UNIVERSITY OF Southampton

Formatting and Analysing a Learner Corpus with CHAT and CLAN

Kevin McManus & Nicole Tracy-Ventura

University of Southampton

Introduction to CHILDES

- **Child Language Data Exchange System** / CHILDES

  (for a useful introduction to CHILDES see MacWhinney 2000 or http://childes.psy.cmu.edu/)

- **Talkbank**– The Database – primarily child language but also some language disorder data and bilingual data
- **CHAT** (Codes for the Human Analysis of Transcripts) – transcription procedures, a system for notation and coding
- **CLAN** (Computerised Language Analysis) – computer programs for searching and manipulating the data.

Why do researchers use CHILDES?

- Well-supported; list-serves available
- International standing (over 1300 published studies)
- Flexible (not language-specific)
- Powerful (over 40 commands)
- Morphosyntactic tagger & direct search for morphosyntactic variables
- It’s free!

Example SLA Projects using CHAT/CLAN

- FLLOC (French Learner Language Oral Corpora) Project
  - 3.5 million words
- SPLLOC (Spanish Learner Language Oral Corpora) Project
  - 330,000 words
- LANG-SNAP (Language and Social Networks Abroad Project)
  - 680,000 words

The FLLOC and SPLLOC projects

flloc.soton.ac.uk & splloc.soton.ac.uk

- Focus in on instructed learners of French and Spanish
- Dual aim:
  - constructing databases of oral learner data freely available to the research community
  - Substantive research agenda on French and Spanish SLA
- Collaboration between the Universities of Essex, Newcastle, Southampton and York
- Team Members: Florence Myles, Rosamond Mitchell, Laura Domínguez, Emma Marsden, Sarah Rule, Annabelle David, Maria Arche, Nicole Tracy-Ventura, Kevin McManus, Christophe dos Santos, Administrative and IT support

LANG-SNAP project:

langsnap.soton.ac.uk

- French and Spanish L2
- Focus on speech and writing
- Longitudinal: 6 data collection cycles over 20 months
- Use of CHILDES Procedures: soundfiles, bulleted transcripts, tagged transcripts
- Advanced learners, native speaker controls
- Eventually accessible on the web via Talkbank and langsnap.soton.ac.uk/
- 680,000 words

CHAT/CLAN Getting Started

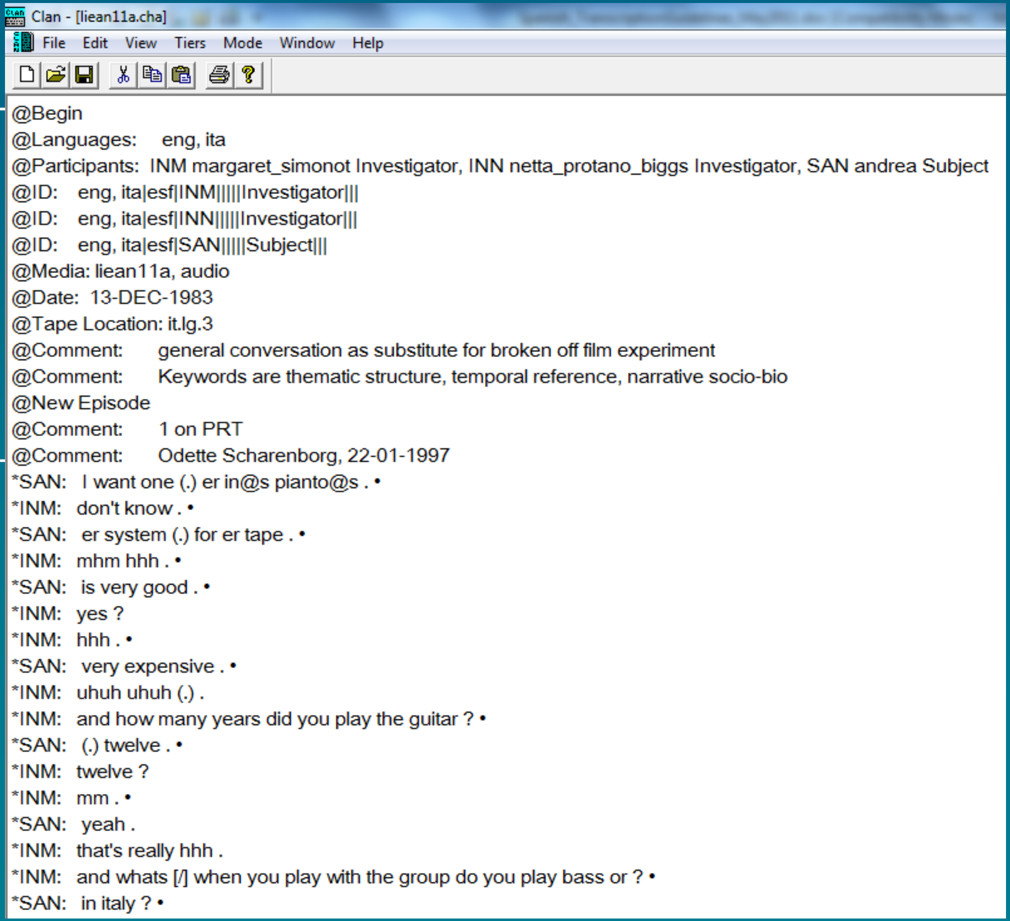
- Download the CLAN program from the Childes website
  - childes.psy.cmu.edu/clan/
  - It is already on the computers in this lab.
- Download both the CHAT & CLAN manuals
  - CHAT: childes.psy.cmu.edu/manuals/chat.pdf
  - CLAN: childes.psy.cmu.edu/manuals/clan.pdf

CHAT: Codes for the Human Analysis of Transcripts

- CHAT is the set of transcription conventions you need to follow.

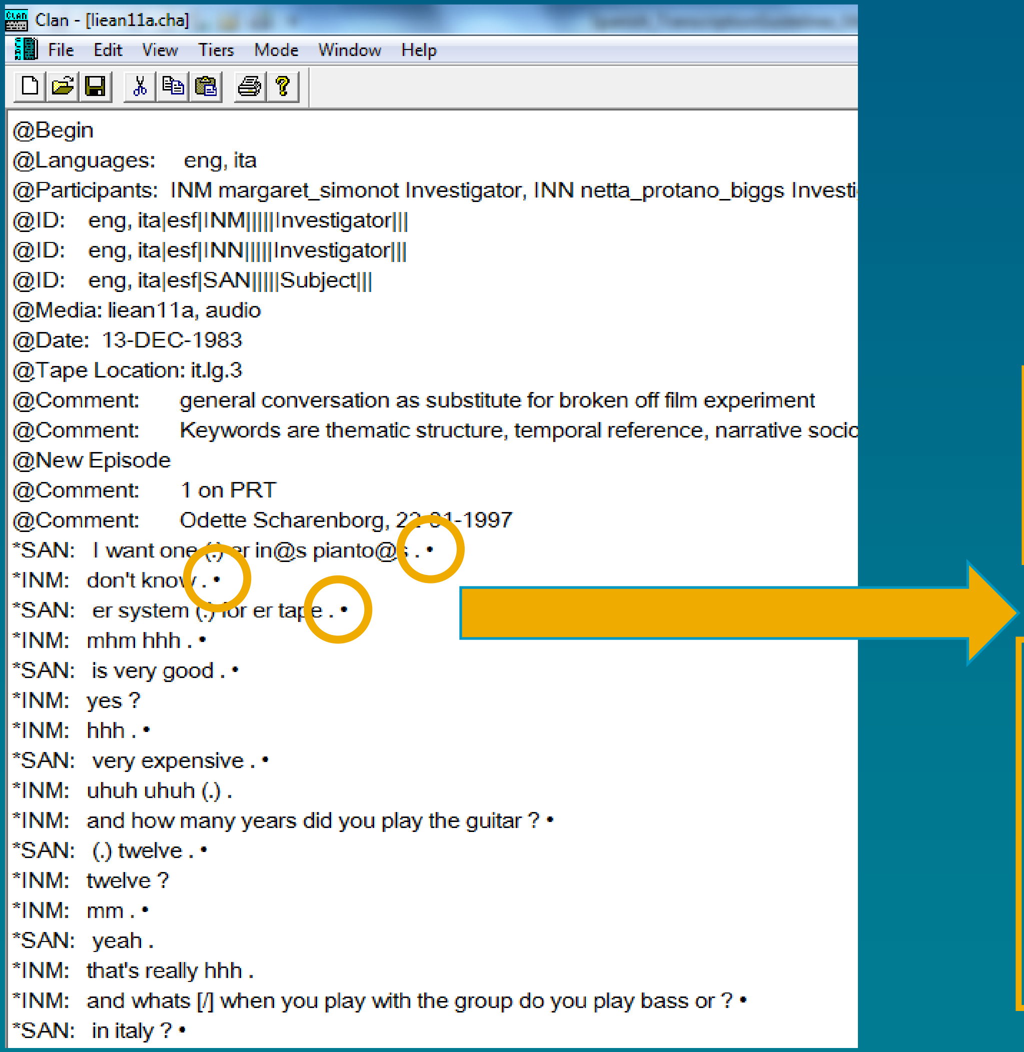
Header

Start of transcript. Each t-unit goes on a separate line



```
@Begin
@Languages:    eng, ita
@Participants:  INM margaret_simonot Investigator, INN netta_protano_biggs Investigator, SAN andrea Subject
@ID:   eng, ita|esf|INM|||||Investigator|||
@ID:   eng, ita|esf|INN|||||Investigator|||
@ID:   eng, ita|esf|SAN|||||Subject|||
@Media: liean11a, audio
@Date:  13-DEC-1983
@Tape Location: it.lg.3
@Comment:      general conversation as substitute for broken off film experiment
@Comment:      Keywords are thematic structure, temporal reference, narrative socio-bio
@New Episode
@Comment:      1 on PRT
@Comment:      Odette Scharenborg, 22-01-1997
*SAN:  I want one (.) er in@s pianto@s . •
*INM:  don't know . •
*SAN:  er system (.) for er tape . •
*INM:  mhm hhh . •
*SAN:  is very good . •
*INM:  yes ?
*INM:  hhh . •
*SAN:  very expensive . •
*INM:  uhuh uhuh (.) .
*INM:  and how many years did you play the guitar ? •
*SAN:  (.) twelve . •
*INM:  twelve ?
*INM:  mm . •
*SAN:  yeah .
*INM:  that's really hhh .
*INM:  and whats [/] when you play with the group do you play bass or ? •
*SAN:  in italy ? •
```

Bulleting: Linking the transcript and audio/video



```
@Begin
@Languages:    eng, ita
@Participants:  INM margaret_simonot Investigator, INN netta_protano_biggs Investi
@ID:    eng, ita|esf|INM|||||Investigator|||
@ID:    eng, ita|esf|INN|||||Investigator|||
@ID:    eng, ita|esf|SAN|||||Subject|||
@Media: liean11a, audio
@Date:  13-DEC-1983
@Tape Location: it.lg.3
@Comment:      general conversation as substitute for broken off film experiment
@Comment:      Keywords are thematic structure, temporal reference, narrative socio
@New Episode
@Comment:      1 on PRT
@Comment:      Odette Scharenborg, 22-01-1997
*SAN:   I want one (.) er in@s pianto@s . •
*INM:   don't know . •
*SAN:   er system (.) for er tape . •
*INM:   mhm hhh . •
*SAN:   is very good . •
*INM:   yes ?
*INM:   hhh . •
*SAN:   very expensive . •
*INM:   uhuh uhuh (.) .
*INM:   and how many years did you play the guitar ? •
*SAN:   (.) twelve . •
*INM:   twelve ?
*INM:   mm . •
*SAN:   yeah .
*INM:   that's really hhh .
*INM:   and whats [/] when you play with the group do you play bass or ? •
*SAN:   in italy ? •
```

Bullets show that the audio and transcript are linked.

