AphasiaBank:

Preliminary Lexical, Morphosyntactic, and Error Analyses

Brian MacWhinney, Audrey Holland,
Davida Fromm, Margaret Forbes, Heather Wright
&
Leonid Spektor
Funded by : NIH-NIDCD grant R01-DC008524 for 2007-2012TalkBank.org

OR.....

Using AphasiaBank to develop and apply a lexicon for Cinderella:
An illustration

What is AphasiaBank?

AphasiaBank is an archival database.
It collects and analyzes samples of the discourse of individuals with aphasia and normal participants across a range of tasks.

The Major Goal of AphasiaBank

To assemble a large repository of video-recorded discourse samples, transcribed in a format that facilitates extensive computerized language analyses

To make it available for use by authorized researchers to answer a variety of questions about aphasic language.

3 separate, but integrated tools for accomplishing this goal

1. CHAT = The transcription & coding format*

2. CLAN = The analysis program *

3. The database

An important secondary goal....

AphasiaBank is already a rich teaching resource via its archived videotapes
and accompanying transcripts

The database for this talk

Non-aphasic adults

Age 23-80
M=58

Gender 16 F, 9 M

Handedness
Right = 23
Left=1
Ambi=1

Education 12-20
M=15

Aphasic adults

Age 30-80
M=64

Gender 8 F, 16 M

Handedness
Right = 21
Left=3

Education 12-20
M=15

WAB aphasia type
Anomic= 5
Conduction= 6
Broca= 3
Wernicke= 3
“Not aphasic”= 2*
TCM= 1

Plan for this talk

Briefly summarize the protocol

Describe a few core analyses and the commands that make them happen

Present data using these analyses with 25 normal and 24 aphasic speakers to describe and compare their Cinderella stories

Protocol

Systematic demographic information
(Including WAB AQ subtest scores)

Some tests:
Short BNT (2001)
Verb Naming Test (NAVS,revised field test)
AphasiaBank Repetition Test (2007)

Discourse measures
Stroke story
Important personal event
Picture descriptions
broken window
cat in tree
refused umbrella
flood
Cinderella narrative*
Pb&j

Many commands: We describe 3 here

GEM=finds areas of text that are marked with text markers (eg, Cinderella story is marked @G by transcriber)

FREQ= produces a list of the words in a file or set of files, along with their frequency counts, and calculates a type-token ratio (command freq)

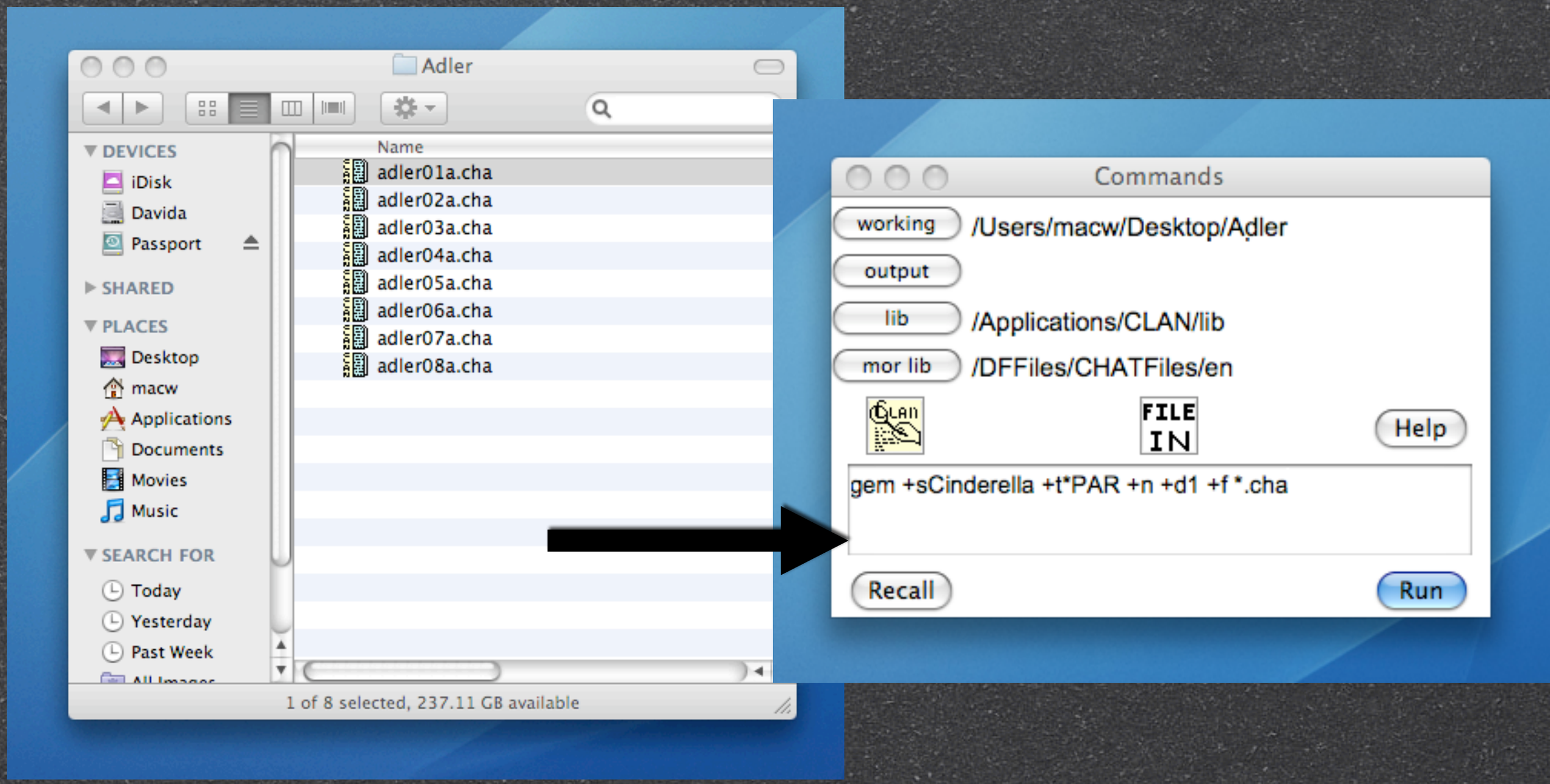
MOR--A Special Case

A grammar for English, downloadable from CHILDES, into the CLAN library (LIB)

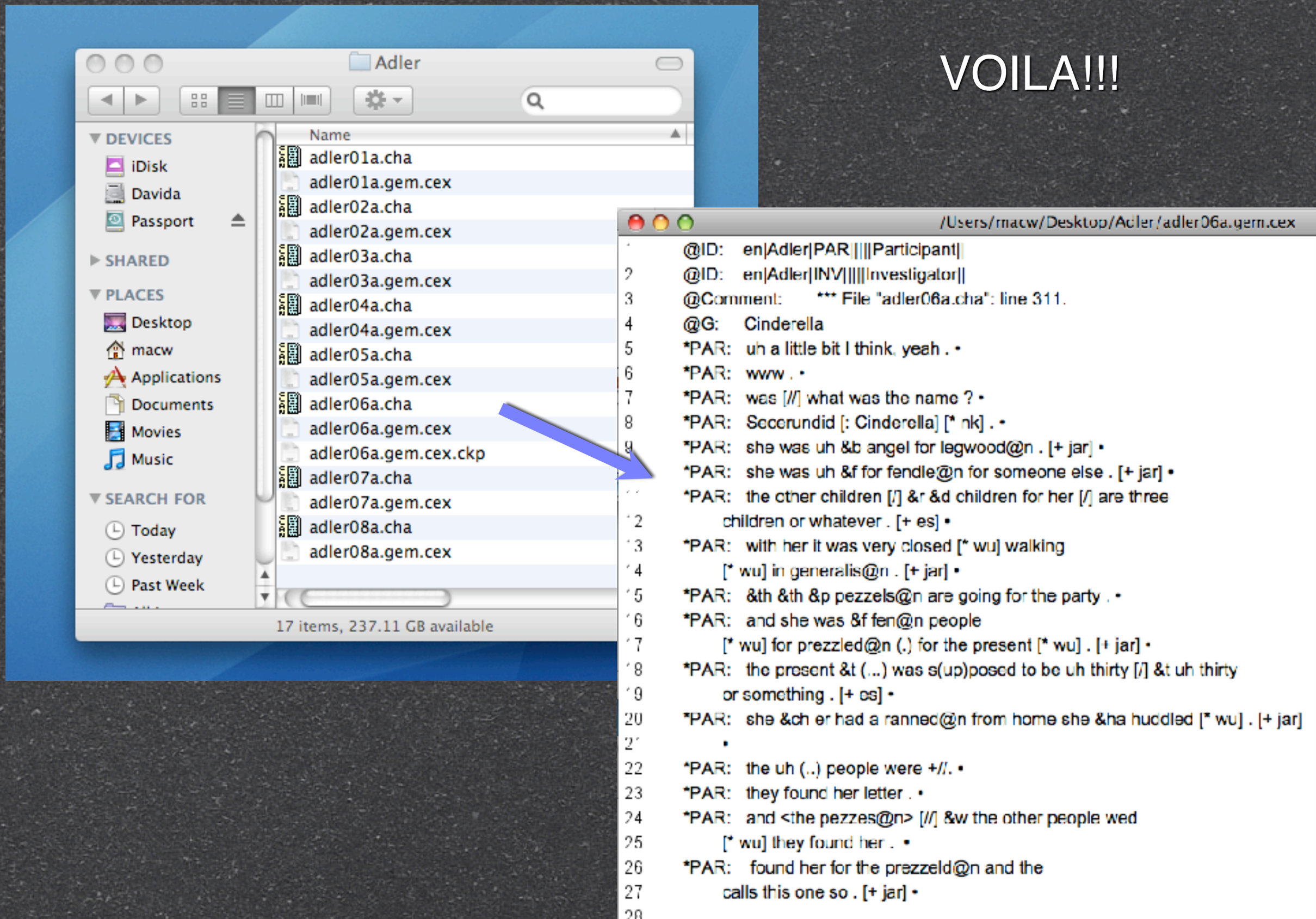
Once installed,

MOR (command MOR) = creates a line below each line of transcribed text showing word stems and parts of speech

GEM Command


Adler
DEVICES
iDisk
Davida
Passport
SHARED
PLACES
Desktop
macw
Applications
Documents
Movies
Music
SEARCH FOR
Today
Yesterday
Past Week
Name
adler01a.cha
adler02a.cha
adler03a.cha
adler04a.cha
adler05a.cha
adler06a.cha
adler07a.cha
adler08a.cha
1 of 8 selected, 237.11 GB available
Commands
working
/Users/macw/Desktop/Adler
output
lib
/Applications/CLAN/lib
mor lib
/DFFiles/CHATFiles/en
FILE IN
Help
gem +sCinderella +t*PAR +n +d1 +f *.cha
Recall
Run

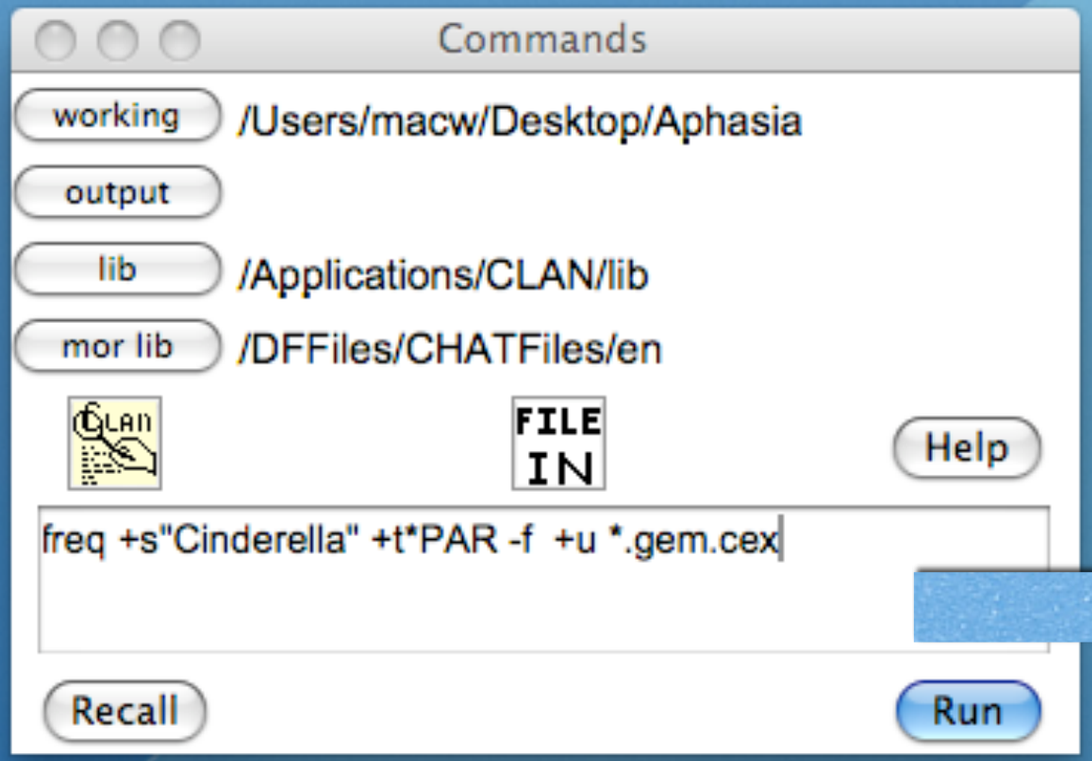
VOILA!!!


