Language Banking for Language Disorders

Brian MacWhinney, CMU
Audrey Holland, U. of Arizona, Emerita
Nan Bernstein Ratner, U. of Maryland
Margie Forbes, CMU
Davida Fromm, CMU
Leanne Togher, U. of Sydney
Michelle Bourgeois, USF
Language Banking for Language Disorders

Introduction

Brian MacWhinney, CMU
Disclosures

- Brian MacWhinney, Davida Fromm, and Margie Forbes receive salaries from the NIH AphasiaBank grant.
- Audrey Holland is a consultant on the AphasiaBank grant.
- Nan Bernstein Ratner is a consultant on the CHILDES grant.
- Leanne Togher: see her summary
- Michelle Bourgeois: see her summary
What is Language Banking?
## TalkBank Components

<table>
<thead>
<tr>
<th>Bank</th>
<th>Collaborators / Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHILDES</td>
<td>Catherine Snow, NICHD</td>
</tr>
<tr>
<td>PhonBank</td>
<td>Yvan Rose, NICHD</td>
</tr>
<tr>
<td>AphasiaBank</td>
<td>Audrey Holland, Davida Fromm, Margie Forbes, NIDCD</td>
</tr>
<tr>
<td>SLABank, BilingBank</td>
<td>Pittsburgh Science of Learning Center, DFG/NEH</td>
</tr>
<tr>
<td>TIBBank</td>
<td>Leanne Togher, NHMRC Australia, Lyn Turkstra</td>
</tr>
<tr>
<td>DementiaBank</td>
<td>Michelle Bourgeois, Dan Kempler</td>
</tr>
<tr>
<td>FluencyBank, ChiDisBank</td>
<td>Nan Ratner, NICHD</td>
</tr>
<tr>
<td>CABank</td>
<td>Johannes Wagner, Southern Denmark University</td>
</tr>
<tr>
<td>--- and six others</td>
<td></td>
</tr>
</tbody>
</table>
http://talkbank.org
Uniform Resources

- CHAT transcription format
- CLAN analysis programs
- Multimedia
- Web-Browsable Database
- Freely downloadable data, programs, manuals
- Community integration: meetings, bboards
- Totally Open Access
Uniform Goals

1. Emphasis on natural spoken communication, but with inclusion of other relevant measures
2. Provide data for research studies:
   • Increased quality
   • Increased quantity
3. Create measurement standards based on publicly available data for:
   • Measuring treatment outcomes
   • Referencing patient to clinical database
4. Samples and illustrations for teaching and training
Human Speechome
Current Status

- CHILDES: 6000 articles, 163 corpora
- AphasiaBank: 100 articles, 371 PWAs, 190 controls
- DementiaBank: 302 dementia, 235 controls Michelle will provide update
- TBIBank: Leanne will provide update
- ChiDisBank: Nan will provide update
Protocol

- Shared protocol for AphasiaBank and TBIBank.
- Will be extended to DementiaBank
- Demographics, WAB, BNT, VNT, repetition task
- Stroke story, recovery, personal event
Picture sequence stories
Cat in the tree
Discourse

Cinderella retelling

How to make a PBJ
Making things easier

• Transcription
  • Linked to audio
  • Online correction in editor
  • Levels of precision
• Morphosyntactic tagging
  • Automatic tagging, automatic disambiguation
• Automatic analyses and profiling through EVAL, KIDEVAL
EVAL

- MLU, TTR
- Verbs/Utt
- % errors
- % N, V, Aux, Adv, Conj, Pro
- % PAST, PASTP, PL
- Retracing, repetition
KIDEVAL

- Implemented also for Spanish and French
- Same measures as EVAL, but also:
  - DSS
  - IPSyn
  - VocD
  - Brown’s 14 morphemes
- Planned comparison sets: Weismer, Hall, HSLLD
KIDEVAL Corpora

- Weismer – SLI, Controls
- Gillam – SLI, Controls
- PEPPER & PhonBank
- Spanish – Iglesias
- Snow – HSLLD
- ......
And now the next presentations
USING TALKBANK UTILITIES TO STUDY CHILD LANGUAGE DISORDERS

- THE CLINICAL GUIDE TO CLAN
- USE OF THE NEW KIDEVAL UTILITY
- FUTURE NORMING STUDIES

Nan Bernstein Ratner
The University of Maryland
Why am I here? And why do I insist that all my students learn to use CLAN?