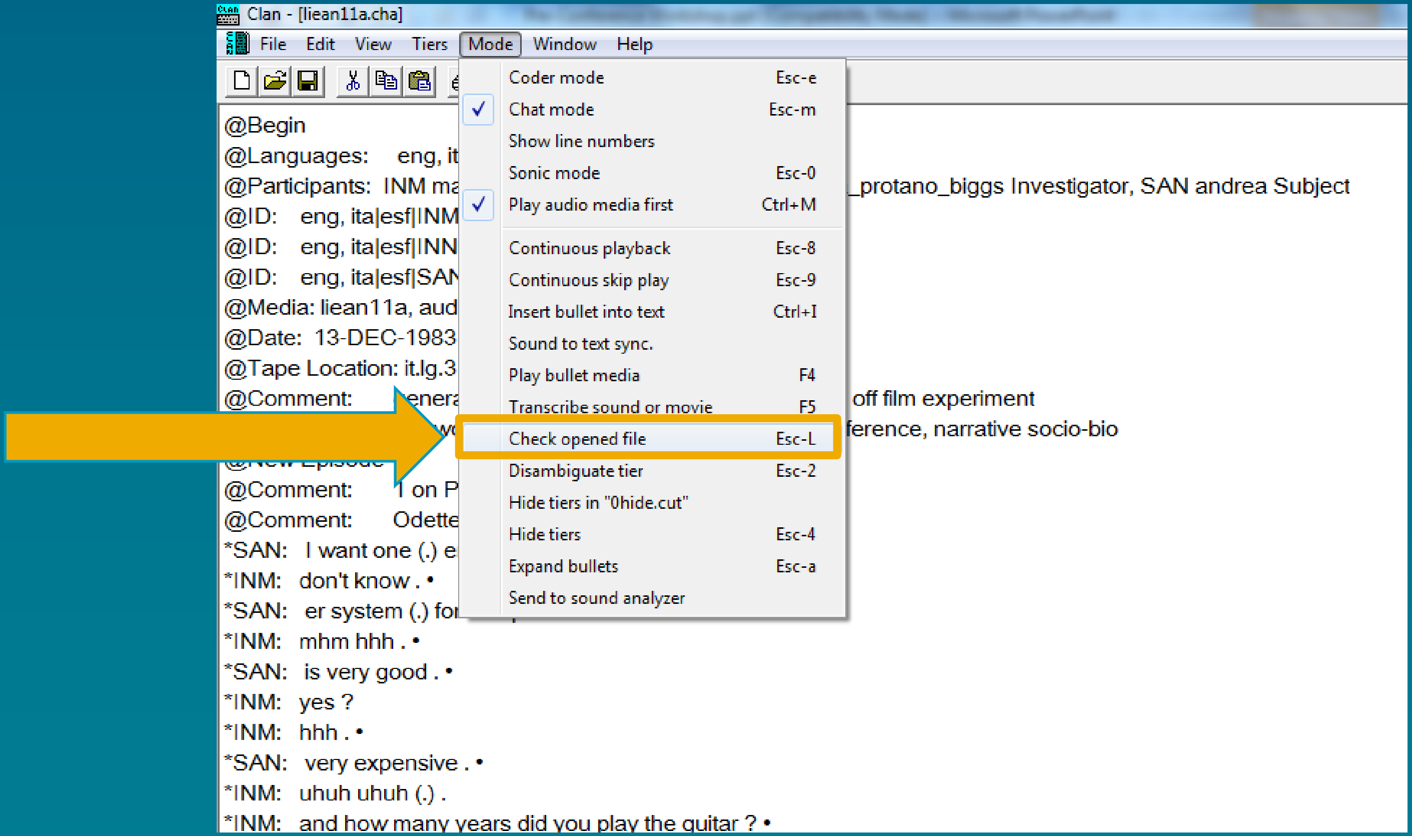
Now you can play any line in the transcript by pressing F4 when the cursor is over that line

Transcribing in the CLAN Program

- ‘Sonic Mode’: transcribe and bullet at the same time.
- ‘Transcriber Mode’: transcribe with video and link already transcribed files with audio.
- ‘Soundwalker’: similar to the old transcriber foot pedal.

Once you've finished transcribing: The 'Check' command

- Used to make sure your transcript doesn't include any technical errors.



```
@Begin
@Languages:    eng, it
@Participants:  INM ma                _protano_biggs Investigator, SAN andrea Subject
@ID:    eng, ita|esf|INM
@ID:    eng, ita|esf|INN
@ID:    eng, ita|esf|SAN
@Media: liean11a, aud
@Date:  13-DEC-1983
@Tape Location: it.lg.3
@Comment:        enera                off film experiment
                  w                   ference, narrative socio-bio
@New Episode
@Comment:        1 on P
@Comment:        Odette
*SAN:   I want one (.) e
*INM:   don't know . •
*SAN:   er system (.) for
*INM:   mhm hhh . •
*SAN:   is very good . •
*INM:   yes ?
*INM:   hhh . •
*SAN:   very expensive . •
*INM:   uhuh uhuh (.) .
*INM:   and how many years did you play the guitar ? •
```

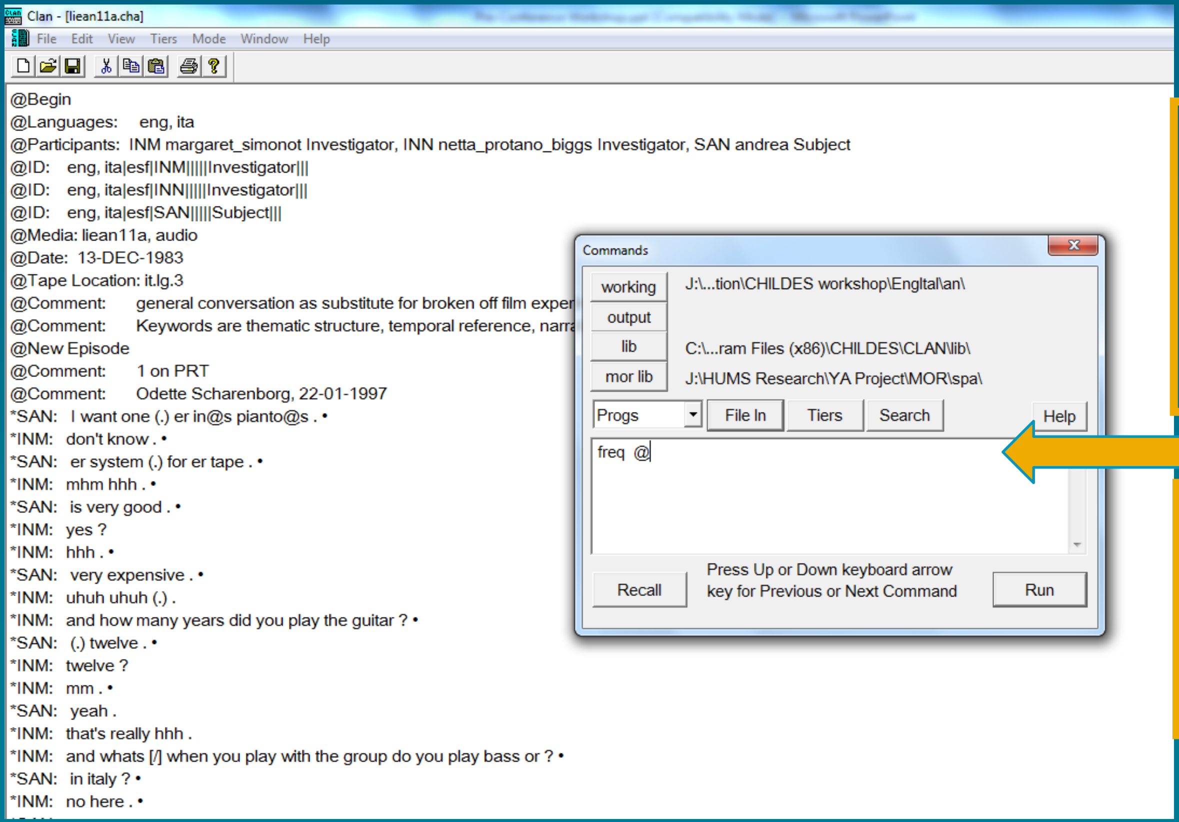
CLAN, the analysis program

- Over 40 built-in programs that work on CHAT files
  - Windows and Mac versions

- Some commands are run on the CHAT file.
- Some commands are run on the ‘tagged’ or ‘MOR’ file

Obtaining a Frequency List

- Command: freq



```
@Begin
@Languages:	eng, ita
@Participants:	INM margaret_simonot Investigator, INN netta_protano_biggs Investigator, SAN andrea Subject
@ID:	eng, ita|esf|INM|||||Investigator|||
@ID:	eng, ita|esf|INN|||||Investigator|||
@ID:	eng, ita|esf|SAN|||||Subject|||
@Media: liean11a, audio
@Date:	13-DEC-1983
@Tape Location: it.lg.3
@Comment:	general conversation as substitute for broken off film exper
@Comment:	Keywords are thematic structure, temporal reference, narra
@New Episode
@Comment:	1 on PRT
@Comment:	Odette Scharenborg, 22-01-1997
*SAN:	I want one (.) er in@s pianto@s . •
*INM:	don't know . •
*SAN:	er system (.) for er tape . •
*INM:	mhm hhh . •
*SAN:	is very good . •
*INM:	yes ?
*INM:	hhh . •
*SAN:	very expensive . •
*INM:	uhuh uhuh (.) .
*INM:	and how many years did you play the guitar ? •
*SAN:	(.) twelve . •
*INM:	twelve ?
*INM:	mm . •
*SAN:	yeah .
*INM:	that's really hhh .
*INM:	and whats [/] when you play with the group do you play bass or ? •
*SAN:	in italy ? •
*INM:	no here . •
```