Adler
DEVICES
iDisk
Davida
Passport
SHARED
PLACES
Desktop
macw
Applications
Documents
Movies
Music
SEARCH FOR
Today
Yesterday
Past Week
Name
adler01a.cha
adler01a.gem.cex
adler02a.cha
adler02a.gem.cex
adler03a.cha
adler03a.gem.cex
adler04a.cha
adler04a.gem.cex
adler05a.cha
adler05a.gem.cex
adler06a.cha
adler06a.gem.cex
adler06a.gem.cex.ckp
adler07a.cha
adler07a.gem.cex
adler08a.cha
adler08a.gem.cex
17 items, 237.11 GB available
/Users/macw/Desktop/Adler/adler06a.gem.cex
@ID: en|Adler|PAR||||Participant||
@ID: en|Adler|INV|||||Investigator||
@Comment: *** File "adler06a.cha": line 311.
@G: Cinderella
*PAR: uh a little bit I think, yeah . •
*PAR: www . •
*PAR: was [//] what was the name ? •
PAR: Sccorundid [: Cinderella] [nk] . •
*PAR: she was uh &b angel for legwood@n . [+ jar] •
*PAR: she was uh &f for fendle@n for someone else . [+ jar] •
*PAR: the other children [/] &r &d children for her [/] are three
children or whatever . [+ es] •
PAR: with her it was very closed [wu] walking
[* wu] in generalis@n . [+ jar] •
*PAR: &th &th &p pezzels@n are going for the party . •
*PAR: and she was &f fen@n people
[* wu] for prezzled@n (.) for the present [* wu] . [+ jar] •
*PAR: the present &t (...) was s(up)posed to be uh thirty [/] &t uh thirty
or something . [+ es] •
PAR: she &ch er had a ranned@n from home she &ha huddled [wu] . [+ jar]
•
*PAR: the uh (..) people were +//. •
*PAR: they found her letter . •
*PAR: and <the pezzes@n> [//] &w the other people wed
[* wu] they found her . •
*PAR: found her for the prezzeld@n and the
calls this one so . [+ jar] •

FREQ

- Using FREQ to ask for a frequency count of the word “Cinderella” in all of the aphasic Cinderella files created by the GEM command, we found 88 uses of “Cinderella”

FREQ Cinderella




```
> freq +s"Cinderella" +t*PAR -f  +u *.gem.cex
freq +sCinderella +t*PAR -f +u *.gem.cex
Tue May 12 14:26:26 2009
freq (11-May-2009) is conducting analyses on:
  ONLY speaker main tiers matching: *PAR;
****************************************
From file <adler01a.gem.cex>
From file <adler02a.gem.cex>
From file <adler03a.gem.cex>
From file <adler04a.gem.cex>
From file <adler05a.gem.cex>
From file <adler06a.gem.cex>
From file <adler07a.gem.cex>
From file <adler08a.gem.cex>
From file <tucson01a.gem.cex>
From file <tucson02a.gem.cex>
From file <tucson03a.gem.cex>
From file <tucson04a.gem.cex>
From file <tucson05a.gem.cex>
From file <tucson06a.gem.cex>
From file <tucson07a.gem.cex>
From file <tucson08a.gem.cex>
From file <tucson09a.gem.cex>
From file <tucson10a.gem.cex>
From file <tucson11a.gem.cex>
From file <tucson12a.gem.cex>
From file <tucson13a.gem.cex>
From file <tucson14a.gem.cex>
From file <tucson15a.gem.cex>
From file <tucson16a.gem.cex>
 88 Cinderella
```

/Users/macw/Desktop/Adler/adler06a.gem.cex

```
@ID:    en|Adler|PAR|||||Participant||
@ID:    en|Adler|INV|||||Investigator||
@Comment:       *** File "adler06a.cha": line 311.
@G:     Cinderella
*PAR:   uh a little bit I think, yeah . •
*PAR:   www . •
*PAR:   was [//] what was the name ? •
*PAR:   Secerundid [: Cinderella] [* nk] . •
*PAR:   she was uh &b angel for legwood@n . [+ jar] •
*PAR:   she was uh &f for fendle@n for someone else . [+ jar] •
*PAR:   the other children [/] &r &d children for her [/] are three
        children or whatever . [+ es] •
*PAR:   with her it was very closed [* wu] walking
        [* wu] in generalis@n . [+ jar] •
*PAR:   &th &th &p pezzels@n are going for the party . •
*PAR:   and she was &f fen@n people
        [* wu] for prezzled@n (.) for the present [* wu] . [+ jar] •
*PAR:   the present &t (...) was s(up)posed to be uh thirty [/] &t uh thirty
        or something . [+ es] •
*PAR:   she &ch er had a ranned@n from home she &ha huddled [* wu] . [+ jar]
        •
*PAR:   the uh (..) people were +//. •
*PAR:   they found her letter . •
*PAR:   and <the pezzes@n> [//] &w the other people wed
        [* wu] they found her .  •
*PAR:    found her for the prezzeld@n and the
        calls this one so . [+ jar] •
```

Davida FREQed all the Cinderella aphasia files for the text associated with the replacement code [: Cinderella].

A few other goodies. . .

Cinderenella
Cinderlella
Cilawella
Cilawilla
Cilawillipa

Non-aphasic speakers

10 Most frequent nouns (in order)

Cinderella
ball
prince
slipper
mother, stepmother
dress
daughter, stepdaughter
fairy
godmother
sister,stepsister

10 Most frequent verbs (in order)

be
go
have
get
come
do
say
try
marry, remarry
know

Aphasic Speakers

10 Most frequent nouns (in order)

Cinderella
girl
ball
prince
mother, stepmother
home
man
slipper
shoe
sister, stepsister

10 Most frequent verbs (in order)

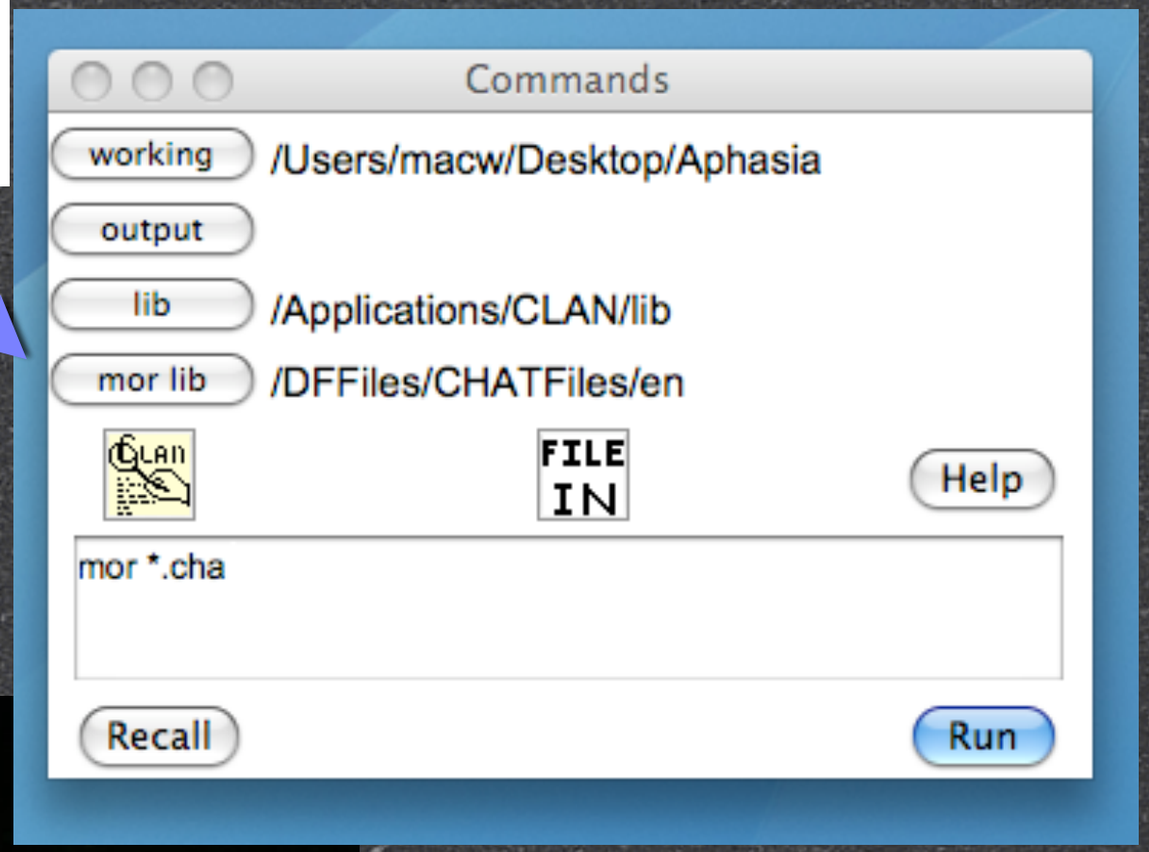
be
go
do
have
get
say
know
find
work
come

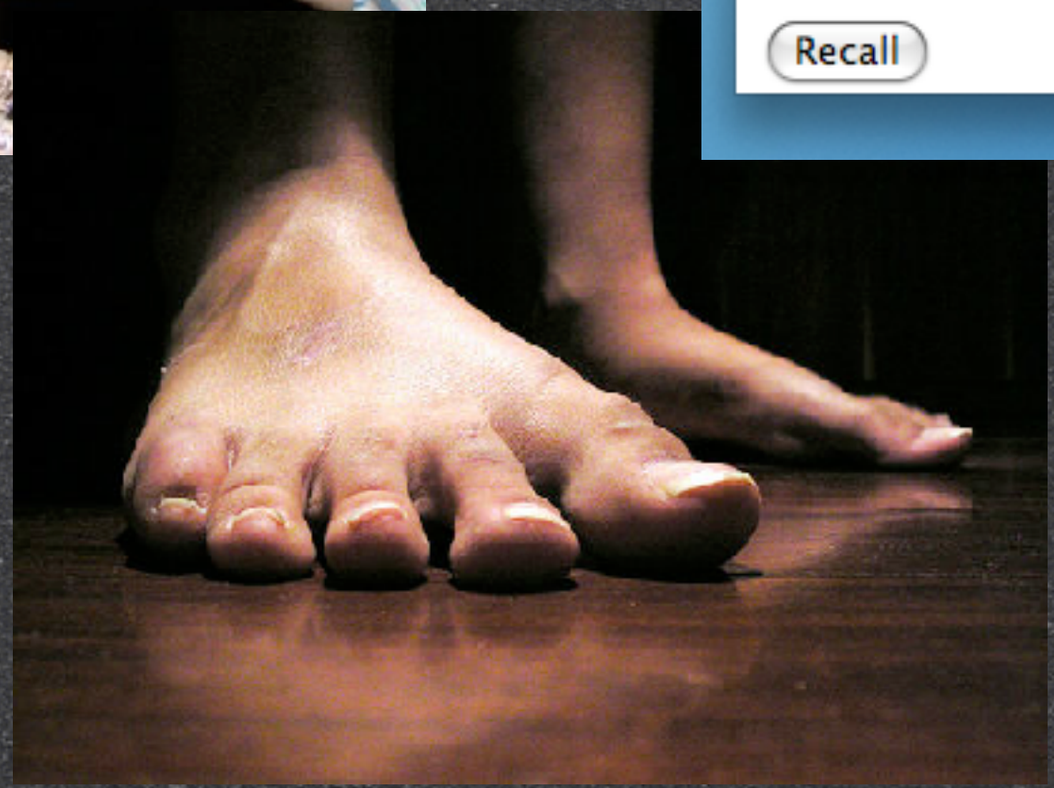
MOR

- Analyzes each utterance for parts of speech and word stems
- Inserts a line listing this information below each utterance in a transcript

MOR example

```
417
421    *PAR:  and the two ugly girls they had big feet . •
424    *PAR:  they didn't fit . •
426    *PAR:  and so she [//] &t he tried it on Cinderella . •
427
```