- Language sample analysis is the “gold standard” for analysis of child language behaviors, and is mandated by many jurisdictions as part of the qualification process to receive school-based services.

- However, language sample analysis:
  - Takes a LOT of time
  - Is subject to errors in labeling/parsing of language structures
  - Is subject to math errors 😞
  - Has norms that are surprisingly limited (especially at younger ages)
  - I WILL ADDRESS EACH OF THESE IN TURN
TIME (there’s never enough of it)

- Just simple transcription of a child language sample takes a lot of time
- Professional transcribers estimate that:
  - “it takes a minimum of four (4) hours to transcribe a one hour tape and could take up to 6 to 8 hours depending on the quality of the tape.” (transcribe-it.inc)
- CLAN dynamically links either audio or video to the transcript as you build it, using free utilities
  - This greatly increases precision
  - Allows for full appreciation of speaker quality as well as content
  - AND SPEEDS UP TRANSCRIPTION BY A FACTOR OF 4-5X.
    - Example from our recent grant: we were able to do the following...
Scope of our recent grant work using CLAN

- ~125 families, 5 play sessions, with both child and mother verbal interaction a focus of analysis
- Using a standard set of toys; ~15 minutes, followed by adult-adult interview
- Total number of 15-30 minute transcripts = ~1250
- Estimate of time required to complete transcription and analysis (pre-award) = up to 5 hrs/transcript to compute multiple measures
- = 6250 hours
- Completion during award period was not predicted
- **We were wrong.** We got it done within 3 years, using typical part-time and volunteer researchers.
Here’s what a typical transcript looks and sounds like…
the new boy has a new pet frog. *1_4740*
while the boy [* was] sleeping and his dog [* was] sleeping the frog got out of <his> [/] his bowl. *4740_16709*
and then he wandered off. *16709_18522*
then when +/. *18522_19635*
talk real loud. *19500_20352*
and then <when he woke up> [/] <the boy woke> [/] when the boy woke up the frog was gone. *20352_26949*
(.) he looked in his shoes. *26949_41662*
and his dog looked in the bowl. *41662_44488*
and then he looked outside. *44488_48993*
(.) then his dog fell through the window. *49877_58425*
then he was so mad at his dog. *58424_62363*
talk real loud. *59200_64831*
he <look> [/] looked in the forest. *64831_71355*
then he looked under [* in] a hole <and his> [/] while his dog was
Practical issues in LSA for clinical purposes

- **Time-consuming to CODE!**
  - E.g., estimates of 8 hours training and 45 mins - 1 hour POST-transcript for computation of DSS (Lee, 1970; Long & Masterson, 1993; Long, 2001)
  - More than 30 mins per sample for IPSYN (Hassanali et al., 2014)
  - Prone to error
  - **Difficult to use same worksheet to compute multiple measures**

- **Result:** although difficult to verify,
  - a sense that LSA rarely used in most clinical settings
  - except when required by local regulations
OK, so transcription is easier, WHAT ELSE?

- CLAN automatically “tags” utterances for part-of-speech and will tally errors that you mark as you go along.

  - *CHI:  and him [*] fall [*] down and the shark eat [*] him all gone.
  - %mor:  coord | and pro:obj | him v | fall adv:loc | down coord | and det | the n | shark v | eat pro:obj | him post | all part | go&PERF.

  - *MOT:  there is no shark in that pond you silly thing.
  - %mor:  adv:loc | there cop | be&3S qn | no n | shark prep | in det | that n | pond pro | you adj | silly n | thing.

  - *MOT:  there’s just the froggy.
  - %mor:  adv:loc | there~cop | be&3S adv:int | just det | the n | frog-DIM.

