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
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

gem +sCinderella +t*PAR +n +d1 +f*cha
VOILA!!!
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
A few other goodies.

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

*PAR: and the two ugly girls they had big feet.
*PAR: they didn't fit.
*PAR: and so she [I] & he tried it on Cinderella.
*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 [] & he tried it on Cinderella.

%mor: conj:coo|and co|so pro|he v|try-PAST pro|it prep|on
  n:prop|Cinderella.
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.

A more complex application of the FREQ command asks for a frequency count of all words in the files containing the Cinderella story.
FREQ (cont’d)
<table>
<thead>
<tr>
<th>Non-aphasic</th>
<th>Aphasic:</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=25</td>
<td>n=24</td>
</tr>
<tr>
<td>total # of different word types used: 839</td>
<td>total # of different word types used: 526 (62% of normal total)</td>
</tr>
<tr>
<td>total # of tokens: 13302</td>
<td>total # of tokens: 5539 (41% of normal total)</td>
</tr>
</tbody>
</table>

**TTR**
- **Mean # of types**
  - Non-aphasic: 165.2, R=68-329
  - Aphasic: 77.54, R=21-155
- **Mean # of tokens**
  - Non-aphasic: 532.36, R=123-1347
  - Aphasic: 222.45, R=38-705
- **Mean TTR**
  - Non-aphasic: .35, R=.24-.56
  - Aphasic: .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

ANOMIC (n=9)

CONDUCTION (n=6)

BROCA (n=3)

WERNICKE (n=3)
Utterance-level errors

ANOMIC (n=9)

BROCA (n=3)

CONDUCTION (n=6)

WERNICKE (n=3)
1. Command below will result in the freq of sentence-level error codes across all CHAT files in data folder.

```
freq +s"[" cir%"]" +s"[" es%"]" +s"[" gram%"]" +s"[" jar%"]" +t"PAR +d2 +fS +2 *.cha"
```

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

```
statfreq stat.out.cex
```
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 |52 |0.058 |
| en|Adler|PAR|.|.|.||Participant|.| | | |160 |0 |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 |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 |

*L*ooks funny, but transfers cleanly
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!!!!!
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!
I have a chest feeding tube in me and little bottles you know. And the guy said, "Hey Mister Vaughn take it easy." Two days I was fine. All of a sudden boom. All the way down here all the way down to my neck. I put the jelly on it. But um other than that. And picture that's the umbrella on him. "Uh and then you put it together and you made a peanut butter sandwich." It's fingers count one two three four five uh... Five weeks? And the device that um I was mowing um leaves um kept me upright. And um with a knife. Ugh! Another piece um you... I don't know. I don't know.
Cinderella Lexicon of nouns and verbs for Non aphasic speakers

<table>
<thead>
<tr>
<th>Nouns</th>
<th>Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cinderella</td>
<td>invite</td>
</tr>
<tr>
<td>ball</td>
<td>become</td>
</tr>
<tr>
<td>prince</td>
<td>help</td>
</tr>
<tr>
<td>slipper</td>
<td>meet</td>
</tr>
<tr>
<td>mother, stepmother</td>
<td>remember</td>
</tr>
<tr>
<td>dress</td>
<td>clean</td>
</tr>
<tr>
<td>daughter, stepdaughter</td>
<td>fall</td>
</tr>
<tr>
<td>fairy</td>
<td>need</td>
</tr>
<tr>
<td>godmother</td>
<td>treat</td>
</tr>
<tr>
<td>sister, stepsister</td>
<td>cry</td>
</tr>
<tr>
<td>glass</td>
<td>see</td>
</tr>
<tr>
<td>home</td>
<td>bring</td>
</tr>
<tr>
<td>girl</td>
<td>give</td>
</tr>
<tr>
<td>time</td>
<td>start</td>
</tr>
<tr>
<td>house</td>
<td>must</td>
</tr>
<tr>
<td>pumpkin</td>
<td>decide</td>
</tr>
<tr>
<td>midnight</td>
<td>fall</td>
</tr>
<tr>
<td>mouse</td>
<td>pass</td>
</tr>
<tr>
<td>carriage</td>
<td>talk</td>
</tr>
<tr>
<td>foot</td>
<td>want</td>
</tr>
<tr>
<td>father</td>
<td>ask</td>
</tr>
<tr>
<td>shoe</td>
<td>belong</td>
</tr>
<tr>
<td>coach</td>
<td>hear</td>
</tr>
<tr>
<td>lady</td>
<td>keep</td>
</tr>
<tr>
<td>animal</td>
<td>sit</td>
</tr>
<tr>
<td>horse</td>
<td>push</td>
</tr>
<tr>
<td>piece</td>
<td>sit</td>
</tr>
<tr>
<td>clock</td>
<td>happens</td>
</tr>
<tr>
<td>kingdom</td>
<td>tear</td>
</tr>
<tr>
<td>chore</td>
<td>end</td>
</tr>
<tr>
<td>king</td>
<td>happen</td>
</tr>
<tr>
<td>love</td>
<td>mean</td>
</tr>
<tr>
<td>story</td>
<td>strike</td>
</tr>
<tr>
<td>word</td>
<td>Verb</td>
</tr>
<tr>
<td>ballroom</td>
<td>Verb</td>
</tr>
<tr>
<td>word</td>
<td>Verb</td>
</tr>
<tr>
<td>realtivity</td>
<td>Verb</td>
</tr>
<tr>
<td>find</td>
<td>Verb</td>
</tr>
<tr>
<td>palace</td>
<td>Verb</td>
</tr>
</tbody>
</table>

Noun N=80

Verb N=71