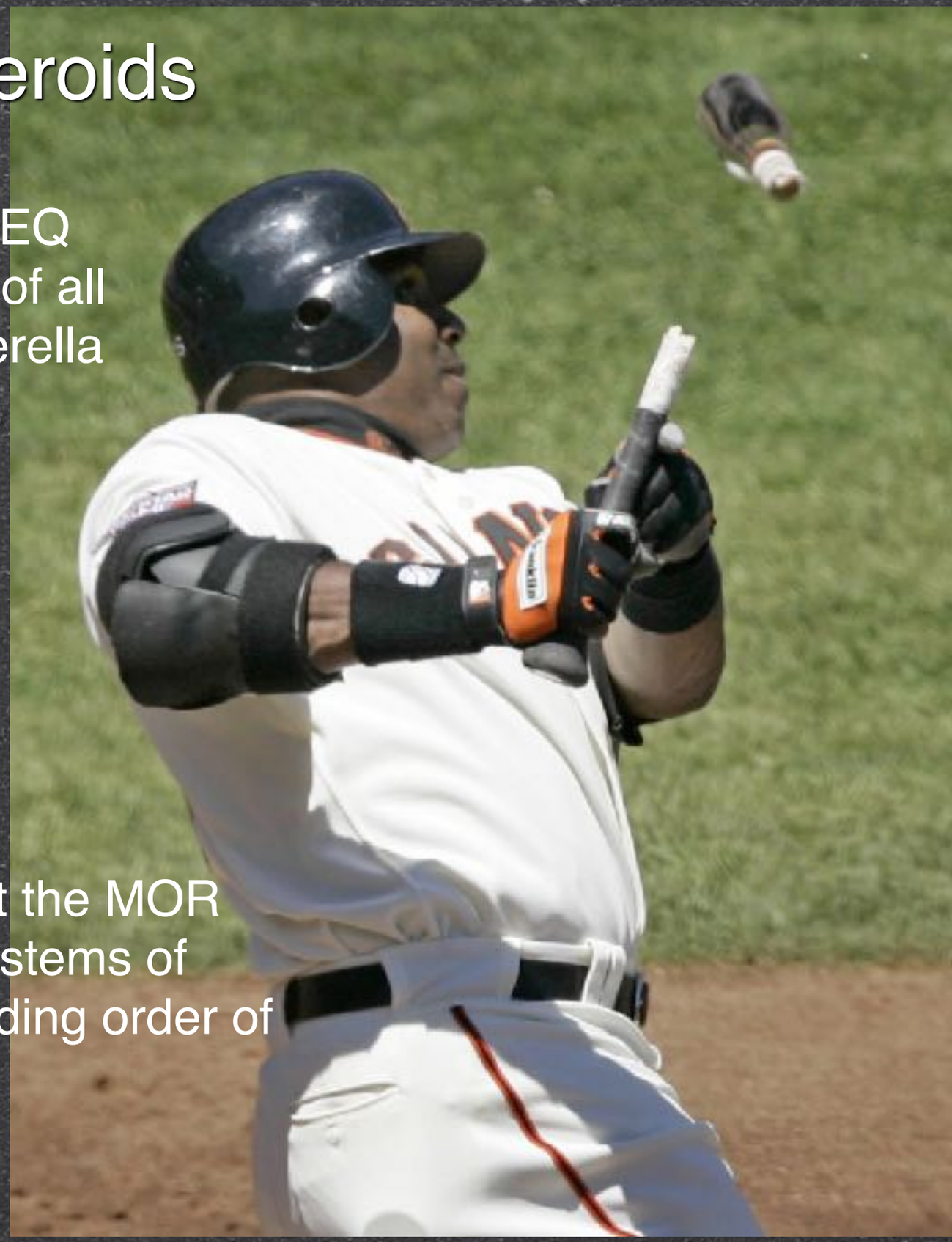




```
*PAR:   and the two ugly girls they had big feet . •
%mor:   conj:coo|and det|the det:num|two adj|ugly n|girl-PL pro|they v|have&PAST
        adj|big n|foot&PL .
*PAR:   they didn't fit . •
%mor:   pro|they aux|do&PAST~neg|not n|fit .
*PAR:   and so she [//] &t he tried it on Cinderella . •
%mor:   conj:coo|and co|so pro|he v|try-PAST pro|it prep|on
        n:prop|Cinderella .
```

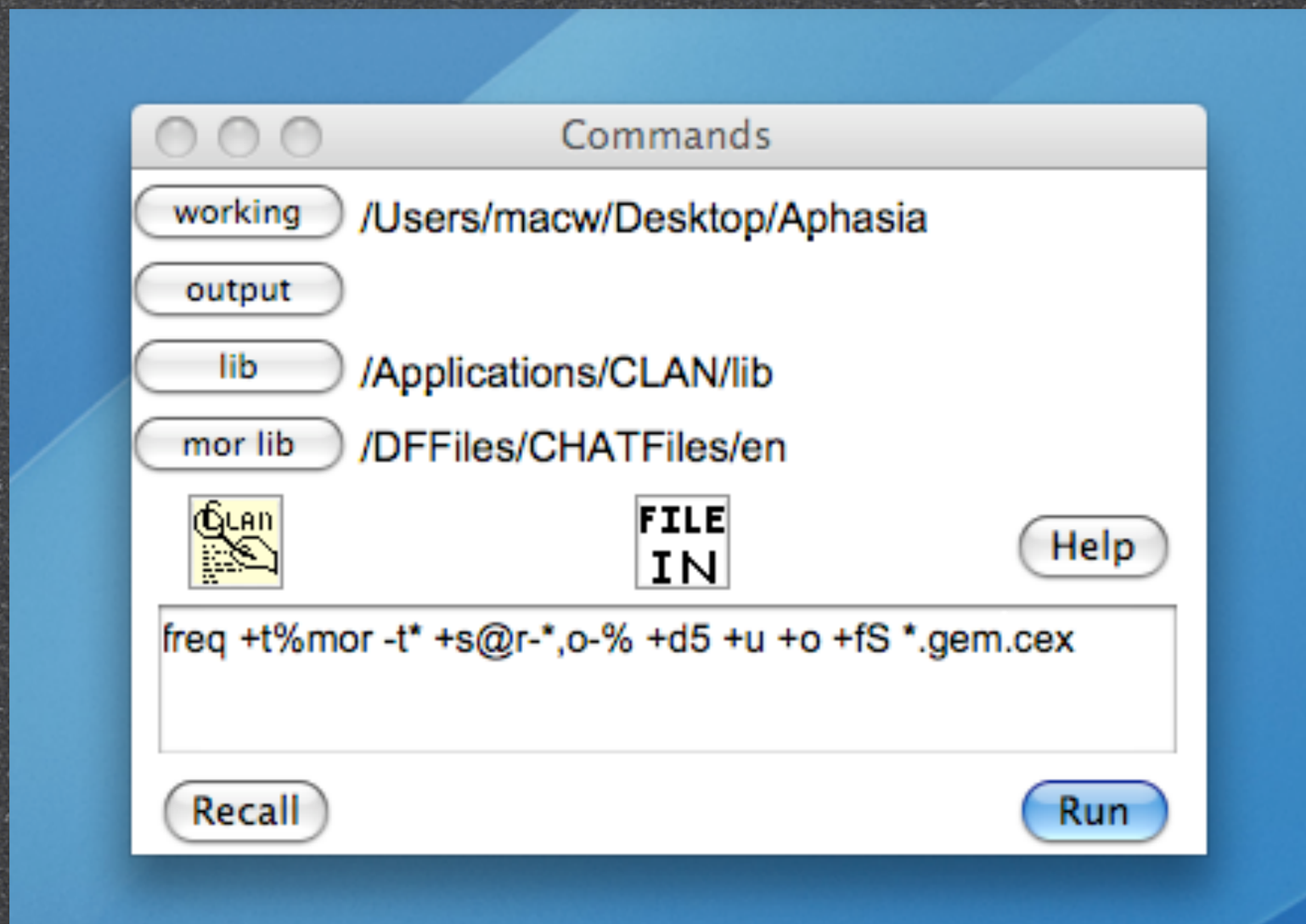
FREQ on Steroids

- A more complex application of the FREQ command asks for a frequency count of all words in the files containing the Cinderella story.



- In this case we asked FREQ to look at the MOR line, showing the parts of speech and stems of words, and to list the stems in descending order of frequency

FREQ (cont'd)



558 uh
551 and
499 the
313 um
200 she
167 I
155 was
142 a
142 to
119 they
107 it
94 he
88 Cinderella
86 her
78 so

Non-aphasic and Aphasic lexicons

Non-aphasic

n=25

total #of different word types used:
839

total # of tokens:
13302

TTR

Mean # of types	165.2
	R=68-329
Mean# of tokens	532.36
	R= 123-1347
Mean TTR	.35
	R= .24-.56

Aphasic:

n=24

total #of different word types used:
526 (62% of normal total)

total # of tokens:
5539 (41% of normal total)

TTR

Mean # of types	77.54
	R=21-155
Mean# of tokens	222.45
	R= 38-705
Mean TTR	.41
	R= .17-.72

“Cinderella” vs “Secerundid”: Word level and utterance level errors

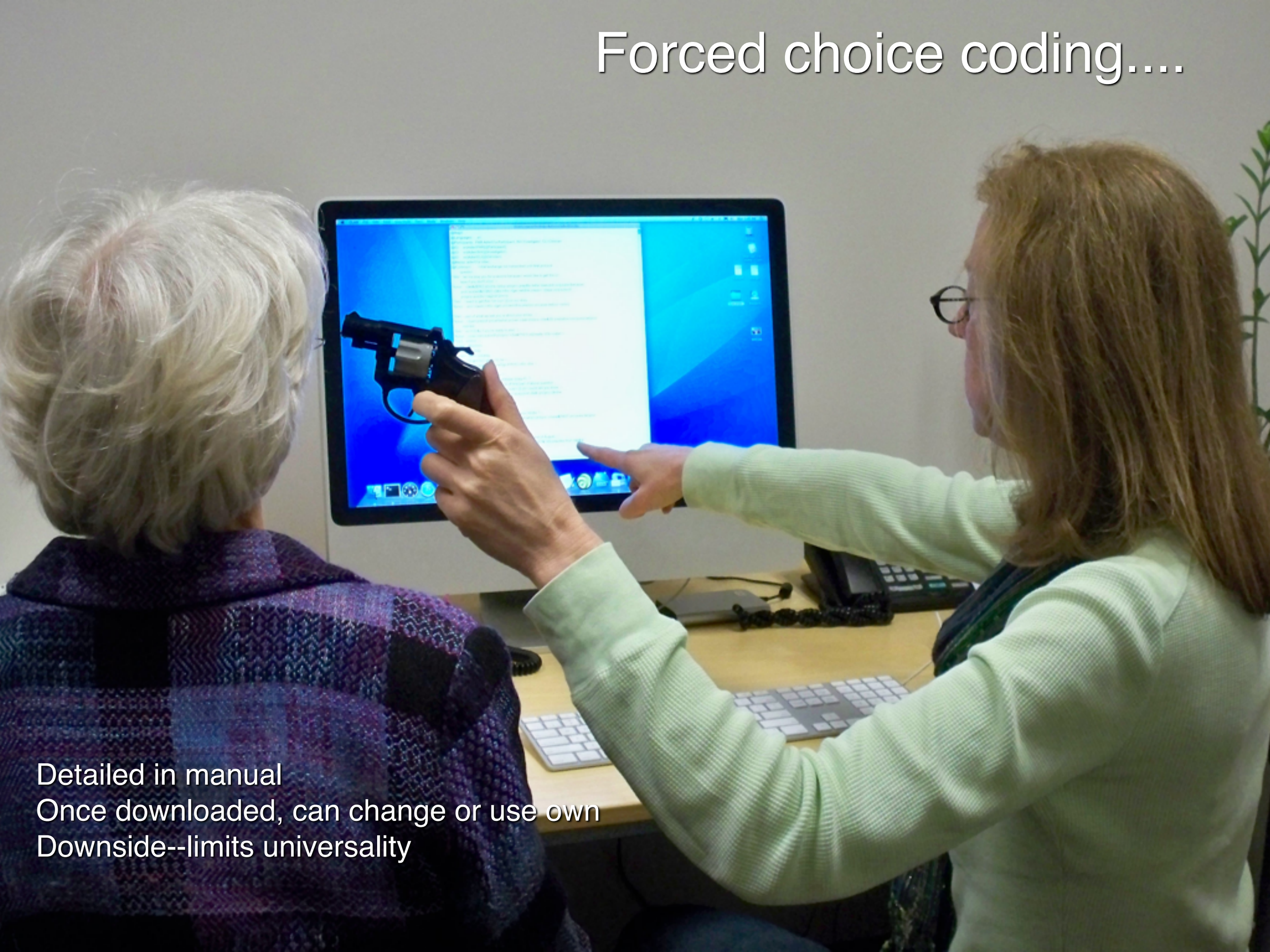
Word level coding:.