working: J:\...tion\CHILDES workshop\Engltal\an\

lib: C:\...ram Files (x86)\CHILDES\CLAN\lib\

mor lib: J:\HUMS Research\YA Project\MOR\spa\

```
freq  @
```

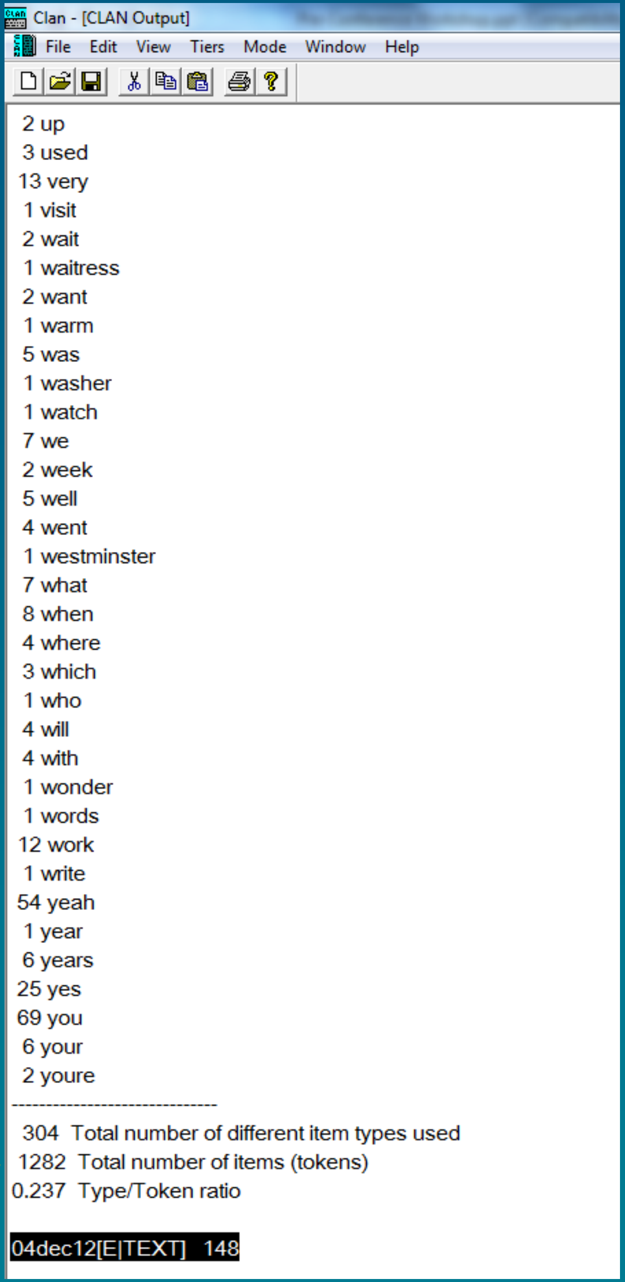
Press Up or Down keyboard arrow key for Previous or Next Command

To run commands you need to type them in the ‘Commands Window’ or select them from the ‘Progs’ menu

‘Freq’ is the command to get a frequency count. You also need to specify the file you want to use.

Freq results

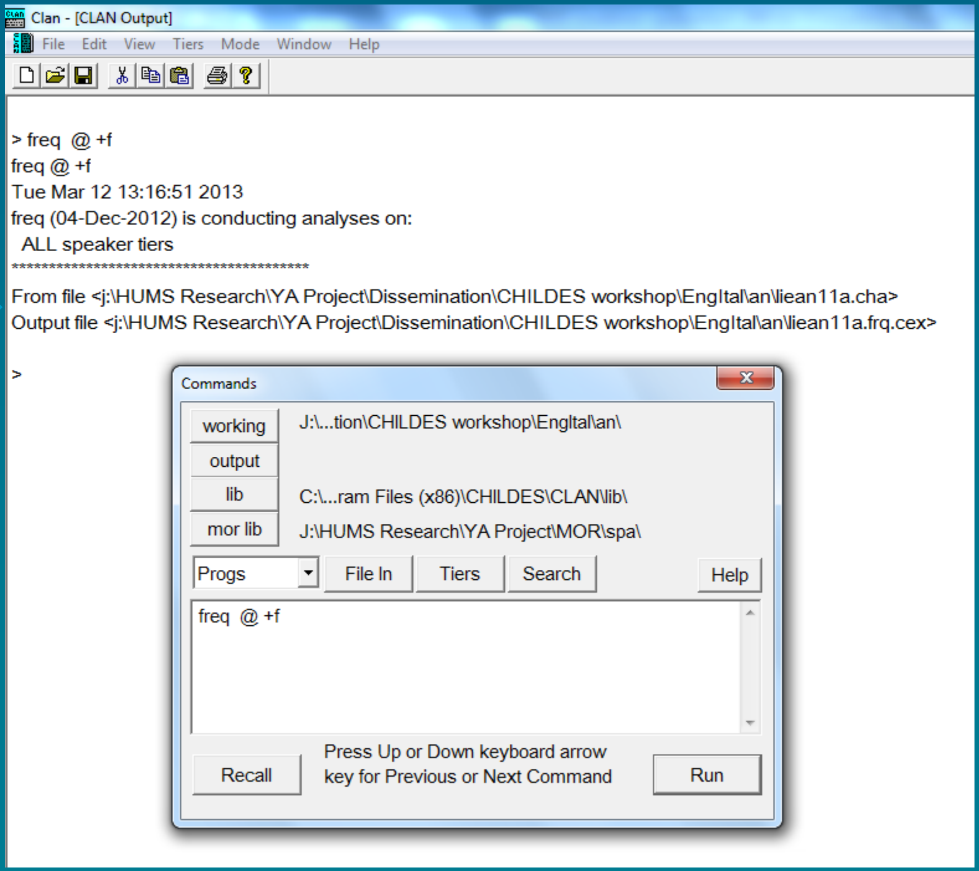
Type and token counts are provided too



```
  2 up
  3 used
 13 very
  1 visit
  2 wait
  1 waitress
  2 want
  1 warm
  5 was
  1 washer
  1 watch
  7 we
  2 week
  5 well
  4 went
  1 westminster
  7 what
  8 when
  4 where
  3 which
  1 who
  4 will
  4 with
  1 wonder
  1 words
 12 work
  1 write
 54 yeah
  1 year
  6 years
 25 yes
 69 you
  6 your
  2 youre
------------------------------
  304  Total number of different item types used
 1282  Total number of items (tokens)
0.237  Type/Token ratio

04dec12[E|TEXT]  148
```

Saving results in a separate file: +f

- By adding +f to your command, the CLAN program will save the results in a separate file.



```
> freq  @ +f
freq @ +f
Tue Mar 12 13:16:51 2013
freq (04-Dec-2012) is conducting analyses on:
  ALL speaker tiers
****************************************
From file <j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean11a.cha>
Output file <j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean11a.frq.cex>

>
```

working: J:\...tion\CHILDES workshop\EngItal\an\

lib: C:\...ram Files (x86)\CHILDES\CLAN\lib\

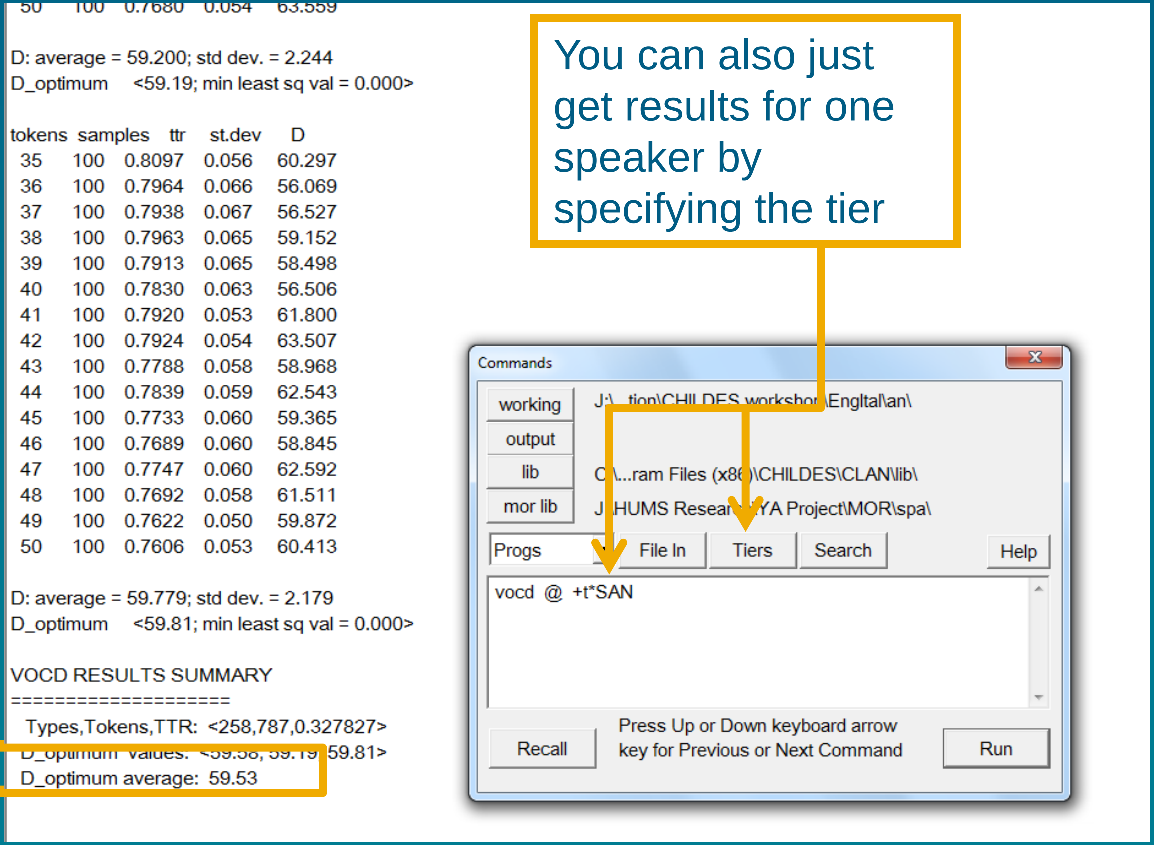
mor lib: J:\HUMS Research\YA Project\MOR\spa\

```
freq  @ +f
```

Press Up or Down keyboard arrow key for Previous or Next Command

Measure of Lexical Diversity: ‘D’

- D is a measure that estimates lexical diversity while accounting for text length
- Command = vocd



```
 50     100  0.7680  0.054    63.559

D: average = 59.200; std dev. = 2.244
D_optimum     <59.19; min least sq val = 0.000>

tokens  samples   ttr     st.dev      D
 35     100  0.8097  0.056    60.297
 36     100  0.7964  0.066    56.069
 37     100  0.7938  0.067    56.527
 38     100  0.7963  0.065    59.152
 39     100  0.7913  0.065    58.498
 40     100  0.7830  0.063    56.506
 41     100  0.7920  0.053    61.800
 42     100  0.7924  0.054    63.507
 43     100  0.7788  0.058    58.968
 44     100  0.7839  0.059    62.543
 45     100  0.7733  0.060    59.365
 46     100  0.7689  0.060    58.845
 47     100  0.7747  0.060    62.592
 48     100  0.7692  0.058    61.511
 49     100  0.7622  0.050    59.872
 50     100  0.7606  0.053    60.413

D: average = 59.779; std dev. = 2.179
D_optimum     <59.81; min least sq val = 0.000>

VOCD RESULTS SUMMARY
====================
  Types,Tokens,TTR:  <258,787,0.327827>
 D_optimum  values:  <59.58, 59.19, 59.81>
 D_optimum average:  59.53
```