  - *CHI:  shark [*] right down there in the pond.
  - %mor:  n | shark adv | right prep | down adv:loc | there prep | in det | the n | pond.
KidEval: the niftiest thing since sliced bread for LSA

- Once a single transcript is made, it is possible, through the syntactic parsing (done by MOR and POST) to generate multiple, simultaneous measures of child language function:
  - The typical measures: MLU (in words and morphemes)
  - Types, tokens, TTR and NDW (these often take longer to do by hand)
    - VOCD for vocabulary diversity (less resistant to sample size bias, but impossible to compute by hand)
  - Developmental Sentence Score (DSS) and IPSYN
    - These are enormously time-consuming to compute, but relatively easy to automate
  - Presence of important developmental morphemes (aka Brown’s morphemes) in English
Here is a sample output:
<table>
<thead>
<tr>
<th></th>
<th>MU1 Utts</th>
<th>MU1 Word</th>
<th>MU1 Morp</th>
<th>MU100 Utts</th>
<th>MU100 Word</th>
<th>MU100 Morp</th>
<th>MTypes</th>
<th>Total Utts</th>
<th>MU1 Utts</th>
<th>MU1 Word</th>
<th>MU1 Morp</th>
<th>MTypes</th>
<th>Total Utts</th>
<th>MU1 Utts</th>
<th>MU1 Word</th>
<th>MU1 Morp</th>
<th>MTypes</th>
<th>Total Utts</th>
<th>MU1 Utts</th>
<th>MU1 Word</th>
<th>MU1 Morp</th>
<th>MTypes</th>
<th>Total Utts</th>
<th>MU1 Utts</th>
<th>MU1 Word</th>
<th>MU1 Morp</th>
<th>MTypes</th>
<th>Total Utts</th>
<th>MU1 Utts</th>
<th>MU1 Word</th>
<th>MU1 Morp</th>
<th>MTypes</th>
<th>Total Utts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Legend:**
- **Total Utts**: Total number of utterances.
- **MU1 Utts**: Utterances with at least one morpheme.
- **MU1 Word**: Utterances with at least one word.
- **MU1 Morp**: Utterances with at least one morpheme.
- **MTypes**: Types of utterances.
- **SUS**: Sentences with unique structures.
- **DSS**: Distinct sentence structures.
- **VOCD**: Vocabulary diversity.
- **c_iP**: Number of unique morphemes.
- **iP**: Number of unique words.
- **PToT**: Percentage of total.
- **AP**: Agreement percentage.
- **PAST**: Past tense percentage.
- **PPR**: Pronoun reference rate.
- **% det**: Percentage of determiners.
- **% past tense**: Percentage of past tense words.
- **% AN**: Percentage of adverbs.
- **% AO**: Percentage of adjectives.
- **% AP**: Percentage of adjectives.
- **% AvT**: Average tense.

**Note:** The table includes cumulative counts and percentages for various linguistic features across different types of utterances. The data is used to evaluate the language development of children.
I have been teaching CLAN in required graduate classes for 3 years

- There is an initial groan when some computer-averse people realize I am serious about this
- And the first sample takes the longest
- But, here is what the students say:
Comments from FIRST SEMESTER MA students (remember how life is for them?)

- “I saved myself a lot of time (outside of class): I use it with clients and dx cases. It gives me concrete numbers that I am not afraid to second guess.”
- “Super useful and time-saving!”
- “Once you learn to transcribe, easy to use – saves time, provides LOTS of analysis.”
- “The Walker/sound loop feature is INVALUABLE!! – saves so much time in transcription!”
- “Time-efficient.”
- “GREAT for ease of transcription!!”
- “A huge time-saver. Very useful!”

- It definitely was not fun for everyone, but the majority would never transcribe or analyze by hand again.
Next: interpreting LSA

We go to all that work doing LSA, but how much do we learn from these measures?
Conceptual issues in LSA for young children

- Normative values are based on relatively few cases at lowest age ranges:
  - **MLU**: Most recent: Hadley, Rispoli et al., 37 children at 24 months
    - Rice et al. (2010) norms – 17 unaffected, 6 affected (LTs) from 2;6 to 2;11
  - **TTR or NDW**: nothing robust under 3;0
    - Sample size dependent, or must have standard 100 word sample
  - **DSS**: 40 children 24-36 mos; 10 children 24-27 mos (Lee, 1972).
  - **IPSyn**: 15 children (Scarborough, 1990)
Our goal: ChiDisBank and reference norms

- We are currently in the process of computing KidEval measures on typical and disordered children in English and other languages.
- This will allow:
  - More robust norms and confidence intervals for both typical development
  - And atypical populations, including children
    - with SLI,
    - hearing loss,
    - bilinguals,
    - low SES, etc.
Future steps:

- “Fluency Bank” to study the onset and developmental features of stuttering
- Including those factors that:
  - May distinguish recovery from persistence in children
  - Language factors that appear to aggravate fluency and require treatment planning consideration
  - Etc.