[* s] semantic paraphasia
[* pw] phonemic paraphasia, real word
[* pn] phonemic paraphasia, non-word
[* wk] word substitution, target known
[* wu] word substitution, target unknown
[* nk] non-word substitution, target known*
@n neologism (non-word substitution, target unknown)
[* agr] agreement error
[* met] metathesis
[* per] perseveration
[* pos] general part of speech
0 missing word or part of speech

Utterance level coding (post-codes)

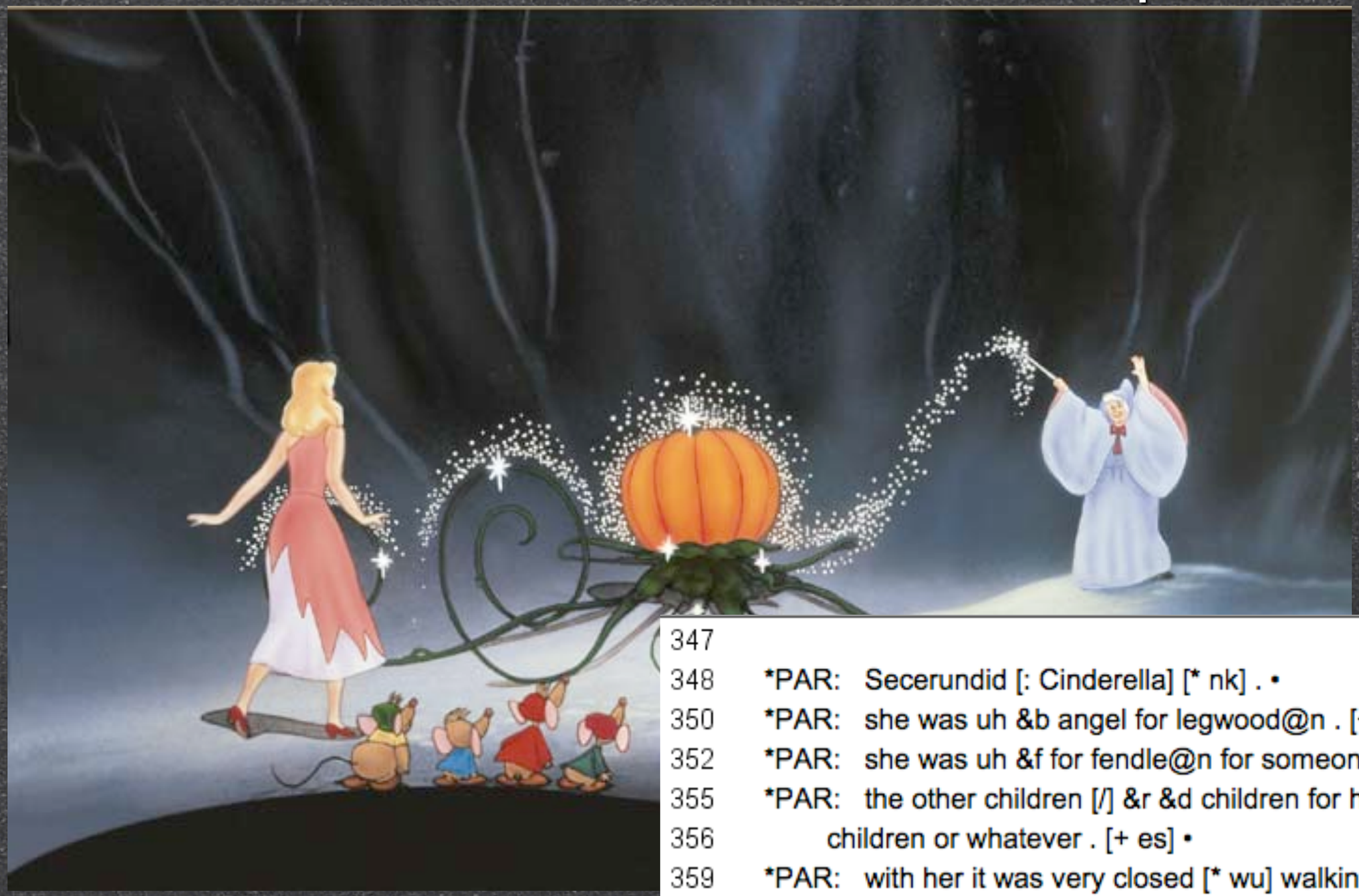
[+ gram] grammatical error
[+ jar] jargon
[+ es] empty speech
[+ per] perseveration
[+ cir] circumlocution

Forced choice coding....



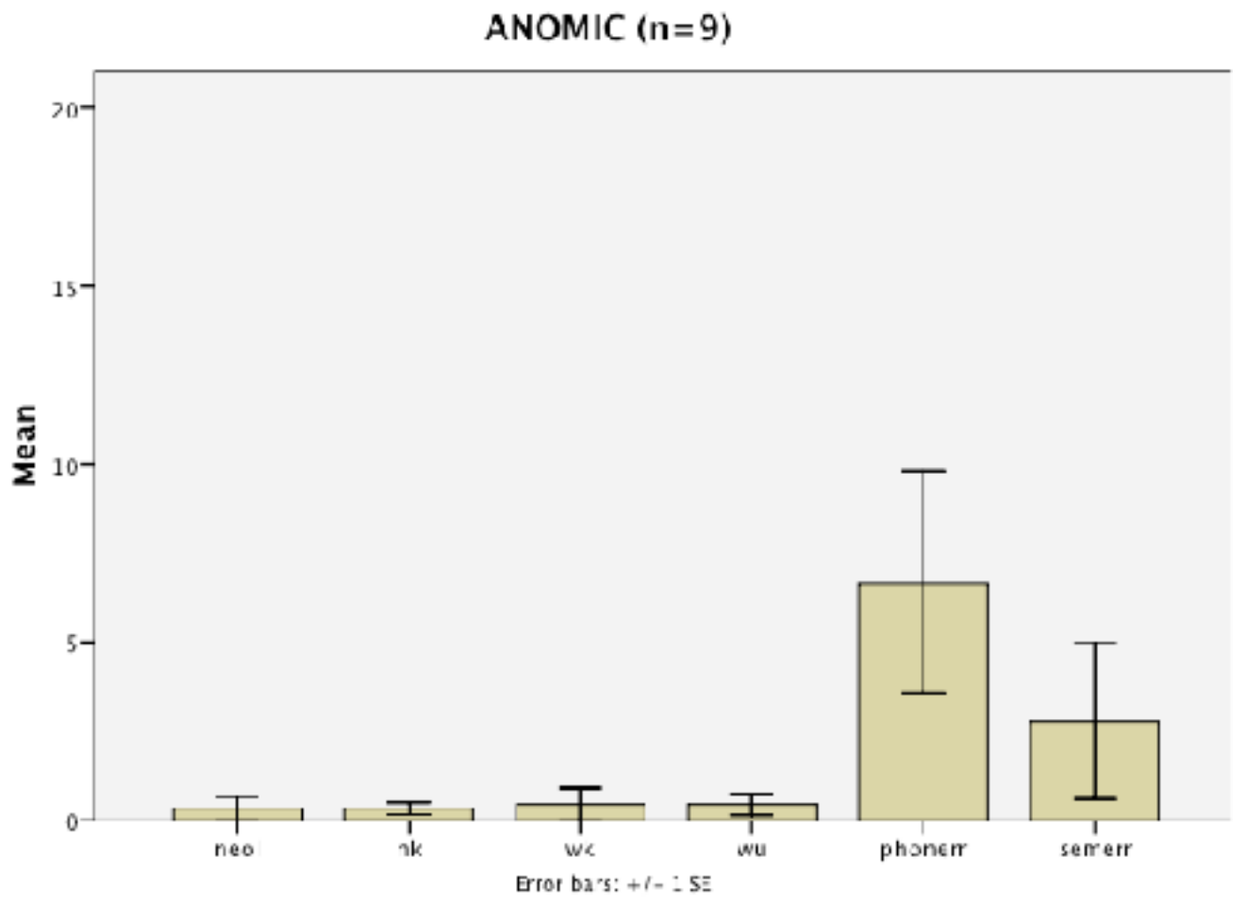
Detailed in manual
Once downloaded, can change or use own
Downside--limits universality

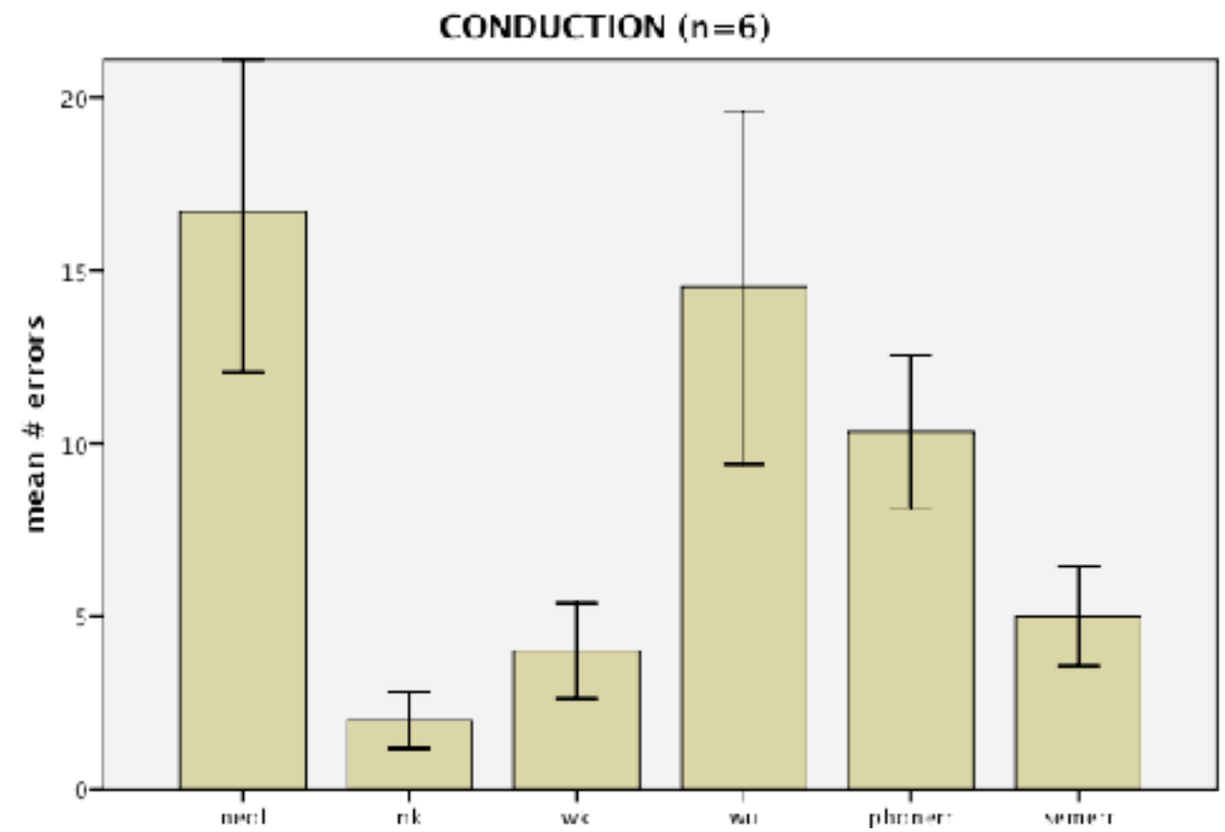
Error coding
Word and utterance level examples

