

# AphasiaBank:

Its role in research and clinical activities

AUDREY HOLLAND  
University of Arizona

SPENCER SUGARMAN  
Carnegie Mellon University

MARGARET FORBES  
Carnegie Mellon University

DAVIDA FROMM  
Carnegie Mellon University

BRIAN MACWHINNEY  
Carnegie Mellon University

HEATHER WRIGHT  
Arizona State University

## WHAT IS APHASIABANK?

AphasiaBank is the largest multimedia database in the world of discourse protocols gathered from persons with aphasia and controls. These speech and language samples are linked and synchronized with reliably coded transcripts.

### APHASIABANK GOALS

AphasiaBank relies on advances in data sharing, transcription, analysis, and web delivery available through the CHILDES and TalkBank Projects.

The goal is to provide researchers with a large corpus of systematically collected and analyzed data that can:

- further knowledge of language in aphasia through using it in research and for teaching
- be of use in improving treatments for aphasia

### THE PROTOCOL

#### Interview Sample

- Stroke Story
- Important Event

#### Picture Descriptions

- Broken Window
- Refused Umbrella
- Cat in the Tree
- Flood

#### Story Narrative

- Cinderella

#### Procedural Discourse

- Peanut Butter and Jelly Sandwich

#### Test Data

- WAB, BNT (short), Repetition Test, Verb Naming Test



## HOW DO I JOIN?

AphasiaBank is already home to 151 members and growing. Members are urged to contribute, but any member can use the database *at no cost*.

Visit <http://talkbank.org/AphasiaBank/> for guidelines and membership forms

## WHY JOIN?

- Access to database for your own research
- Information on ideal recording techniques and protocol materials
- Instructions for transcription and analysis
- Workshops on transcription and analysis
- Membership in AphasiaBank Googlegroups for interactions with other members, and new information concerning the database
- Help available by telephoning AphasiaBank staff

## WHO IS IN THE DATABASE?

TABLE 1. PARTICIPANT DEMOGRAPHICS:

	APHASIA (n=99)	NON-APHASIA (n=102)
AGE RANGE (YRS)	35.9 - 90.7 (mean=64.3)	23.0 - 87.8 (mean=60.9)
GENDER	34 females 65 males	55 females 47 males
EDUCATION RANGE (YRS)	12-25 (mean=15.6)	10-22 (mean=15.1)
HANDEDNESS	85 right 8 left 5 ambidextrous 1 unavailable	89 right 10 left 3 ambidextrous

TABLE 2. PARTICIPANT WAB TYPES:

Anomic	33
Broca	25
Conduction	14
Wernicke	9
Above WAB Cutoff	9
Transcortical Motor	5
Global	2
Transcortical Sensory	1
Not Available	1

## DATA ANALYSIS TOOLS

**CLAN** The editor provides functions such as media playback, data validation, linkage to audio and video, and a comprehensive set of data analysis programs.

**CHAT** A standardized format for transcription, coding, and detailed morphological analysis of conversational interactions.

**CODES FOR APHASIC ERRORS** New codes capture paraphasia, word substitution, neologism, agreement, metathesis, perseveration, part of speech, grammar, jargon, empty speech, and circumlocution.

## SAMPLE CHAT-CODED TRANSCRIPT

The MOR program has been used on the following excerpt to add a line after each tier, identifying the part of speech of each word and specifying the morphology of affixes.

\*INV: do you remember when you had your stroke ?

\*PAR: oh &um ninety four .  
%mor: co|oh det:num|ninety det:num|four .

\*PAR: &um &mar I don't know the date but I think so &um maybe [ / ] m:aybe &eh &sep &um March twenty fifth . [+ gram]

%mor: pro|I aux|do~neg|not v|know det|the n|date conj:coo|but pro|I v|think co|so adv|maybe n:prop|March det:num|twenty adj|fifth .

\*PAR: but I don't know the precise brate [: date] [\* pn] but &uh end o(f) the month yeah .

%mor: conj:coo|but pro|I aux|do~neg|not v|know det|the adj|precise n|date conj:coo|but n|end prep|of det|the n|month co|yeah .

## EXAMPLES OF DATA ANALYSIS

The FREQ program is used for frequency analysis. Following are three examples of the powerful uses of this simple program applied to a small subsection of the AphasiaBank database.

**EXAMPLE 1.** To find the total number of nouns (@|-n,-n:\*,o-%), stems only, collapsed across files in a folder (+u), in descending order (+o) :

**freq +t%mor +t\*PAR -t\* +d5 +o +s"@|-n,-n:\*,o-%" +u \*.cha**

This command would output the following:

```
292 n
31 n:prop
2 n:gerund
2 n:pt
1 n|+n+n
1 n|+v+ptl
```

**EXAMPLE 2.** To count and list the nouns (r-\*,|-n,-n:\*,o-%), use:

**freq +t%mor +t\*PAR -t\* +d5 +o +s"@r-\*,|-n,-n:\*,o-%" +u \*.cha**

```
31 n:prop|Cinderella      7 n|person
21 n|sudden              7 n|shoe
17 n|ball                 6 n|glass
14 n|home                 6 n|mouse
13 n|girl                 6 n|woman
11 n|slipper              5 n|fairy
10 n|house                 5 n|horse
10 n|prince                4 n|child
9 n|mother                 4 n|family
8 n|dress                  4 n|godmother
7 n|boy
7 n|man
7 n|o'clock                ...and so on
```

**EXAMPLE 3.** If you want paraphasic errors, parts of speech, and bound morpheme information (+d6), use:

**freq +t%mor +t\*PAR -t\* +d6 +o +s"@r-\*,|-n,-n:\*,o-%" +u \*.cha**

```
31 n:prop|Cinderella
30 n:prop|Cinderella
1 n:prop|Cinderella@Secerundid*nk
21 n|sudden
21 n|sudden
17 n|ball
17 n|ball
14 n|home
14 n|home
13 n|girl
2 n|girl
9 n|girl-PL
```

## WHAT QUESTIONS CAN APHASIABANK DATA ANSWER?

### PROJECTS UNDERWAY:

- Developing a lexicon for Cinderella story -- an objective metric for content.

MacWhinney, B., Fromm, Holland, A., Forbes, M., Wright, H. (2010). *Automated analysis of the Cinderella story*. Aphasiology, iFirst, 1-13.

- Developing data-based new approaches to oral language-based aphasia classification

### SOME MORE IDEAS (AMONG MANY):

- Verb argument structure in discourse samples of agrammatic speakers
- Word usage errors in aphasia, related to traditional aphasia syndromes
- Comparisons of discourse measures used in the protocol

- Correlational studies of test data with measures in protocol
- Changes over the course of treatment
- Nature of paraphasic errors: are traditional categories useful?

### IN THE FUTURE:

- TBIBank (NHRSC funding, Leanne Togher, PI, University of Sydney, Australia)
- DementiaBank (proposal pending)

This project is funded by NIH\_NIDCD grant R01-DC008524 (2007-2012).