You can also just get results for one speaker by specifying the tier

working: J:\...tion\CHILDES workshop\EngItal\an\

lib: C:\...ram Files (x86)\CHILDES\CLAN\lib\

mor lib: J:\HUMS Resear...YA Project\MOR\spa\

```
vocd  @  +t*SAN
```

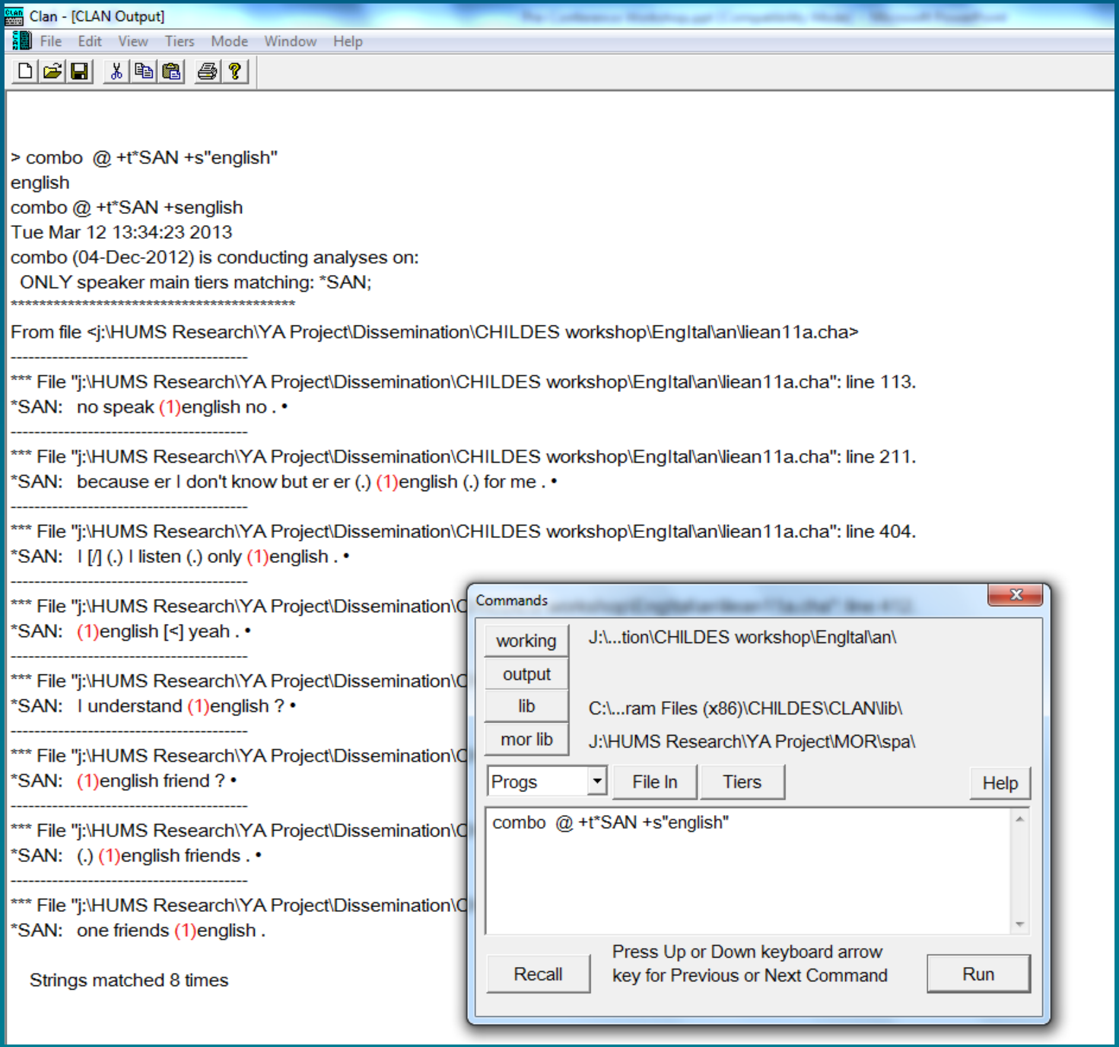
Press Up or Down keyboard arrow key for Previous or Next Command

Considerations for D

- Run freq to see what words get counted ‘as words’. Spelling mistakes could impact the result
- You can have some ‘words’ excluded from the D calculation using some additional commands
- You may only want the result from a specific speaker so you need to specify in the command

Search for words or sequence of words: COMBO

- Command= combo
- Use +s (string) and add the word in quotes: e.g., "english"



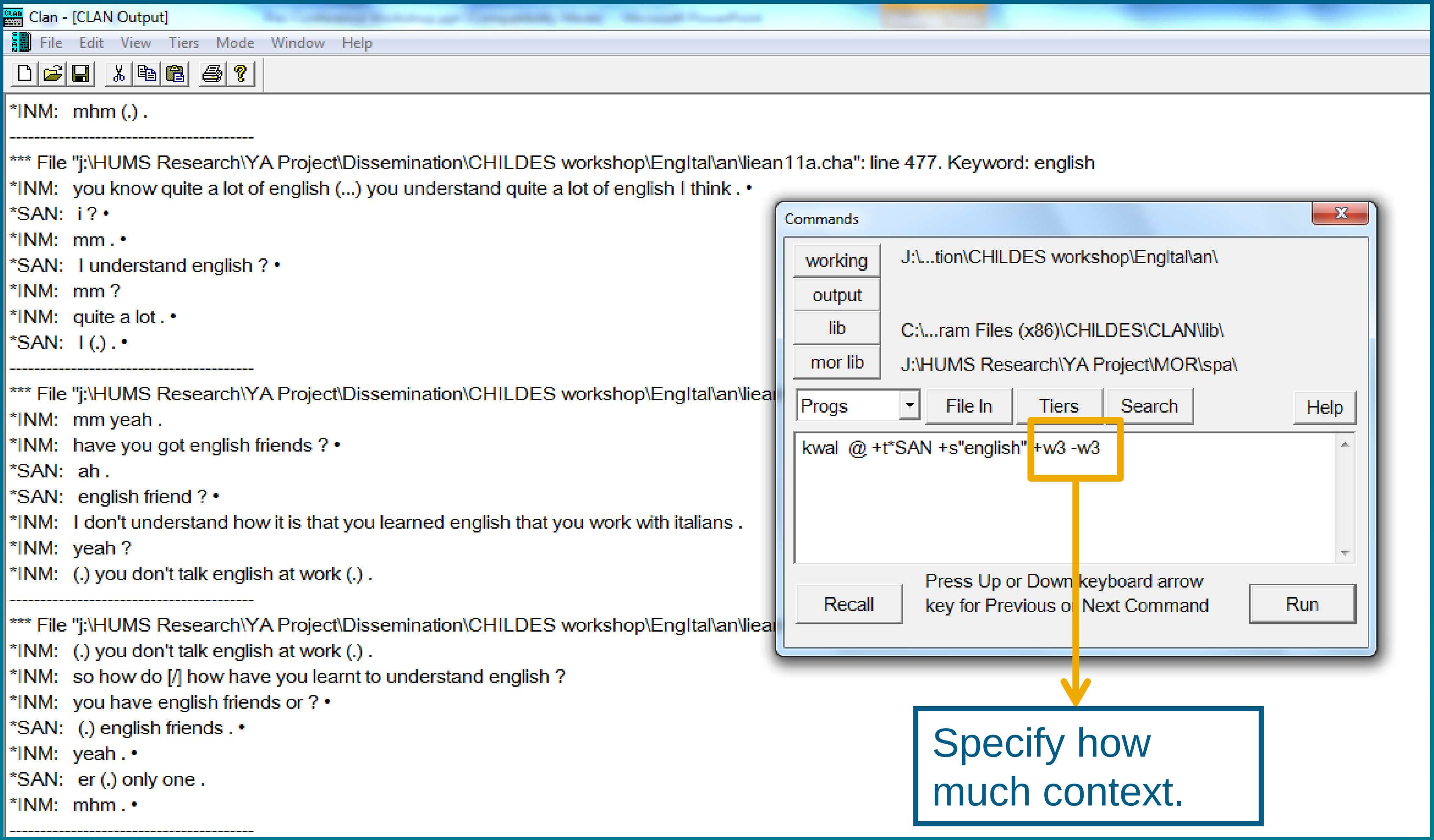
```
> combo  @ +t*SAN +s"english"
english
combo @ +t*SAN +senglish
Tue Mar 12 13:34:23 2013
combo (04-Dec-2012) is conducting analyses on:
  ONLY speaker main tiers matching: *SAN;
*******************************************
From file <j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean11a.cha>
----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean11a.cha": line 113.
*SAN:   no speak (1)english no . •
----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean11a.cha": line 211.
*SAN:   because er I don't know but er er (.) (1)english (.) for me . •
----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean11a.cha": line 404.
*SAN:   I [/] (.) I listen (.) only (1)english . •
----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\C
*SAN:   (1)english [<] yeah . •
----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\C
*SAN:   I understand (1)english ? •
----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\C
*SAN:   (1)english friend ? •
----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\C
*SAN:   (.) (1)english friends . •
----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\C
*SAN:   one friends (1)english .

    Strings matched 8 times
```

working J:\...tion\CHILDES workshop\EngItal\an\
output
lib C:\...ram Files (x86)\CHILDES\CLAN\lib\
mor lib J:\HUMS Research\YA Project\MOR\spa\

```
combo  @ +t*SAN +s"english"
```

Recall
Press Up or Down keyboard arrow
key for Previous or Next Command
Run

Search for words in context: KWAL



```
*INM:   mhm (.) .
-----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean11a.cha": line 477. Keyword: english
*INM:   you know quite a lot of english (...) you understand quite a lot of english I think . •
*SAN:   i ? •
*INM:   mm . •
*SAN:   I understand english ? •
*INM:   mm ?
*INM:   quite a lot . •
*SAN:   I (.) . •
-----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liea
*INM:   mm yeah .
*INM:   have you got english friends ? •
*SAN:   ah .
*SAN:   english friend ? •
*INM:   I don't understand how it is that you learned english that you work with italians .
*INM:   yeah ?
*INM:   (.) you don't talk english at work (.) .
-----------------------------------------
*** File "j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liea
*INM:   (.) you don't talk english at work (.) .
*INM:   so how do [/] how have you learnt to understand english ?
*INM:   you have english friends or ? •
*SAN:   (.) english friends . •
*INM:   yeah . •
*SAN:   er (.) only one .
*INM:   mhm . •
-----------------------------------------
```

working: J:\...tion\CHILDES workshop\EngItal\an\

output:

lib: C:\...ram Files (x86)\CHILDES\CLAN\lib\

mor lib: J:\HUMS Research\YA Project\MOR\spa\

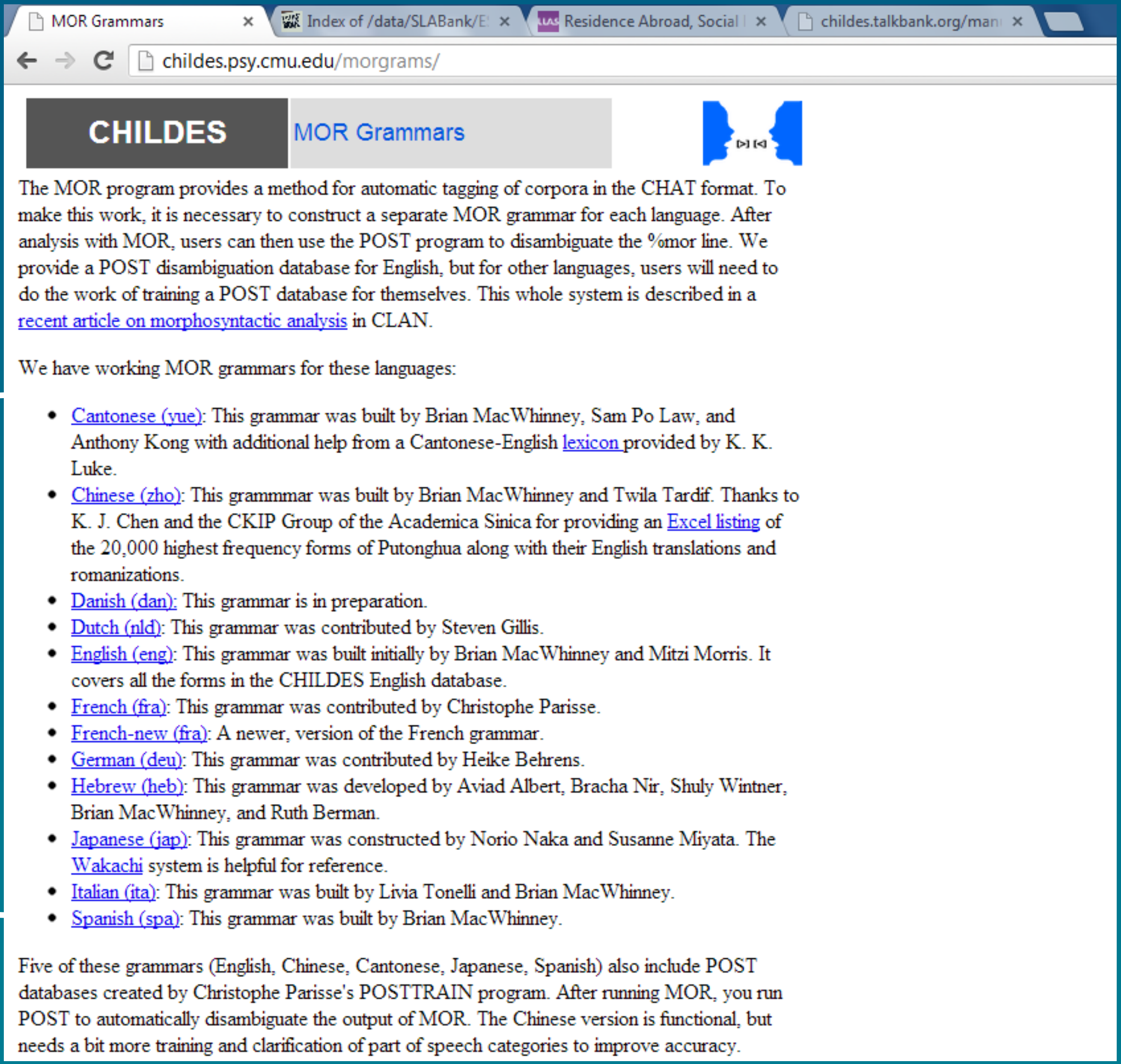
```
kwal  @ +t*SAN +s"english" +w3 -w3
```

Press Up or Down keyboard arrow key for Previous or Next Command

Specify how much context.

Part of Speech Tagger: MOR

Available in a number of languages



The MOR program provides a method for automatic tagging of corpora in the CHAT format. To make this work, it is necessary to construct a separate MOR grammar for each language. After analysis with MOR, users can then use the POST program to disambiguate the %mor line. We provide a POST disambiguation database for English, but for other languages, users will need to do the work of training a POST database for themselves. This whole system is described in a recent article on morphosyntactic analysis in CLAN.

We have working MOR grammars for these languages:

- Cantonese (yue): This grammar was built by Brian MacWhinney, Sam Po Law, and Anthony Kong with additional help from a Cantonese-English lexicon provided by K. K. Luke.
- Chinese (zho): This grammmar was built by Brian MacWhinney and Twila Tardif. Thanks to K. J. Chen and the CKIP Group of the Academica Sinica for providing an Excel listing of the 20,000 highest frequency forms of Putonghua along with their English translations and romanizations.
- Danish (dan): This grammar is in preparation.
- Dutch (nld): This grammar was contributed by Steven Gillis.
- English (eng): This grammar was built initially by Brian MacWhinney and Mitzi Morris. It covers all the forms in the CHILDES English database.
- French (fra): This grammar was contributed by Christophe Parisse.
- French-new (fra): A newer, version of the French grammar.
- German (deu): This grammar was contributed by Heike Behrens.
- Hebrew (heb): This grammar was developed by Aviad Albert, Bracha Nir, Shuly Wintner, Brian MacWhinney, and Ruth Berman.
- Japanese (jap): This grammar was constructed by Norio Naka and Susanne Miyata. The Wakachi system is helpful for reference.
- Italian (ita): This grammar was built by Livia Tonelli and Brian MacWhinney.
- Spanish (spa): This grammar was built by Brian MacWhinney.

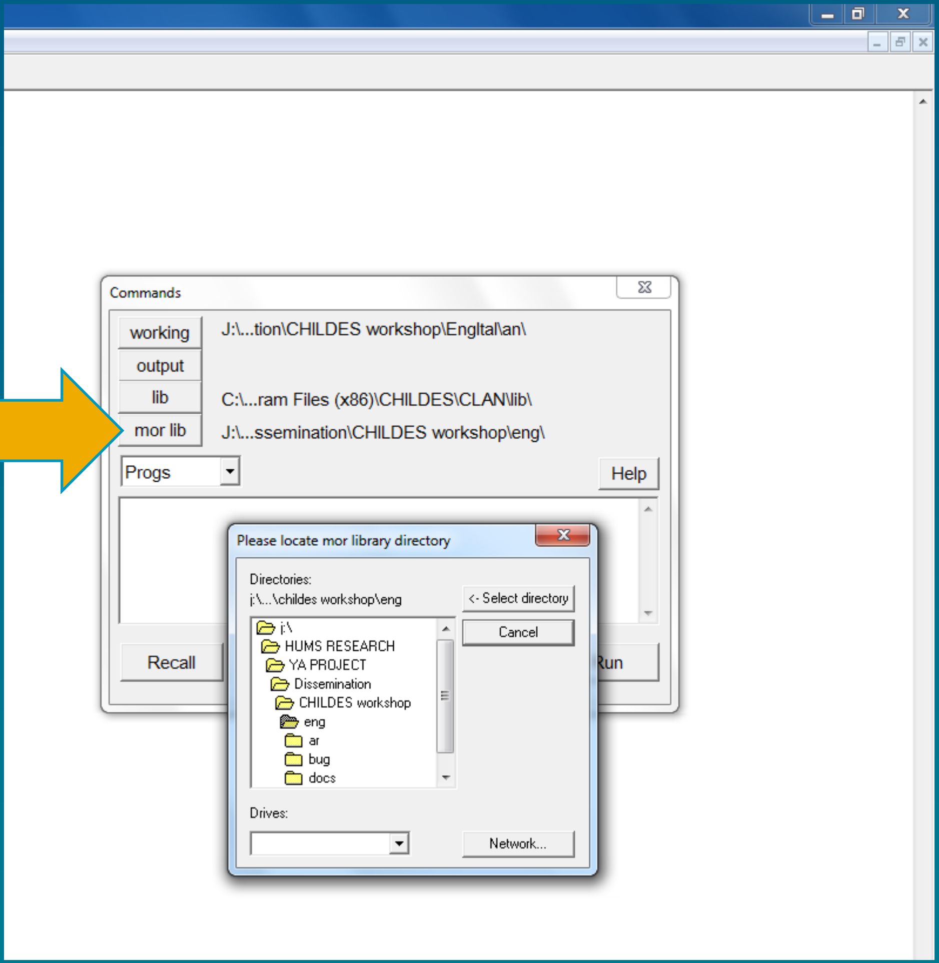
Five of these grammars (English, Chinese, Cantonese, Japanese, Spanish) also include POST databases created by Christophe Parisse's POSTTRAIN program. After running MOR, you run POST to automatically disambiguate the output of MOR. The Chinese version is functional, but needs a bit more training and clarification of part of speech categories to improve accuracy.

English MOR

- You must first download the MOR onto your computer. Make sure you know where to locate it.
- MOR provides a morphosyntactic ‘tag’ for each word in the corpus, adding an additional layer of grammatical information
- The benefits of having a tagged text is that you can search by grammatical category (e.g., noun, verb, etc).

Selecting the MOR Library

- Make sure you select the right location



Commands

working J:\...tion\CHILDES workshop\Engltal\an\

output

lib C:\...ram Files (x86)\CHILDES\CLAN\lib\

mor lib J:\...ssemination\CHILDES workshop\eng\

Please locate mor library directory

Directories:
j:\...\childes workshop\eng

```
j:\
 HUMS RESEARCH
  YA PROJECT
   Dissemination
    CHILDES workshop
     eng
      ar
      bug
      docs
```

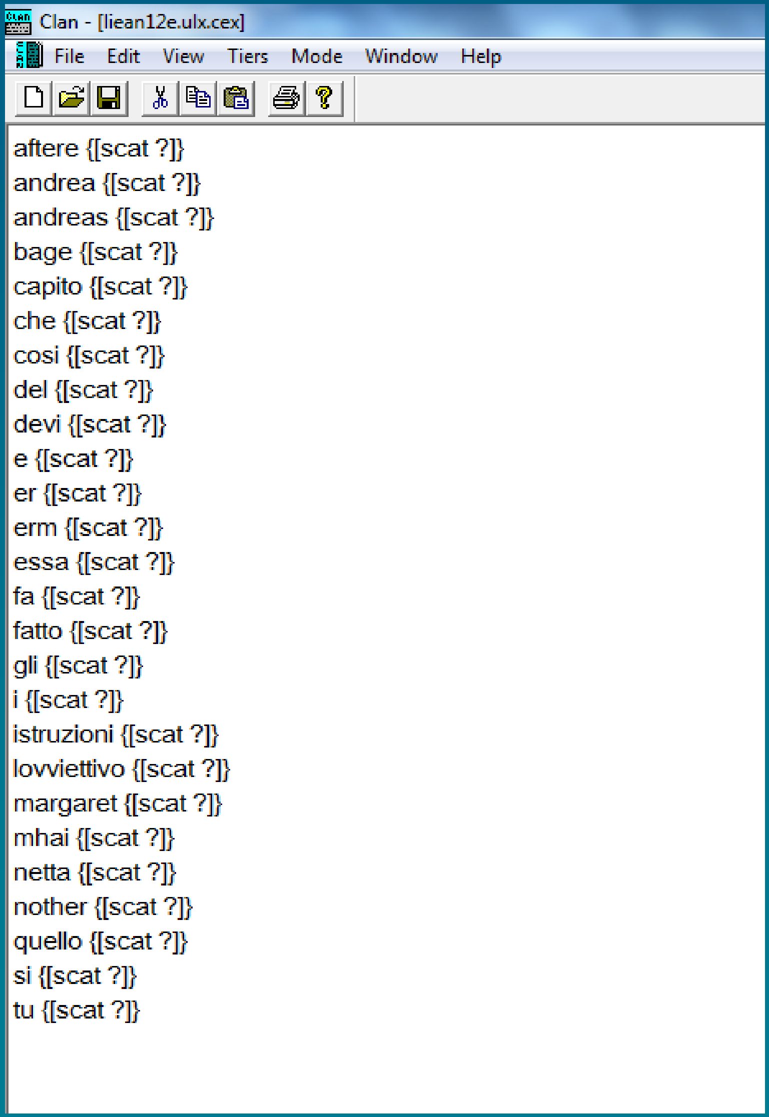
Drives:

MOR–ing Files: Step 1

- Check for ‘new’ vocabulary
- The MOR dictionary or ‘lexicon’ that the program uses has been built word by word.
- Therefore, some real words in your corpus may not already be in the lexicon and you will need to add them manually.