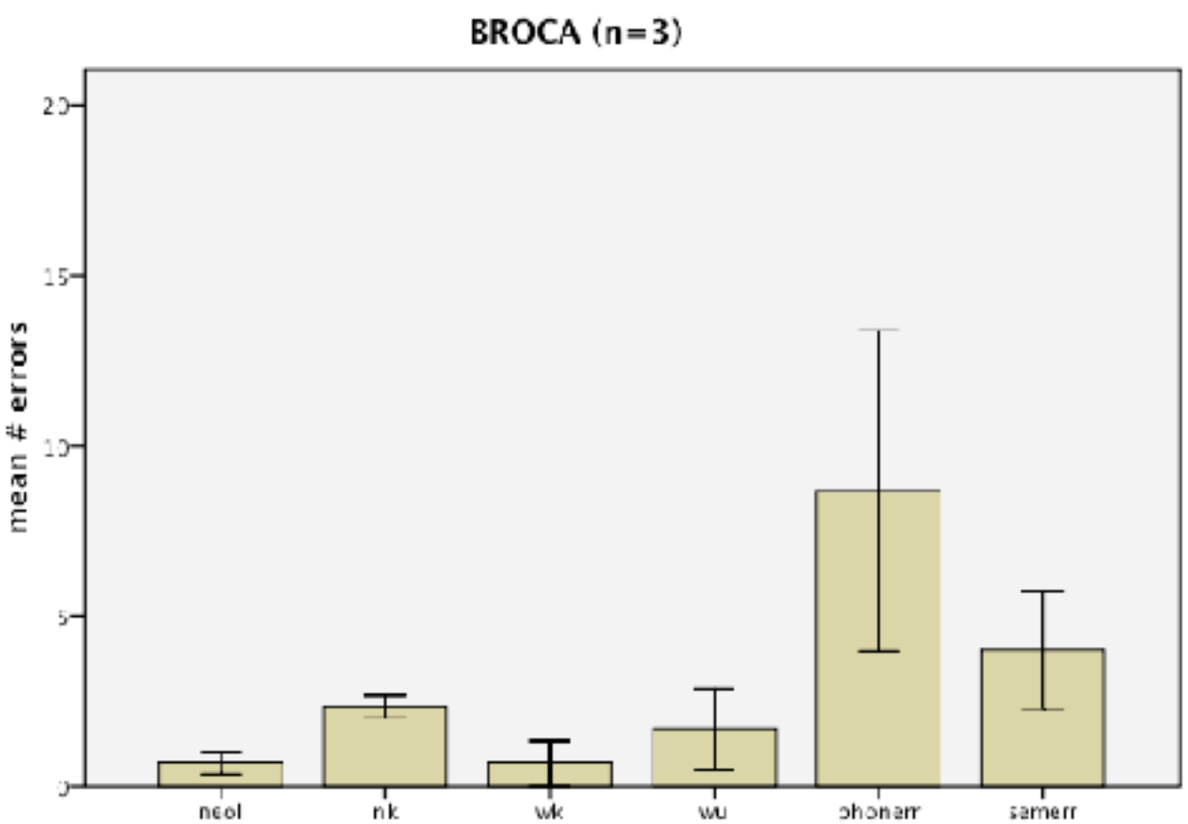


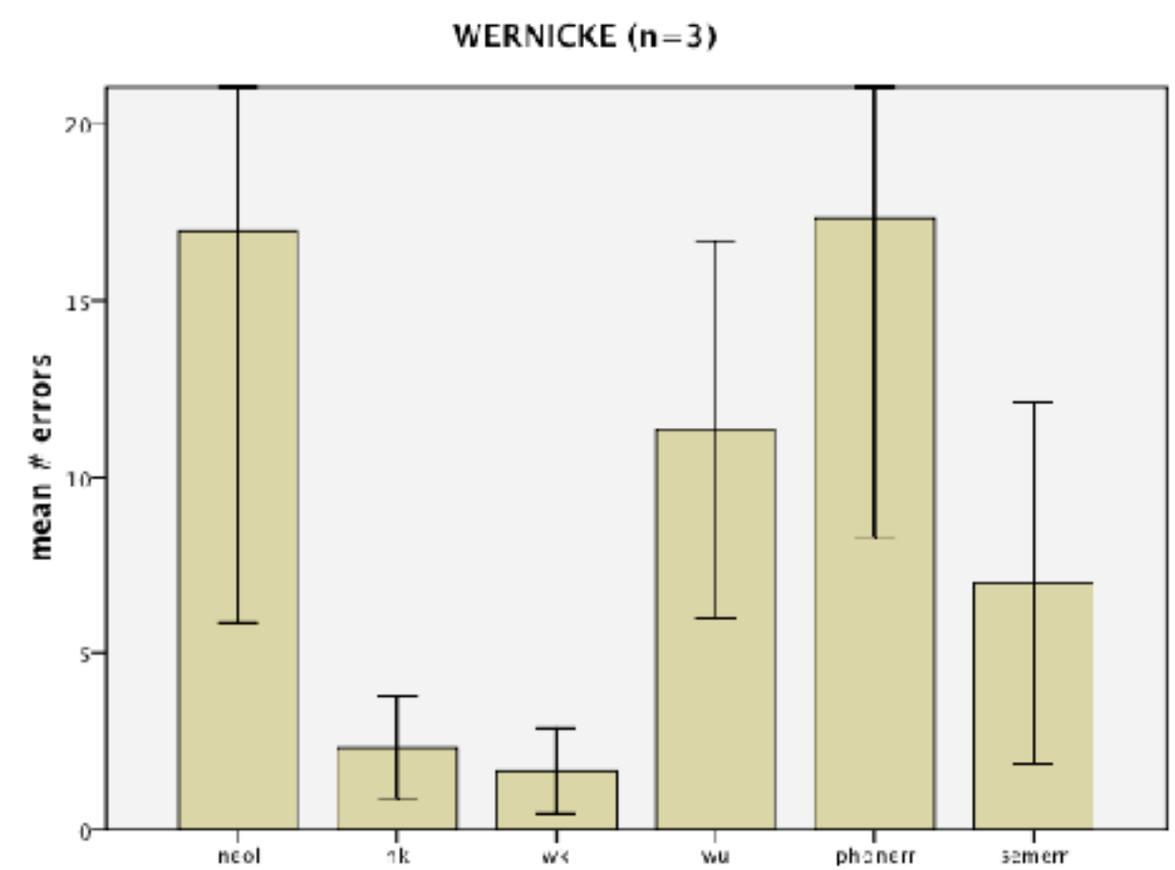
```
*PAR:   Secerundid [: Cinderella] [* nk] . •
*PAR:   she was uh &b angel for legwood@n . [+ jar] •
*PAR:   she was uh &f for fendle@n for someone else . [+ jar] •
*PAR:   the other children [/] &r &d children for her [/] are three
        children or whatever . [+ es] •
*PAR:   with her it was very closed [* wu] walking
        [* wu] in generalis@n . [+ jar] •
*PAR:   &th &th &p pezzels@n are going for the party . •
*PAR:   and she was &f fen@n people
        [* wu] for prezzled@n (.) for the present [* wu] . [+ jar] •
```

