expand TalkBank
- CLARIN can make wider use of TalkBank methods
- We can promotate TalkBank-CLARIN integration