Checking for ‘new’ words

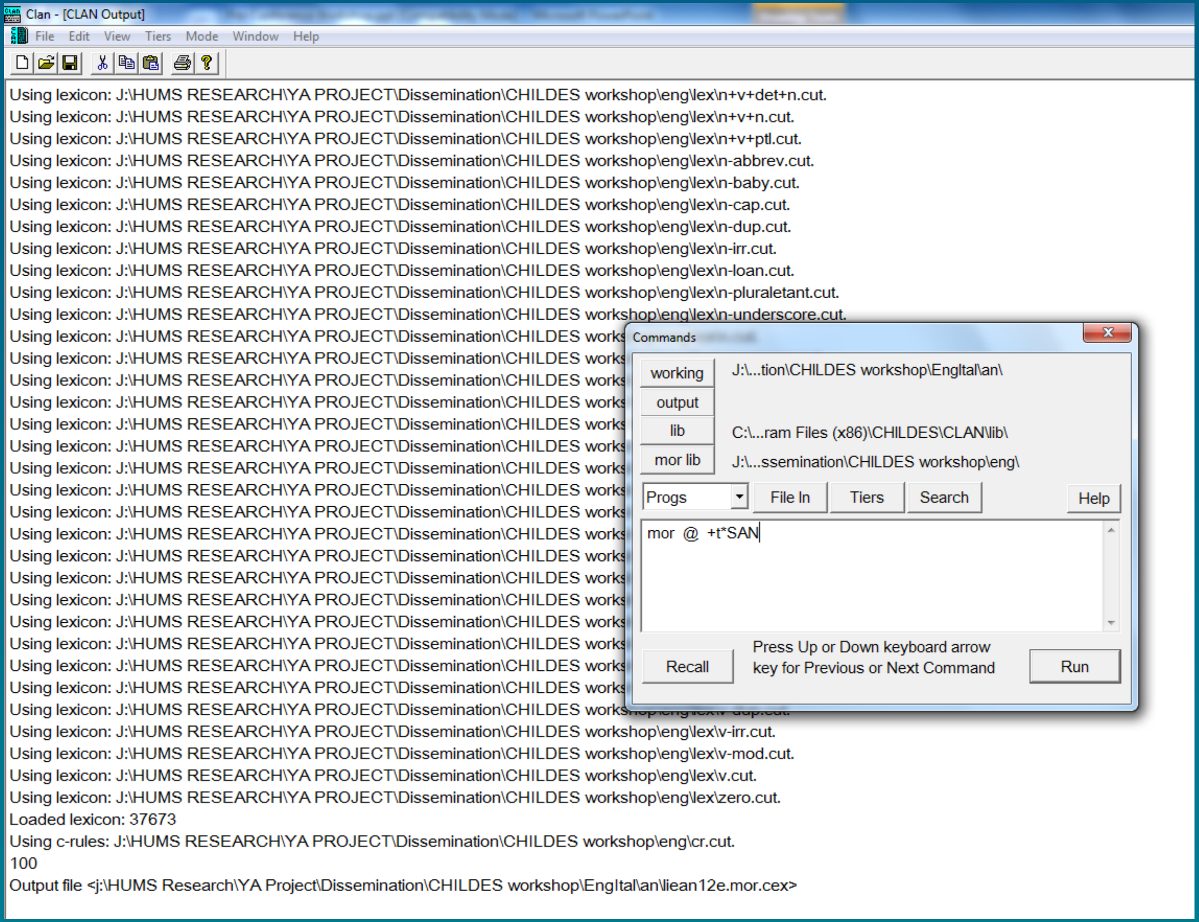
- Command=
  mor +xl
- A list of these words is automatically created in a new file.
- Some will be words to add to the lexicon. Others might be transcription errors



```
aftere {[scat ?]}
andrea {[scat ?]}
andreas {[scat ?]}
bage {[scat ?]}
capito {[scat ?]}
che {[scat ?]}
cosi {[scat ?]}
del {[scat ?]}
devi {[scat ?]}
e {[scat ?]}
er {[scat ?]}
erm {[scat ?]}
essa {[scat ?]}
fa {[scat ?]}
fatto {[scat ?]}
gli {[scat ?]}
i {[scat ?]}
istruzioni {[scat ?]}
lovviettivo {[scat ?]}
margaret {[scat ?]}
mhai {[scat ?]}
netta {[scat ?]}
nother {[scat ?]}
quello {[scat ?]}
si {[scat ?]}
tu {[scat ?]}
```

Step 2: MOR–ing

- Command: mor
- Specify the file (or files) and the tier.
- A new file is created with a new extension



```
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n+v+det+n.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n+v+n.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n+v+ptl.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n-abbrev.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n-baby.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n-cap.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n-dup.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n-irr.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n-loan.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n-pluraletant.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\n-underscore.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES works
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\v-dup.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\v-irr.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\v-mod.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\v.cut.
Using lexicon: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\lex\zero.cut.
Loaded lexicon: 37673
Using c-rules: J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\cr.cut.
100
Output file <j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean12e.mor.cex>
```

working: J:\...tion\CHILDES workshop\EngItal\an\

lib: C:\...ram Files (x86)\CHILDES\CLAN\lib\

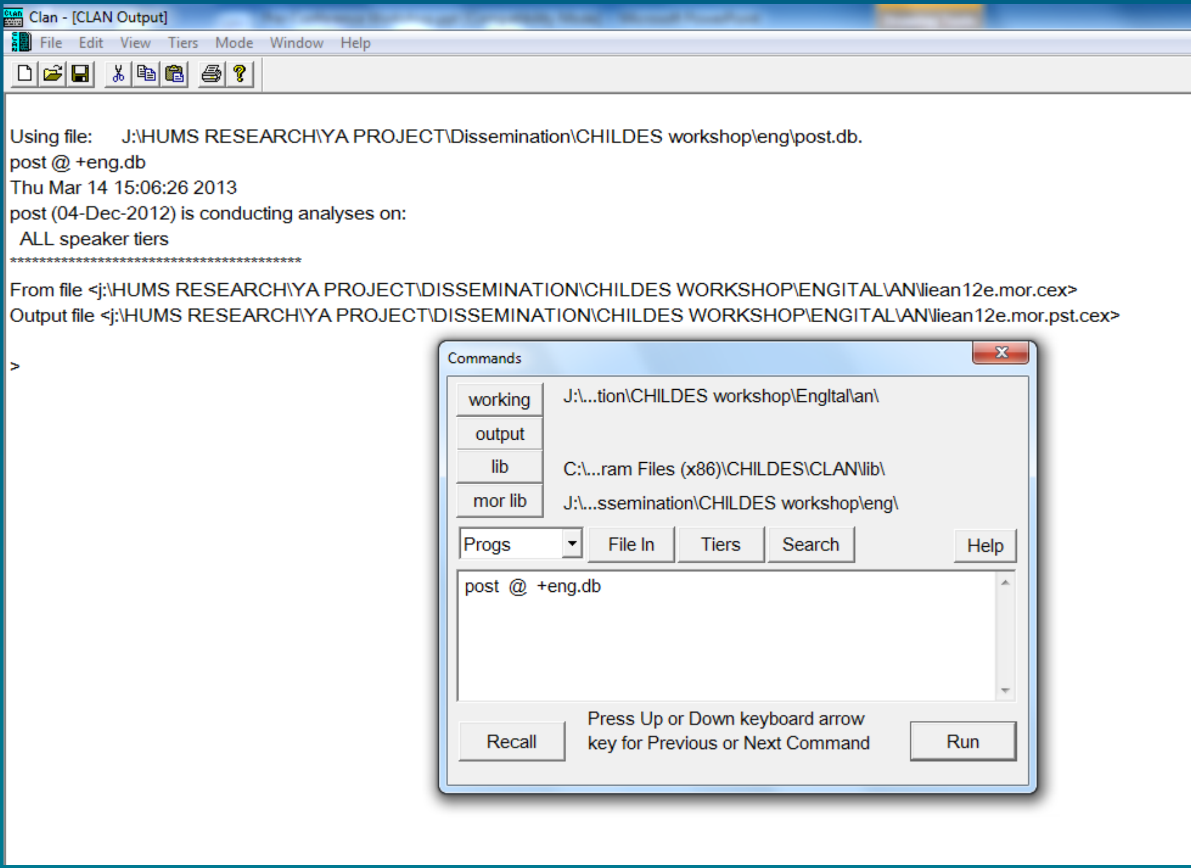
mor lib: J:\...ssemination\CHILDES workshop\eng\

```
mor  @  +t*SAN
```

Press Up or Down keyboard arrow key for Previous or Next Command

Step 3: Automatic Disambiguation

- Command: post +eng.db
- The program creates a new file with a new extension



```
Using file:     J:\HUMS RESEARCH\YA PROJECT\Dissemination\CHILDES workshop\eng\post.db.
post @ +eng.db
Thu Mar 14 15:06:26 2013
post (04-Dec-2012) is conducting analyses on:
  ALL speaker tiers
******************************************
From file <j:\HUMS RESEARCH\YA PROJECT\DISSEMINATION\CHILDES WORKSHOP\ENGITAL\AN\lean12e.mor.cex>
Output file <j:\HUMS RESEARCH\YA PROJECT\DISSEMINATION\CHILDES WORKSHOP\ENGITAL\AN\lean12e.mor.pst.cex>

>
```

working: J:\...tion\CHILDES workshop\EngItal\an\

output:

lib: C:\...ram Files (x86)\CHILDES\CLAN\lib\

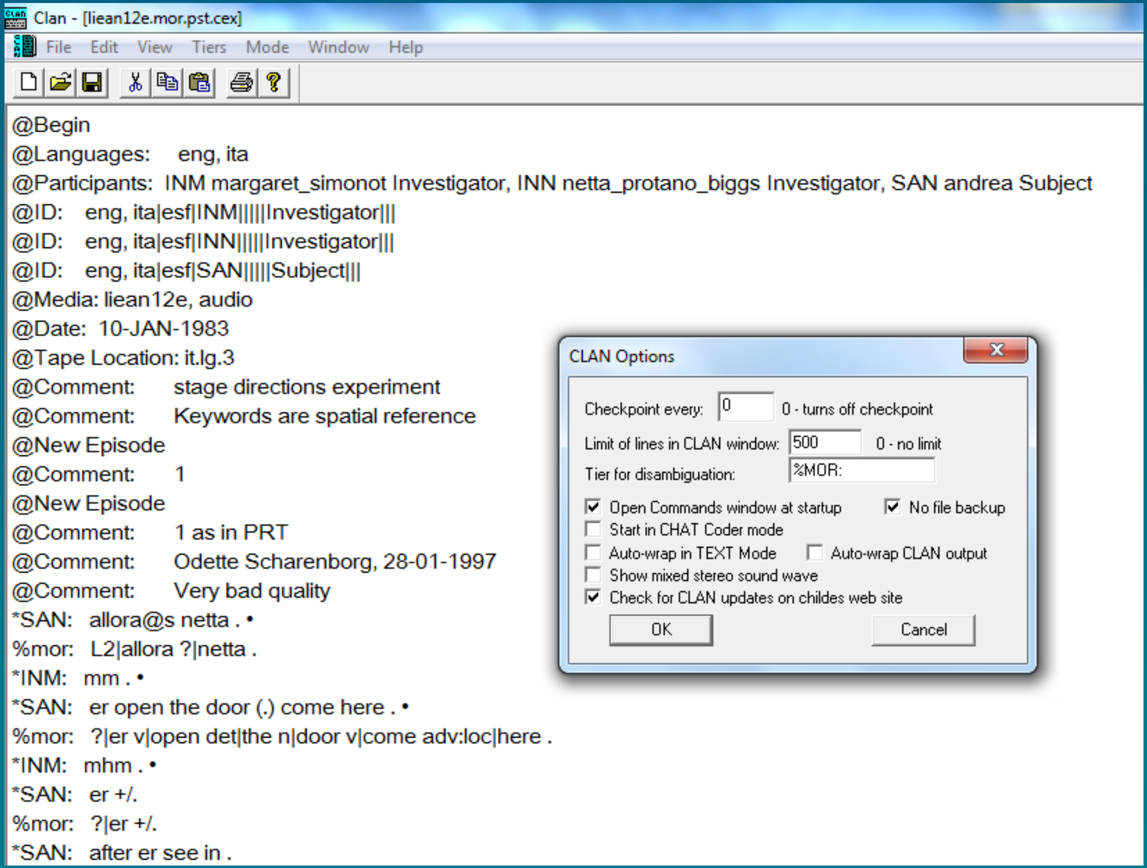
mor lib: J:\...ssemination\CHILDES workshop\eng\

```
post  @  +eng.db
```