Word-level errors

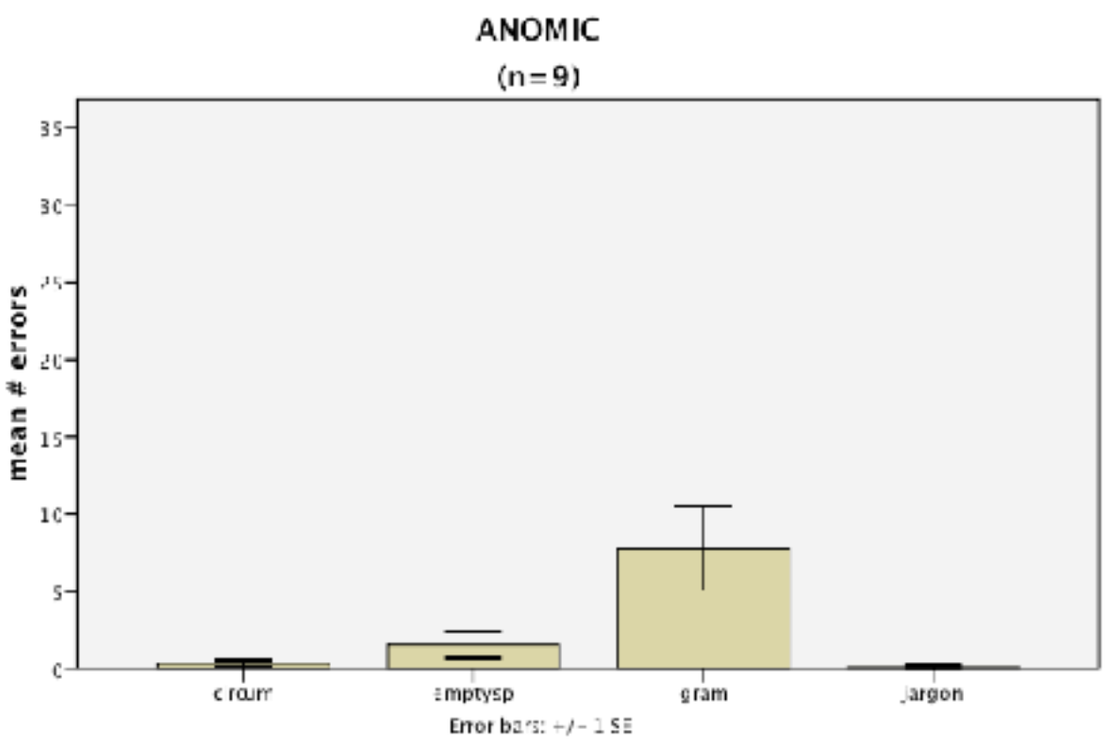


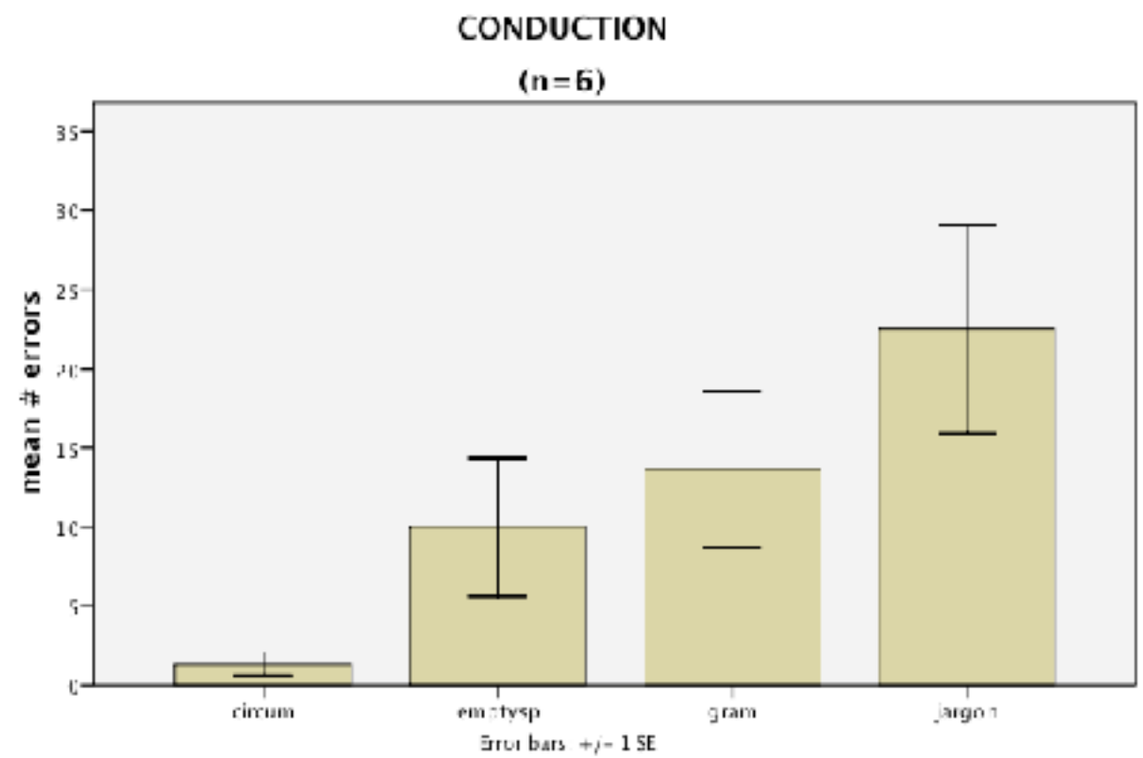


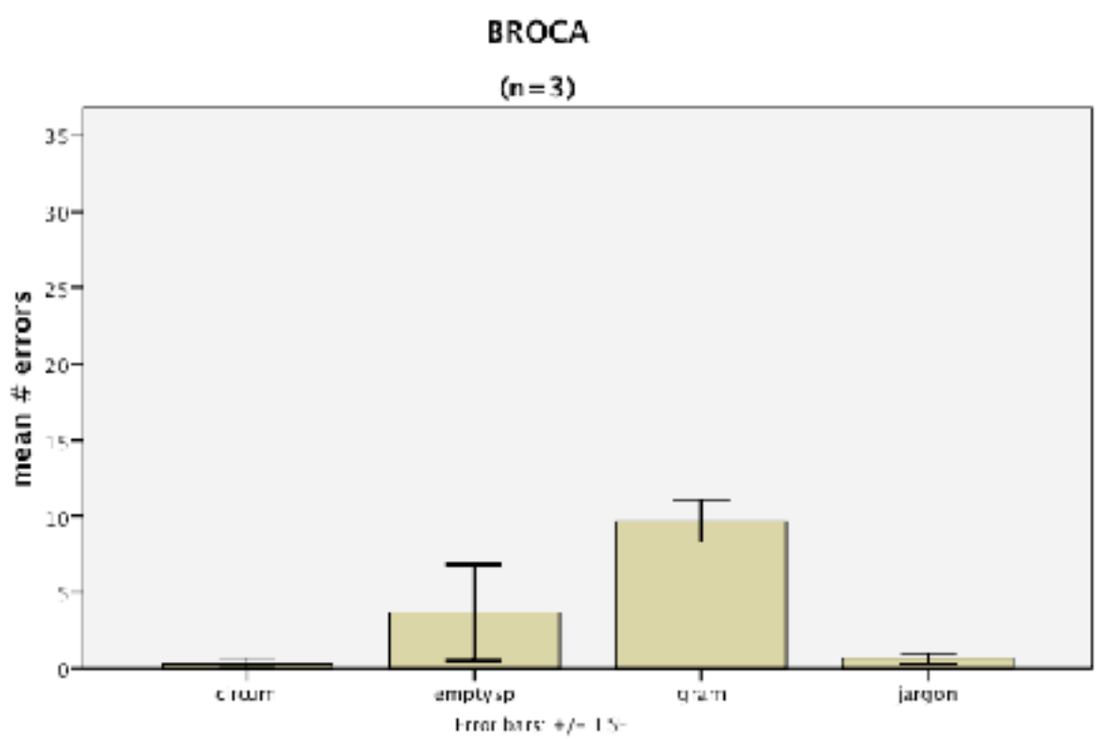


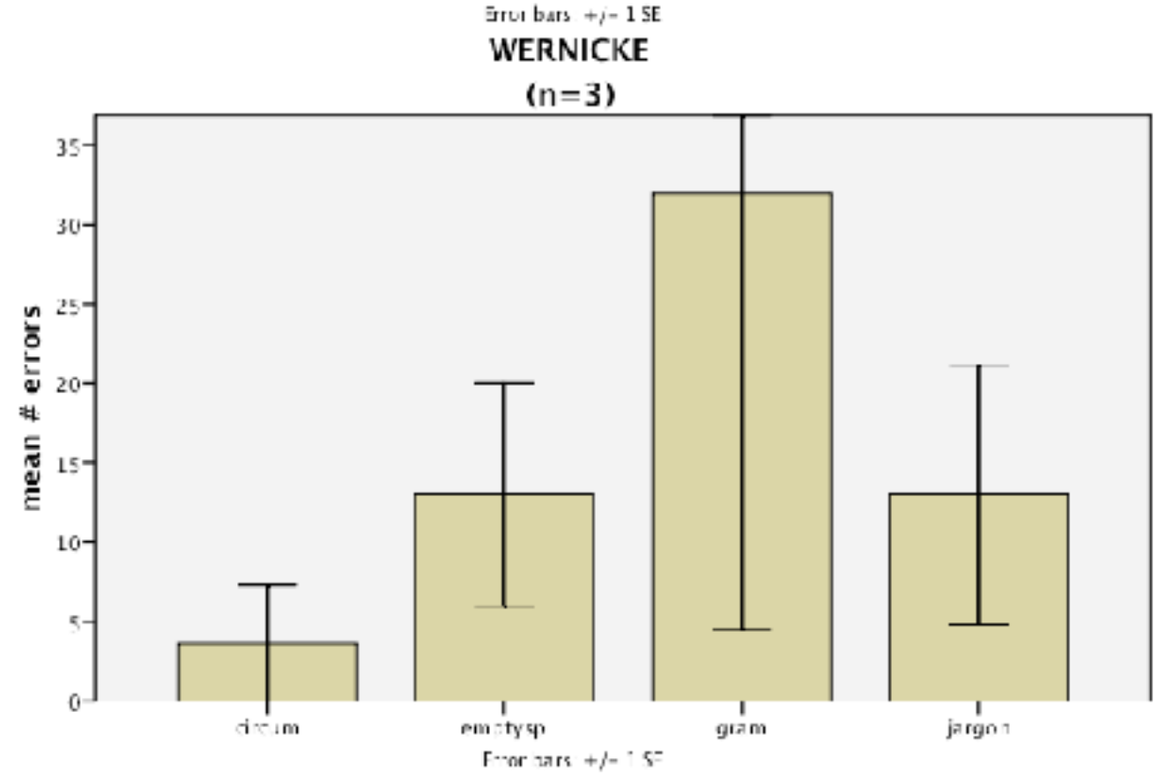


Utterance-level errors



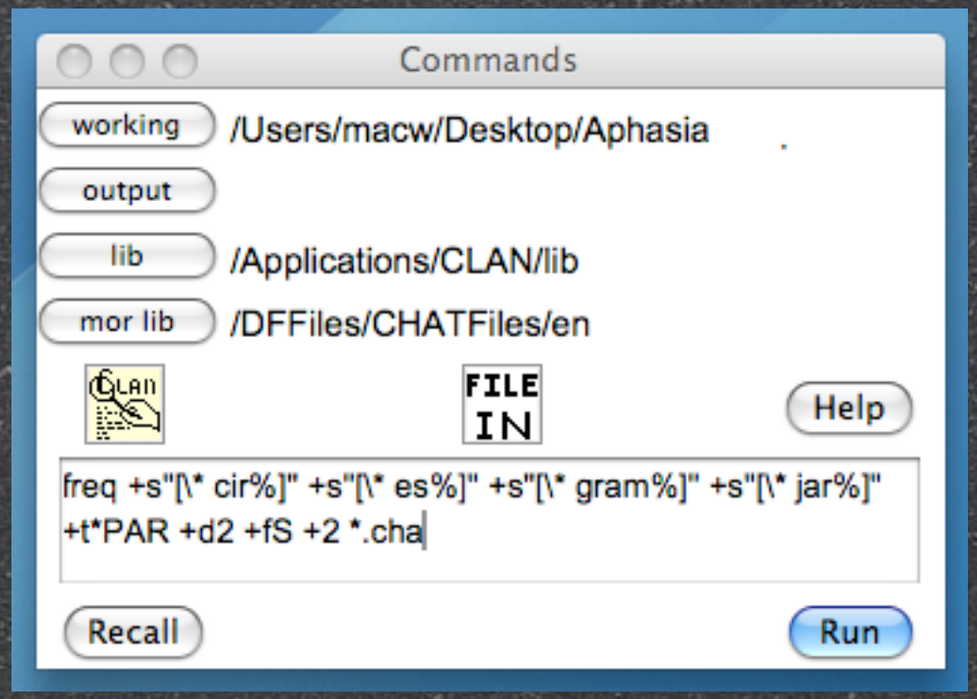




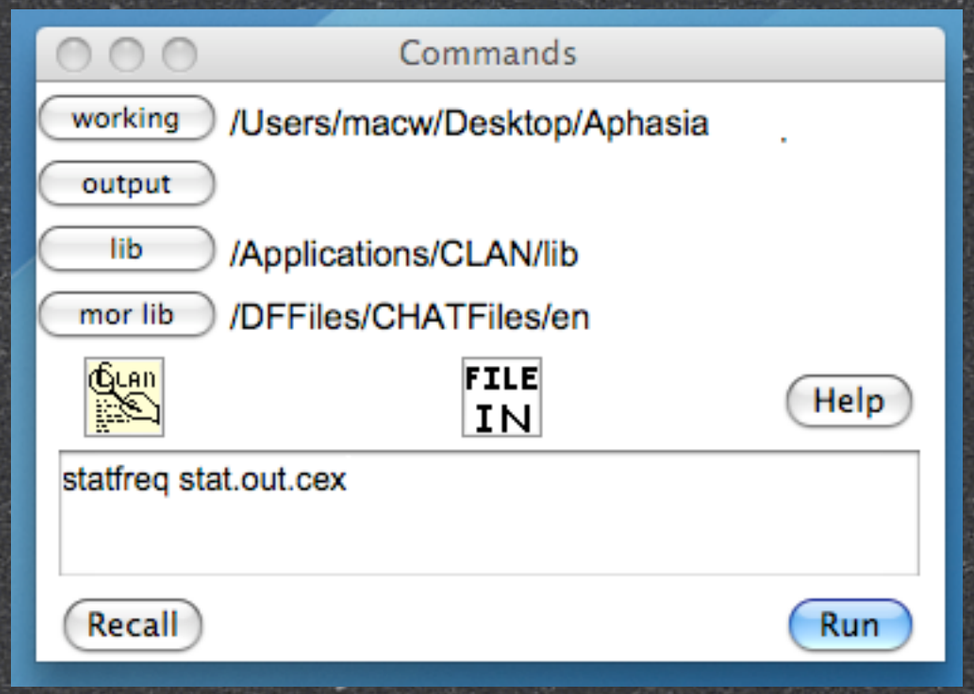


Going STATFREQing

1. Command below will result in the freq of sentence-level error codes across all CHAT files in data folder.



2. Then this command puts the data into a format that will transfer it to EXCEL

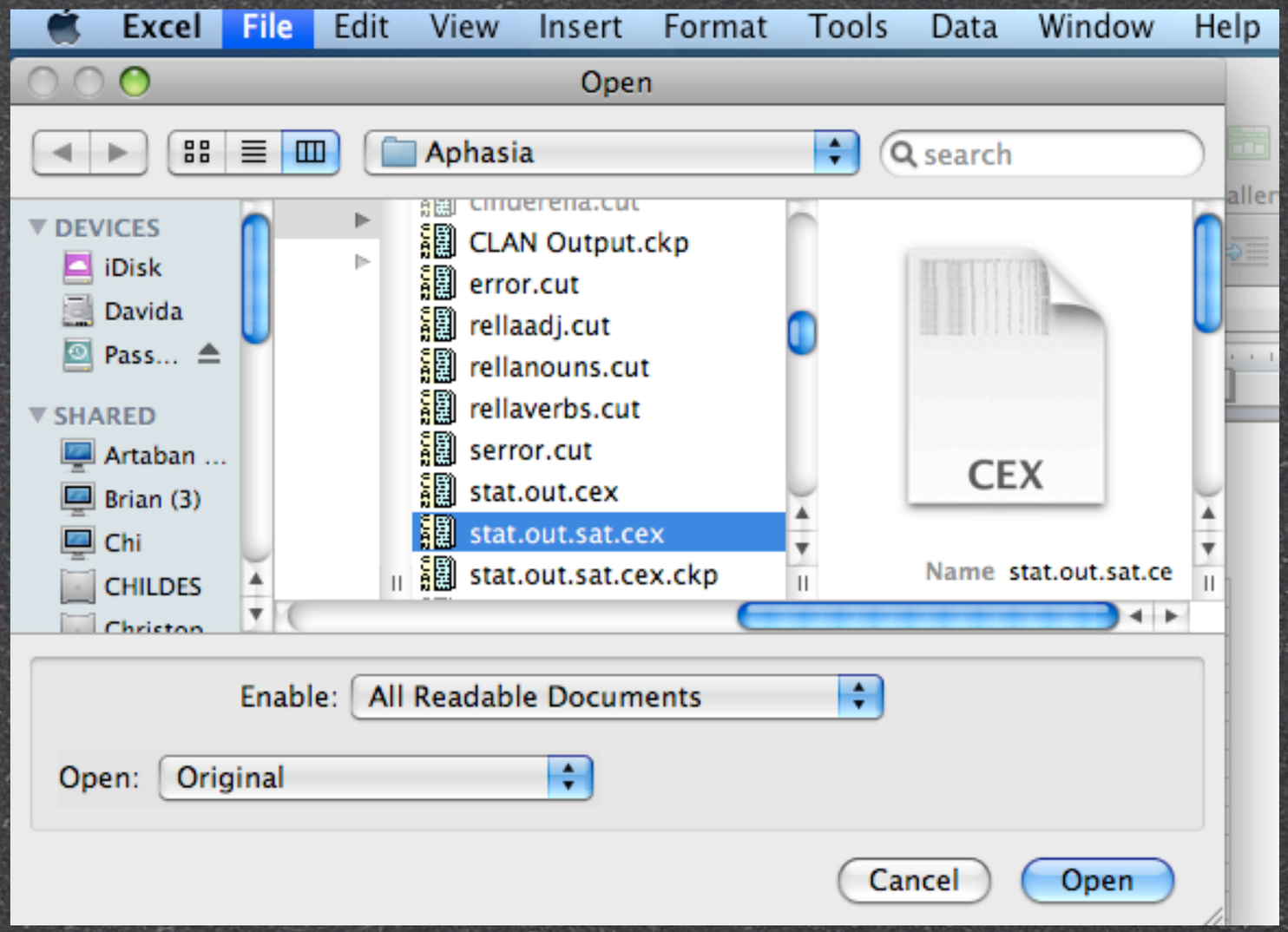


Here’s what you get

en|Adler|PAR|.|.|.|.|Participant|.|	219	0	0	8	1	2	9	0.222
en|Adler|PAR|.|.|.|.|Participant|.|	363	0	12	37	3	3	52	0.058
en|Adler|PAR|.|.|.|.|Participant|.|	160	0	0	0	0	0	0	-
en|Adler|PAR|.|.|.|.|Participant|.|	115	0	0	3	0	1	3	0.333
en|Adler|PAR|.|.|.|.|Participant|.|	184	1	2	11	38	4	52	0.077
en|Adler|PAR|.|.|.|.|Participant|.|	89	0	11	5	29	3	45	0.067
en|Adler|PAR|.|.|.|.|Participant|.|	443	0	0	2	0	1	2	0.500
en|Adler|PAR|.|.|.|.|Participant|.|	144	0	7	6	0	2	13	0.154
en|Tucson|PAR|.|.|.|.|Participant|.|	124	0	0	1	0	1	1	1.000
en|Tucson|PAR|.|.|.|.|Participant|.|	132	0	10	11	1	3	22	0.136
en|Tucson|PAR|.|.|.|.|Participant|.|	174	11	26	4	8	4	49	0.082
en|Tucson|PAR|.|.|.|.|Participant|.|	150	0	0	0	0	0	0	-
en|Tucson|PAR|.|.|.|.|Participant|.|	169	0	0	15	0	1	15	0.067
en|Tucson|PAR|.|.|.|.|Participant|.|	114	0	1	4	0	2	5	0.400
en|Tucson|PAR|.|.|.|.|Participant|.|	186	2	5	6	0	3	13	0.231
en|Tucson|PAR|.|.|.|.|Participant|.|	134	0	0	13	3	2	16	0.125
en|Tucson|PAR|.|.|.|.|Participant|.|	115	1	0	7	0	2	8	0.250
en|Tucson|PAR|.|.|.|.|Participant|.|	103	0	0	26	0	1	26	0.038
en|Tucson|PAR|.|.|.|.|Participant|.|	126	1	9	2	24	4	36	0.111
en|Tucson|PAR|.|.|.|.|Participant|.|	152	5	7	11	28	4	51	0.078
en|Tucson|PAR|.|.|.|.|Participant|.|	313	0	2	87	2	3	91	0.033
en|Tucson|PAR|.|.|.|.|Participant|.|	126	0	1	11	1	3	13	0.231
en|Tucson|PAR|.|.|.|.|Participant|.|	215	1	30	8	39	4	78	0.051
en|Tucson|PAR|.|.|.|.|Participant|.|	78	1	1	3	0	3	5	0.600