Press Up or Down keyboard arrow key for Previous or Next Command

Step 4: Manual Disambiguation

- Open CLAN
- Under ‘Edit’, select ‘CLAN Option’
- Make sure the ‘tier for disambiguation’ is %MOR



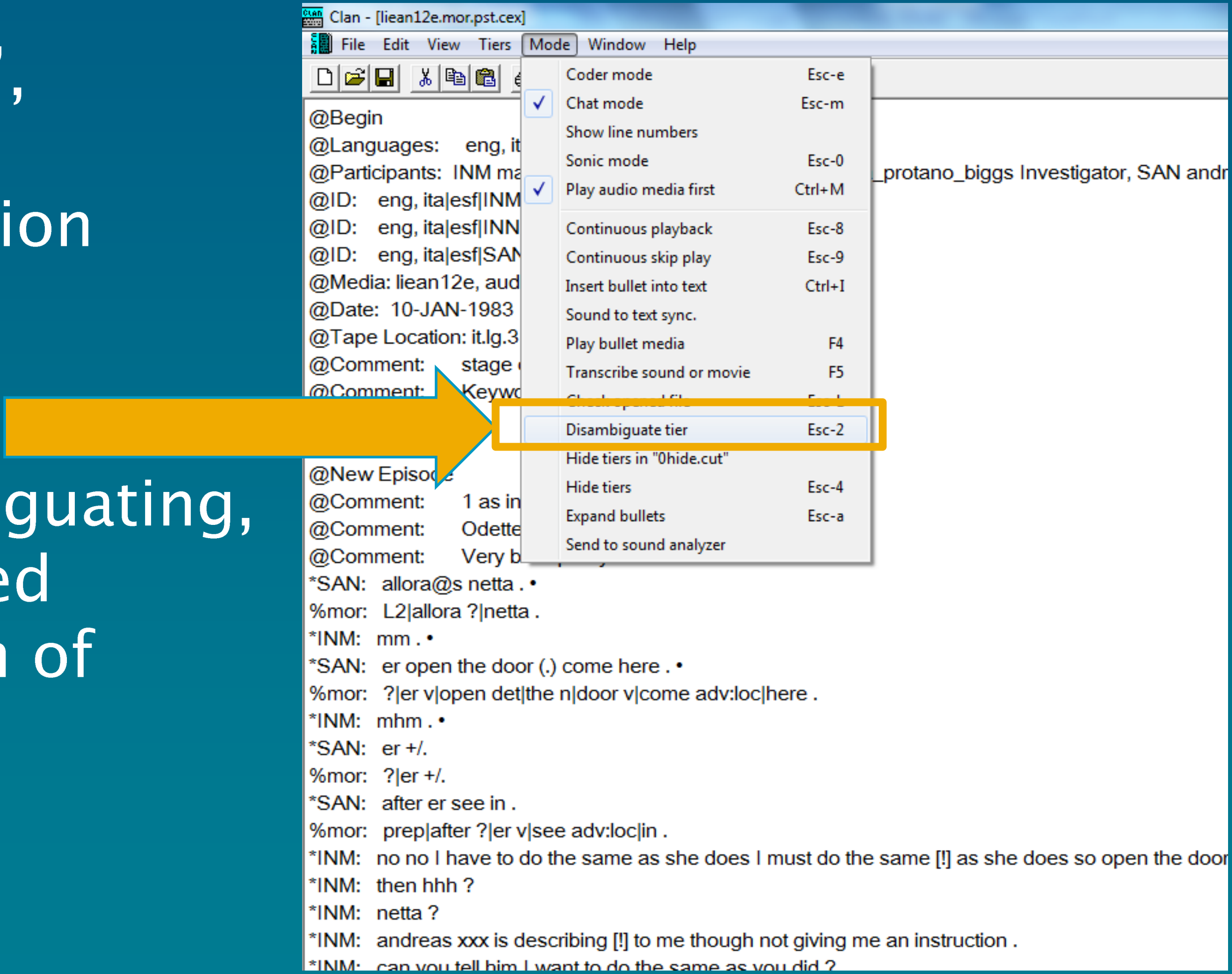
```
@Begin
@Languages:	eng, ita
@Participants:	INM margaret_simonot Investigator, INN netta_protano_biggs Investigator, SAN andrea Subject
@ID:	eng, ita|esf|INM|||||Investigator|||
@ID:	eng, ita|esf|INN|||||Investigator|||
@ID:	eng, ita|esf|SAN|||||Subject|||
@Media: liean12e, audio
@Date:	10-JAN-1983
@Tape Location: it.lg.3
@Comment:	stage directions experiment
@Comment:	Keywords are spatial reference
@New Episode
@Comment:	1
@New Episode
@Comment:	1 as in PRT
@Comment:	Odette Scharenborg, 28-01-1997
@Comment:	Very bad quality
*SAN:	allora@s netta . •
%mor:	L2|allora ?|netta .
*INM:	mm . •
*SAN:	er open the door (.) come here . •
%mor:	?|er v|open det|the n|door v|come adv:loc|here .
*INM:	mhm . •
*SAN:	er +/.
%mor:	?|er +/.
*SAN:	after er see in .
```

Checkpoint every: 0	0 - turns off checkpoint
Limit of lines in CLAN window: 500	0 - no limit
Tier for disambiguation: %MOR:

☑ Open Commands window at startup	☑ No file backup
☐ Start in CHAT Coder mode
☐ Auto-wrap in TEXT Mode	☐ Auto-wrap CLAN output
☐ Show mixed stereo sound wave
☑ Check for CLAN updates on childes web site

Manual Disambiguation

- Under ‘mode’, select ‘Disambiguation tier’
- If any tags need disambiguating, you’ll be asked at the bottom of the page.



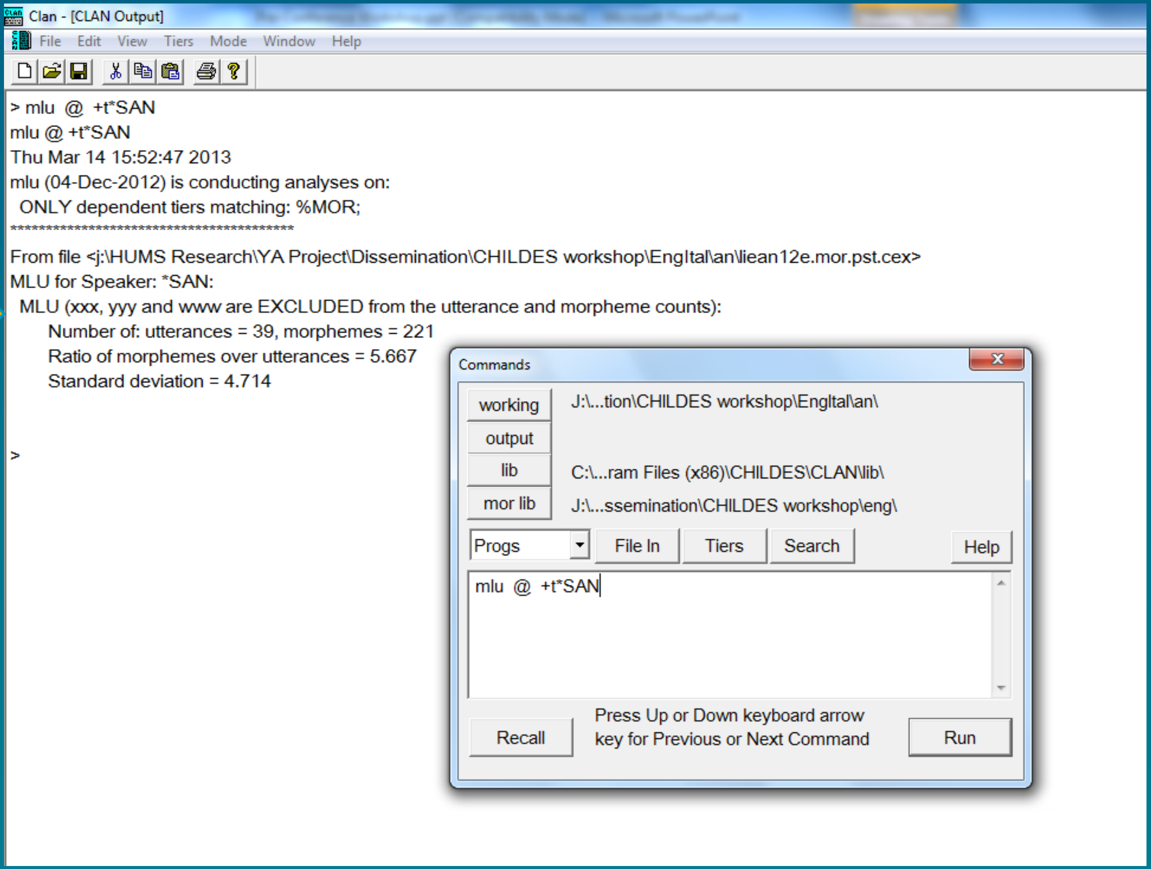
Clan - [liean12e.mor.pst.cex]
File Edit View Tiers Mode Window Help
Coder mode Esc-e
✓ Chat mode Esc-m
Show line numbers
Sonic mode Esc-0
✓ Play audio media first Ctrl+M
Continuous playback Esc-8
Continuous skip play Esc-9
Insert bullet into text Ctrl+I
Sound to text sync.
Play bullet media F4
Transcribe sound or movie F5
Disambiguate tier Esc-2
Hide tiers in "0hide.cut"
Hide tiers Esc-4
Expand bullets Esc-a
Send to sound analyzer

```
@Begin
@Languages:	eng, it
@Participants:	INM ma                    _protano_biggs Investigator, SAN andr
@ID:	eng, ita|esf|INM
@ID:	eng, ita|esf|INN
@ID:	eng, ita|esf|SAN
@Media: liean12e, aud
@Date:	10-JAN-1983
@Tape Location: it.lg.3
@Comment:	stage
@Comment:	Keywo
@New Episode
@Comment:	1 as in
@Comment:	Odette
@Comment:	Very b
*SAN:	allora@s netta . •
%mor:	L2|allora ?|netta .
*INM:	mm . •
*SAN:	er open the door (.) come here . •
%mor:	?|er v|open det|the n|door v|come adv:loc|here .
*INM:	mhm . •
*SAN:	er +/.
%mor:	?|er +/.
*SAN:	after er see in .
%mor:	prep|after ?|er v|see adv:loc|in .
*INM:	no no I have to do the same as she does I must do the same [!] as she does so open the door
*INM:	then hhh ?
*INM:	netta ?
*INM:	andreas xxx is describing [!] to me though not giving me an instruction .
*INM:	can you tell him I want to do the same as you did ?
```