*Looks funny,but transfers cleanly

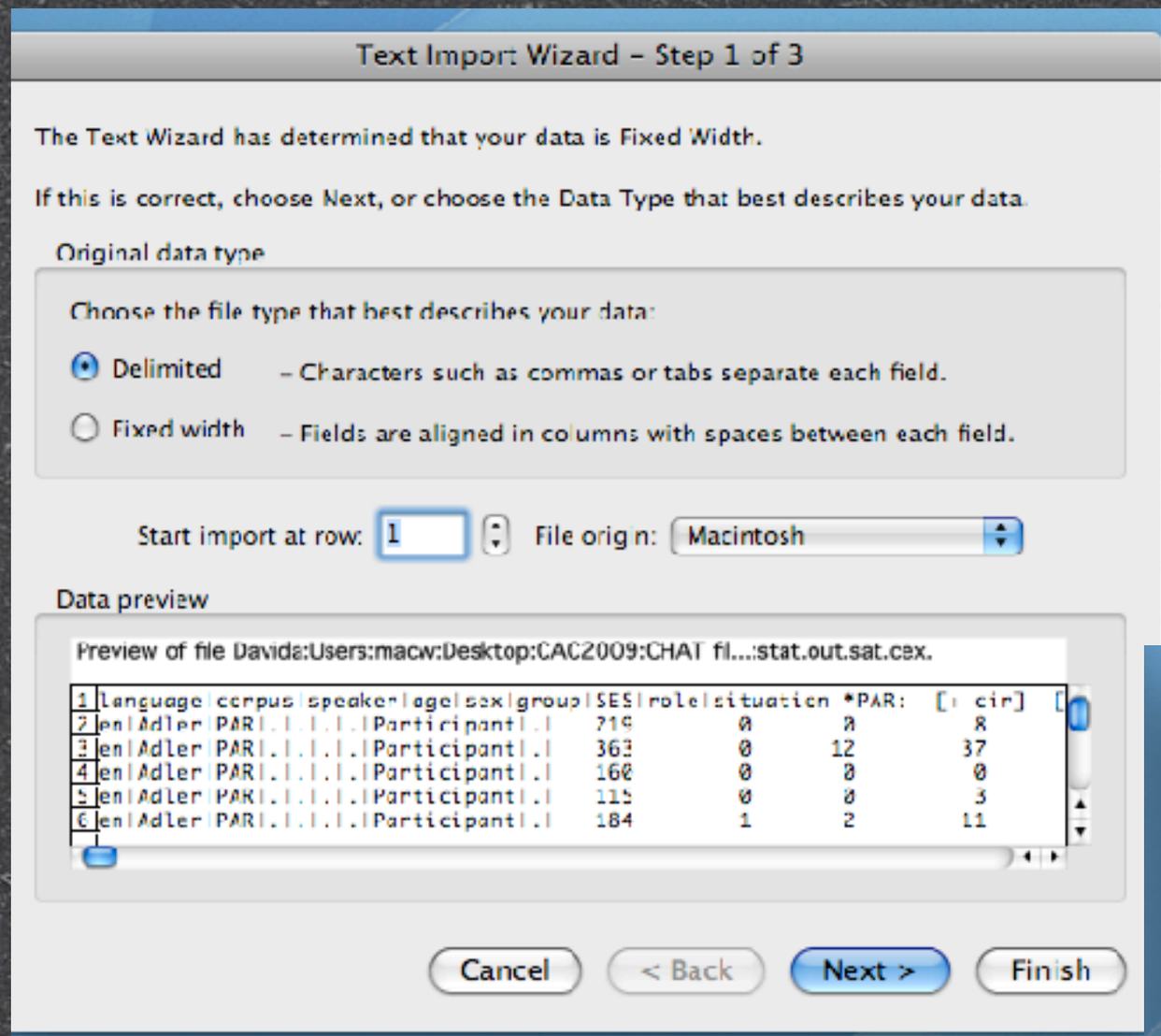
THEN.....

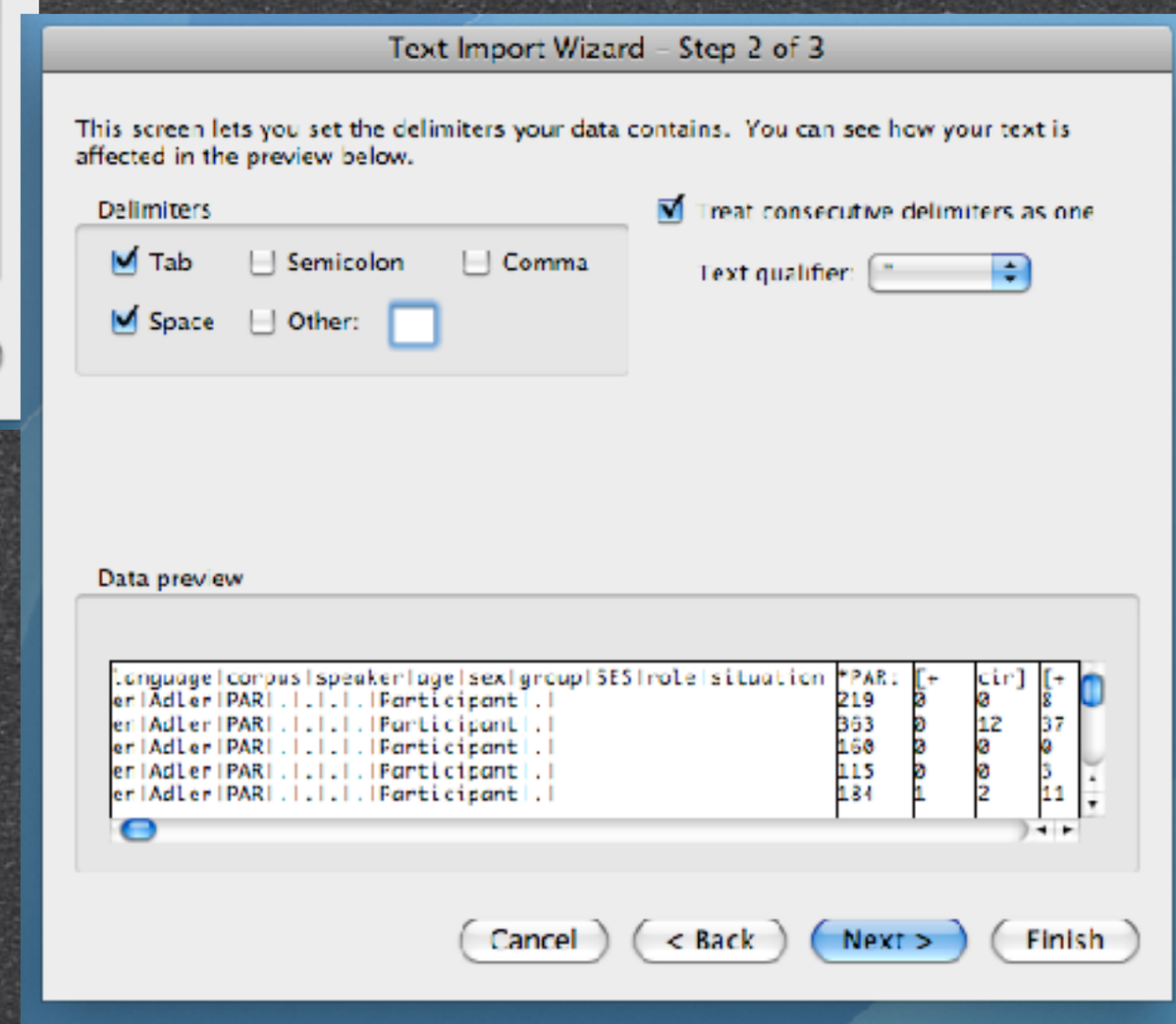


1. Open up Excel (see menu)

2. Go to File, then Open the highlighted one

2 more steps and then.....





And look what you have!!!!!

	A	B	C	D	E	F	G	H	I	J
1	partic	wabtype	*PAR:	cir	es	gram	jar	types	token	ttr
2	adler01a	Anomic	219	0	0	8	1	2	9	
3	adler02a	Conduction	363	0	12	37	3	3	52	
4	adler03a	Anomic	160	0	0	0	0	0	0	-
5	adler04a	Trans Motor	115	0	0	3	0	1	3	
6	adler05a	Conduction	184	1	2	11	38	4	52	
7	adler06a	Broca	89	0	11	5	29	3	45	
8	adler07a	Anomic	443	0	0	2	0	1	2	
9	adler08a	Anomic	144	0	7	6	0	2	13	
10	tucson01a	not aphasic	124	0	0	1	0	1	1	
11	tucson02a	Broca	132	0	10	11	1	3	22	
12	tucson03a	Wernicke	174	11	26	4	8	4	49	
13	tucson04a	not aphasic	150	0	0	0	0	0	0	-
14	tucson05a	Anomic	169	0	0	15	0	1	15	
15	tucson06a	Anomic	114	0	1	4	0	2	5	
16	tucson07a	Anomic	186	2	5	6	0	3	13	
17	tucson08a	Conduction	134	0	0	13	3	2	16	
18	tucson09a	Broca	115	1	0	7	0	2	8	
19	tucson10a	Anomic	103	0	0	26	0	1	26	
20	tucson11a	Conduction	126	1	9	2	24	4	36	
21	tucson12a	Conduction	152	5	7	11	28	4	51	
22	tucson13a	Wernicke	313	0	2	87	2	3	91	
23	tucson14a	Broca	126	0	1	11	1	3	13	
24	tucson15a	Conduction	215	1	30	8	39	4	78	
25	tucson16a	Anomic	78	1	1	3	0	3	5	
26										

To finish up this romp through CLAN...

We essentially used the aphasic corpus and the error analyses, then STATFREQed the data to answer the following questions:

Is there a significant correlation of WAB scores & total number of words in Cinderella stories?

NO $r = .153$

Is there a significant correlation between WAB scores & word-level errors in Cinderella stories?

YES $r = -.415 < .05$

Is there a significant correlation between WAB scores & utterance-level errors in Cinderella stories?

YES $r = -.528 < .01$

Research thoughts.....

REALLY develop a “normal speaker” template across
a protocol discourse task (or tasks) .
Compare aphasic performances against
the lexicon(s).
Use this to develop a objective metric for severity,
or even a classification system.

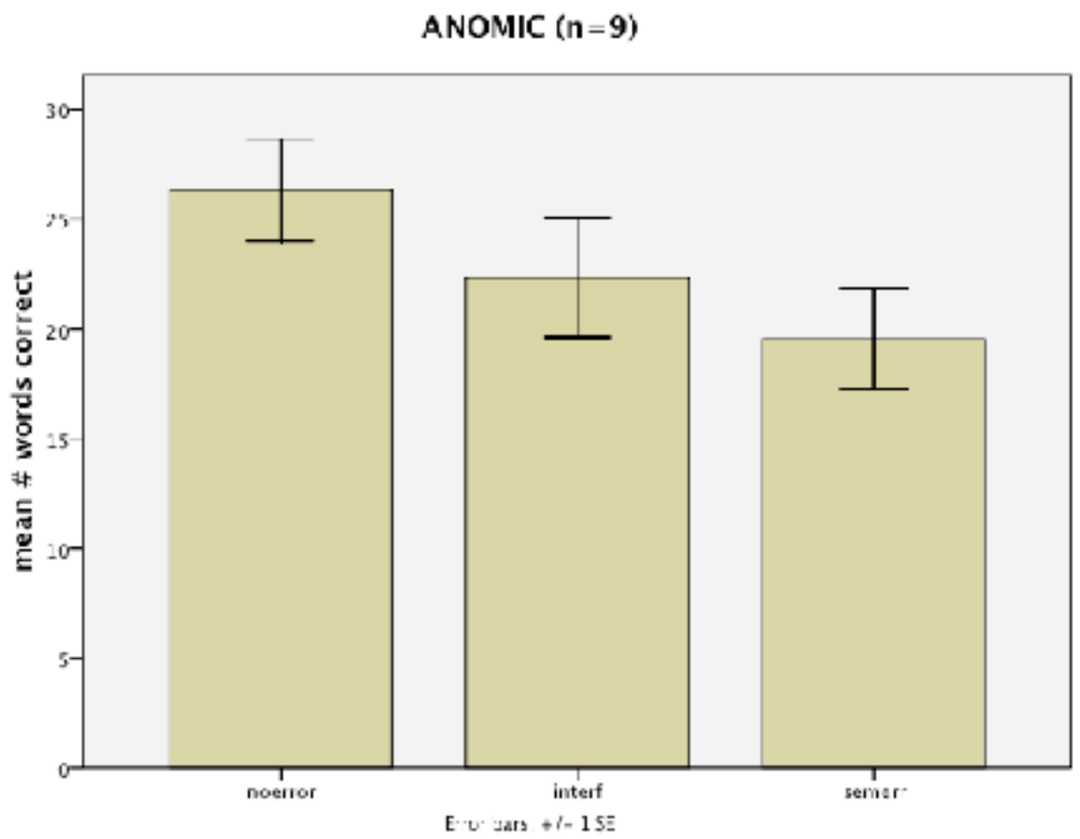
Catalog and compare answers to
one (or more) free speech questions

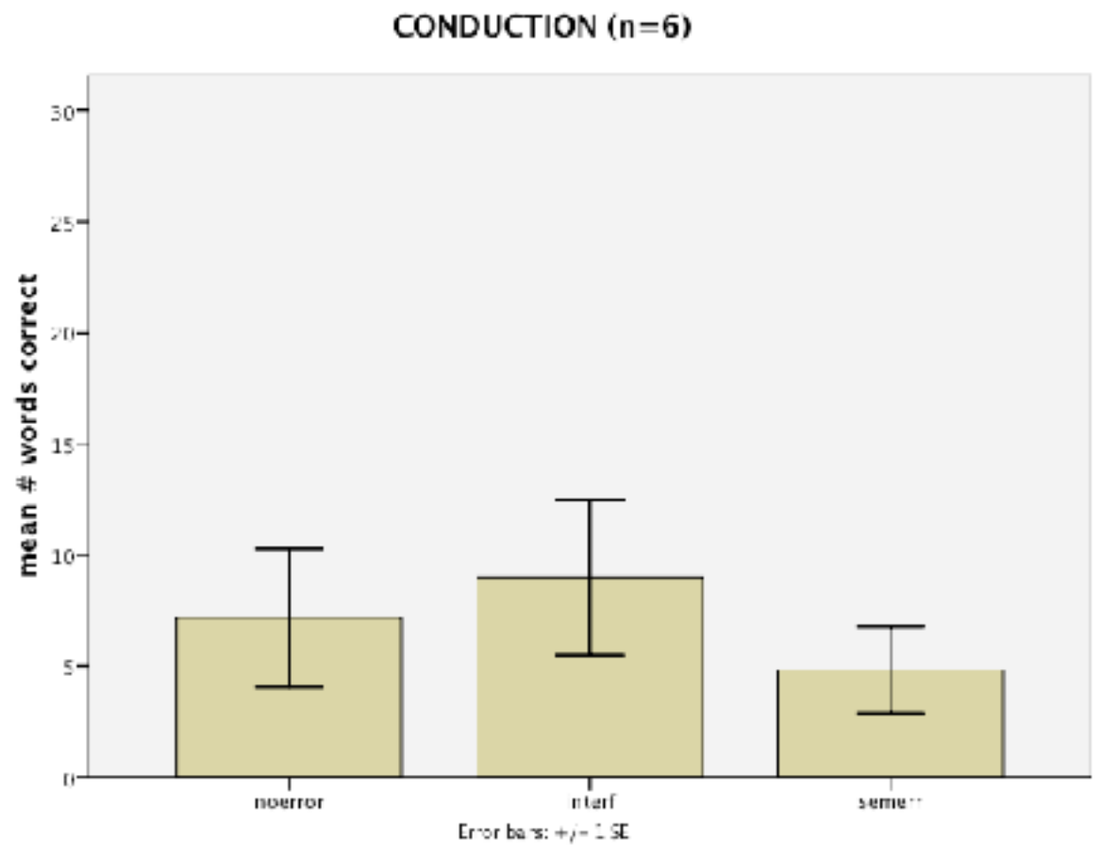
Develop data-based definitions for categorizing
aphasic speech errors

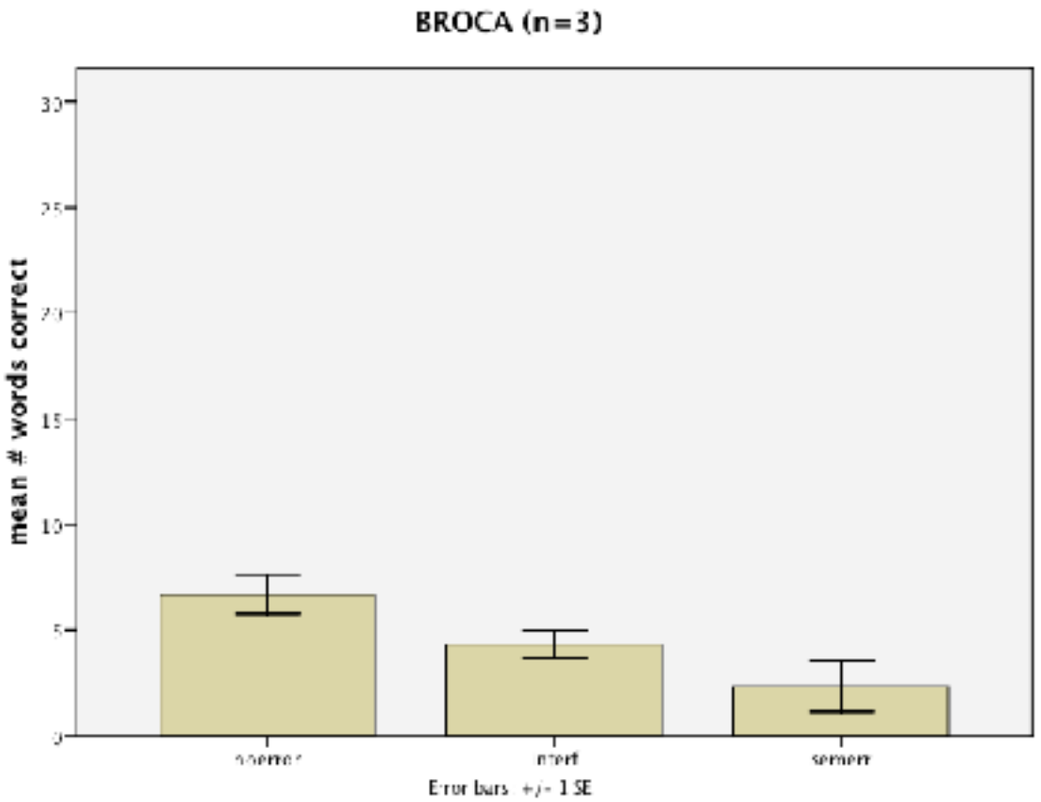
AND many, many more.....

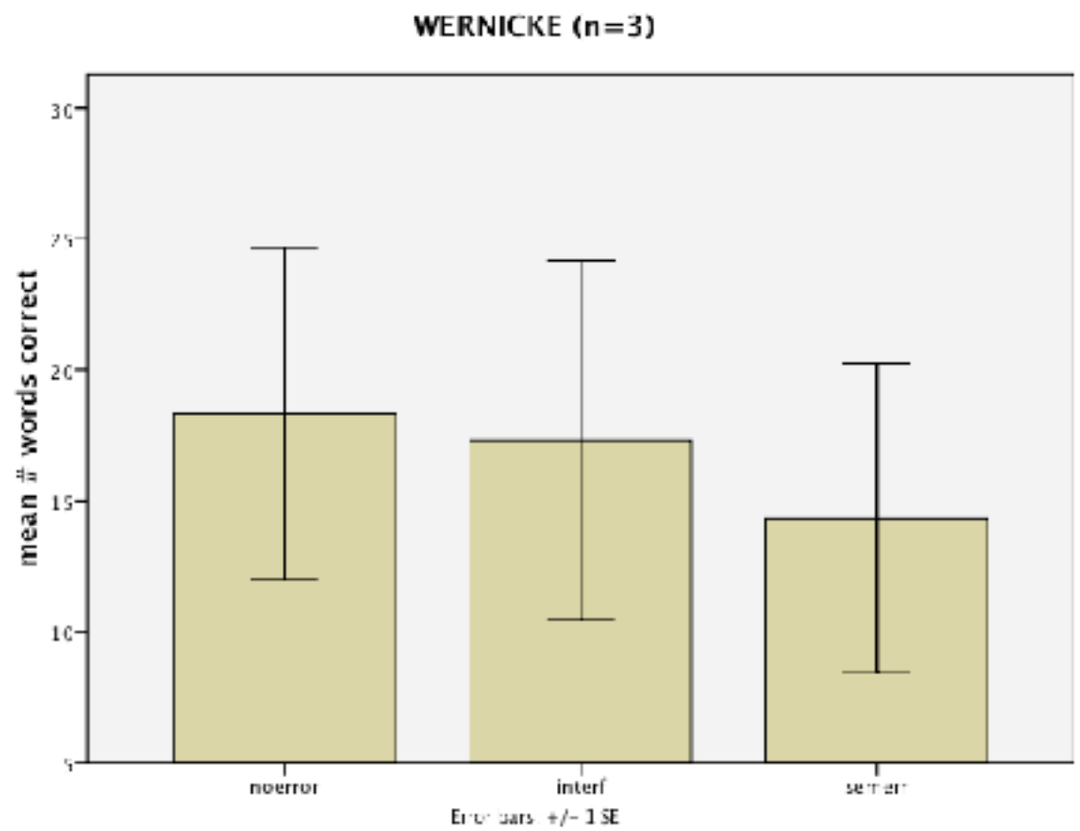

And we all lived happily ever after!
Thanks for listening!

*PAR: I have uh a &=ges:chest feeding tube in me and little &=ges:arm &ji bottles you know . •336700_341554•*PA
&=ges:knock knockin(g) on the +...*PAR: and the guy said +"/.*PAR: +" &=ges:easy hey Mister Vaughn take it easy .*P
days I was fine .*PAR: all of a sudden boom@o .*PAR: uh some [//] &=points:leg all [/] all the way down
here &=points:chest here all the way down to &=points:neck .*PAR: and &=ges:spread I (.) cover the one slice .*PAR
I put the jelly on it.*PAR: but um other than that &=shrugs .*PAR: &=ges:umbrella <he had a> [//] <he had it> [//]
&a on him . •487719_493839•*PAR: and &=points:picture that's the umbrella on him .*PAR: uh and then you &=hands:clap put it t
peanut butter sandwich .*PAR: it's &=fingers:count one two three four five uh uh +...*INV: five weeks ?*PA
&=ges:mowing the device that um I was mowing um um leaves um kept me upright .*PAR: and um um with a knife &=g
ugh !*PAR: um &=ges:spread spread some peanut butter on it .*PAR: I don't know &=head:shake .*PAR: um (a)nother pieces um you
&=laughs I don't know .


ANOMIC (n=9)
mean # words correct
30
25
20
15
10
5
0
noerror
interf
semerr
Error bars: +/- 1 SE


CONDUCTION (n=6)
mean # words correct
30
25
20
15
10
5
0
noerror
interf
semerr
Error bars: +/- 1 SE


BROCA (n=3)
mean # words correct
30
25
20
15
10
5
0
noerror
interf
semerr
Error bars: +/- 1 SE


WERNICKE (n=3)
mean # words correct
30
25
20
15
10
5
noerror
interf
semerr
Error bars: +/- 1 SE

Cinderella Lexicon of nouns and verbs for Non aphasic speakers

Nouns

Cinderella
ball
prince
slipper
mother, stepmother
dress
daughter, stepdaughter
fairy
godmother
sister, stepsister
glass
home
girl
time
house
pumpkin
midnight
mouse
carriage
foot
father
shoe
coach
lady
animal
horse
clock
kingdom
chore
king
love
story
wife
castle
invitation
person
servant
day
palace

wand
Prince Charming
clothes
course
cat
land
magic
party
stair
thing
friend
life
man
dance
door
end
footman
princess
gown
hair
maid
night
room
dog
family
piece
scene
son
step
stroke
word
ballroom
child, stepchild
meantime
messenger
o’clock

Noun
N=80

Verbs

be
go
have
get
come
do
say
try
marry, remarry
know
make
work
fit
find
see
take
dance
leave
run
lose
live
look
turn
think
appear, disappear, reappear
strike
send
tell
wear
excite
put
realize
make
let
like
find

invite
become
help
meet
remember
clean
fall
need
treat
cry
see
bring
give
start
must
decide
fall
pass
talk
want
ask
belong
hear
keep
push
sit
tear
happen
end
happen
mean
strike

Verb
N=71