Programs with a MOR file: Mean Length of Utterance (MLU)

- Command: mlu

Results are provided automatically or you can save in a separate file with +f command.



```
> mlu  @  +t*SAN
mlu @ +t*SAN
Thu Mar 14 15:52:47 2013
mlu (04-Dec-2012) is conducting analyses on:
  ONLY dependent tiers matching: %MOR;
*******************************************
From file <j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean12e.mor.pst.cex>
MLU for Speaker: *SAN:
  MLU (xxx, yyy and www are EXCLUDED from the utterance and morpheme counts):
      Number of: utterances = 39, morphemes = 221
      Ratio of morphemes over utterances = 5.667
      Standard deviation = 4.714


>
```

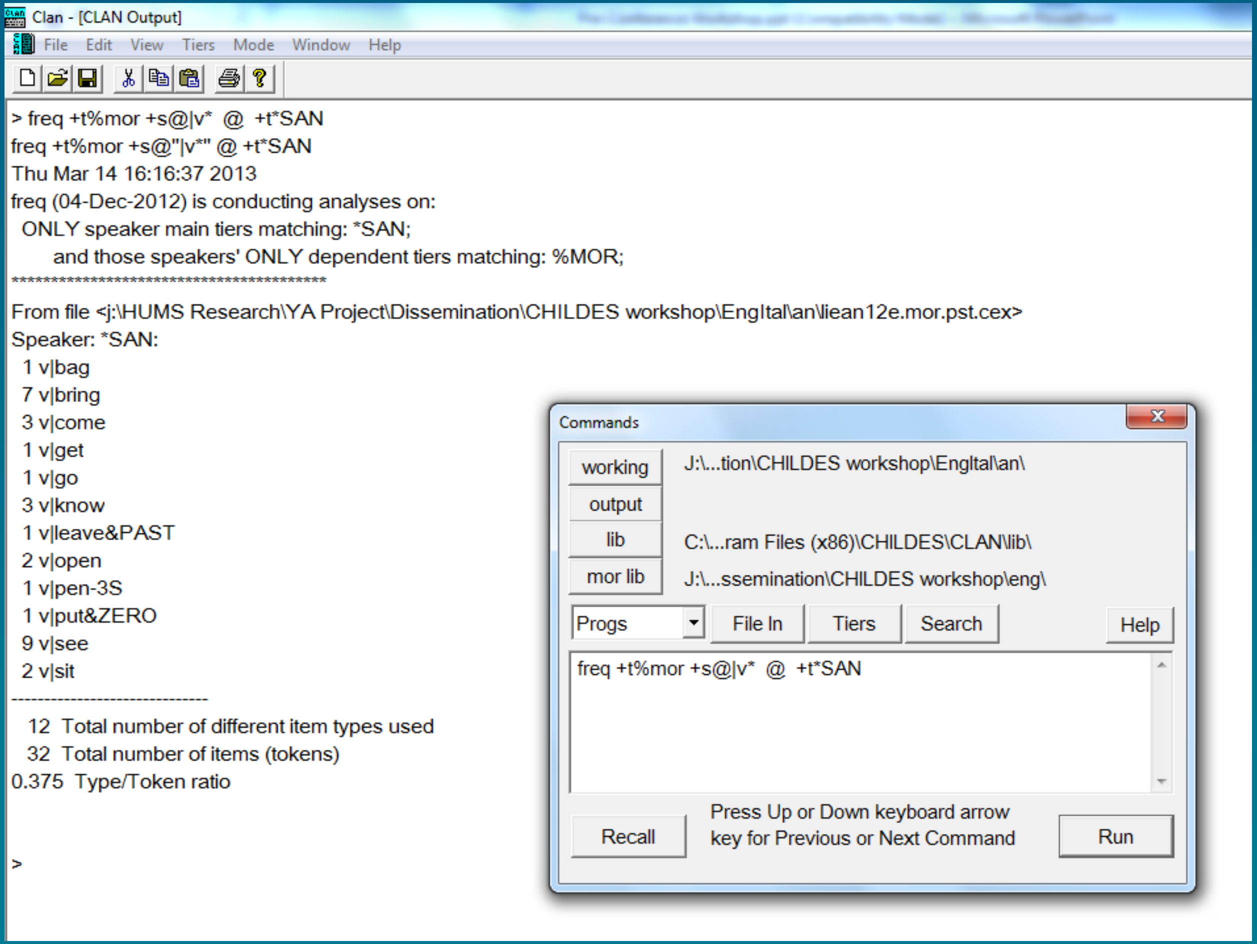
working J:\...tion\CHILDES workshop\EngItal\an\
output
lib C:\...ram Files (x86)\CHILDES\CLAN\lib\
mor lib J:\...ssemination\CHILDES workshop\eng\

```
mlu  @  +t*SAN
```

Press Up or Down keyboard arrow key for Previous or Next Command

Frequency list of tags

- For example: We can use the freq command to get a list of verbs



```
> freq +t%mor +s@|v*  @  +t*SAN
freq +t%mor +s@"|v*" @ +t*SAN
Thu Mar 14 16:16:37 2013
freq (04-Dec-2012) is conducting analyses on:
  ONLY speaker main tiers matching: *SAN;
     and those speakers' ONLY dependent tiers matching: %MOR;
****************************************
From file <j:\HUMS Research\YA Project\Dissemination\CHILDES workshop\EngItal\an\liean12e.mor.pst.cex>
Speaker: *SAN:
  1 v|bag
  7 v|bring
  3 v|come
  1 v|get
  1 v|go
  3 v|know
  1 v|leave&PAST
  2 v|open
  1 v|pen-3S
  1 v|put&ZERO
  9 v|see
  2 v|sit
------------------------------
  12  Total number of different item types used
  32  Total number of items (tokens)
0.375  Type/Token ratio

>
```

working J:\...tion\CHILDES workshop\EngItal\an\

output

lib C:\...ram Files (x86)\CHILDES\CLAN\lib\

mor lib J:\...ssemination\CHILDES workshop\eng\

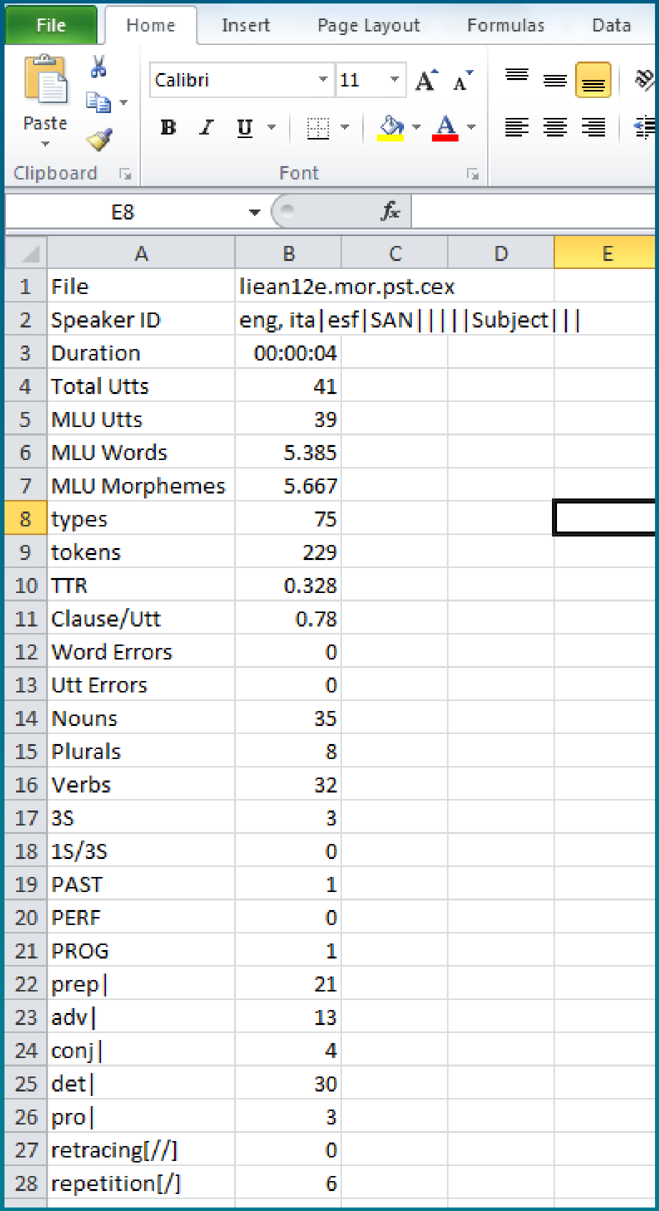
```
freq +t%mor +s@|v*  @  +t*SAN
```

Press Up or Down keyboard arrow key for Previous or Next Command

EVAL: results of many commands all at once

- Currently works for English files only

  type, tokens, TTR, number of nouns, amount of retracing, repetition, etc



	A	B	C	D	E
1	File	liean12e.mor.pst.cex			
2	Speaker ID	eng, ita|esf|SAN||||Subject||			
3	Duration	00:00:04			
4	Total Utts	41			
5	MLU Utts	39			
6	MLU Words	5.385			
7	MLU Morphemes	5.667			
8	types	75			
9	tokens	229			
10	TTR	0.328			
11	Clause/Utt	0.78			
12	Word Errors	0			
13	Utt Errors	0			
14	Nouns	35			
15	Plurals	8			
16	Verbs	32			
17	3S	3			
18	1S/3S	0			
19	PAST	1			
20	PERF	0			
21	PROG	1			
22	prep|	21			
23	adv|	13			
24	conj|	4			
25	det|	30			
26	pro|	3			
27	retracing[//]	0			
28	repetition[/]	6			

Exporting files to/from other programs

- Commands: CHAT2ELAN, CHAT2PRAAT
- Also, text files can be imported into CHAT:
  - Command: TEXTIN

Let’s do some practice!

- Download ‘EngItal.zip’ from Talkbank:
  - talkbank.org/data/SLABank/ESF/

- Move the files to an appropriate folder
- Open CLAN and click on ‘working’. Locate the folder with these files.

Run some commands!

- Try the ones we've demonstrated on the plain transcripts:
  - Freq
  - Vocd
  - Combo
  - Kwal
- You can also try MORing. First, download the English MOR:
  - childes.psy.cmu.edu/morgrams/

To learn more:

- Join the CHILDES online community
  - childes.psy.cmu.edu/tools/email.html
  - Visit the FLLOC and SPLLOC websites

Thank you!

K.McManus@soton.ac.uk

N.Tracy-Ventura@soton.ac.uk