- SO....
THANKS for your attention!
## AphasiaBank – Clinical Populations

**EVAL**

<table>
<thead>
<tr>
<th>Famous People</th>
<th>Margaret Forbes</th>
</tr>
</thead>
</table>

EVAL

For adult language samples
(KidEval is its offspring)

Initially created as a tool for busy aphasia clinicians
Can also be used for research
Designed to produce maximum useful information in minimum time

- Based on discourse level - the level of real communication
- Can be done quickly - streamlined transcription and error coding
- Computes 30 items in one pass
- Language profile is displayed in an EXCEL spreadsheet
EVAL

Analyzes any discourse sample

Must use AphasiaBank tasks (at the website) to compare to database

Choices: free speech, picture description, story retell, procedural discourse
EVAL

Four-step Process

- Administer and record a brief discourse sample - option to use AphasiaBank tasks
- Transcribe and code using streamlined methods
- Run EVAL - option to compare your sample to the database
- Open the EXCEL spreadsheet with 30 language measures side-by-side with comparison group results
EXAMPLE: Compare PWA with the database

Administer, record and transcribe the picture description task (Window) according to EVAL conventions (manual at the AphasiaBank website)

In CLAN, run the EVAL program
EVAL
EVAL

EVAL spreadsheet shows PWA results side-by-side with the reference group.

It also displays for database:
Mean
Minimum
Maximum
Standard deviation

Indicates where your sample differs from the database by one or two standard deviations.
## EVAL

<table>
<thead>
<tr>
<th>File/DB</th>
<th>Speaker ID</th>
<th>Duration</th>
<th>Total Utts</th>
<th>MLU Utts</th>
<th>MLU Words</th>
<th>MLU Morph</th>
<th>FREQ types</th>
<th>FREQ tokens</th>
<th>FREQ TTR</th>
<th>Verbs/Utt</th>
</tr>
</thead>
<tbody>
<tr>
<td>eval_demo.eng</td>
<td>EVALdev1</td>
<td>0:00:00</td>
<td>5</td>
<td>5</td>
<td>7.6</td>
<td>7.8</td>
<td>24</td>
<td>38</td>
<td>0.632</td>
<td>0.8</td>
</tr>
<tr>
<td>+/- SD</td>
<td>-0:00:00:00.99</td>
<td>-1.83</td>
<td>-1.836</td>
<td>-0.968</td>
<td>-1.41</td>
<td>-3.05</td>
<td>-1.99</td>
<td>6.355</td>
<td>-2.663</td>
<td></td>
</tr>
<tr>
<td>Mean Database</td>
<td></td>
<td>0:17:28</td>
<td>204.352</td>
<td>203.78</td>
<td>9.255</td>
<td>10.555</td>
<td>516.522</td>
<td>1836.363</td>
<td>0.302</td>
<td>1.502</td>
</tr>
<tr>
<td>Min Database</td>
<td></td>
<td>0:04:22</td>
<td>51</td>
<td>51</td>
<td>6.573</td>
<td>7.529</td>
<td>198</td>
<td>461</td>
<td>0.177</td>
<td>1.073</td>
</tr>
<tr>
<td>Max Database</td>
<td></td>
<td>3:28:59</td>
<td>739</td>
<td>734</td>
<td>18.758</td>
<td>21.556</td>
<td>1241</td>
<td>6610</td>
<td>0.461</td>
<td>3.293</td>
</tr>
<tr>
<td>SD Database</td>
<td></td>
<td>17:28.5</td>
<td>108.934</td>
<td>108.259</td>
<td>1.71</td>
<td>1.954</td>
<td>161.478</td>
<td>903.543</td>
<td>0.052</td>
<td>0.263</td>
</tr>
</tbody>
</table>

+/- SD  * = 1 SD, ** = 2 SD

Database keywords: control

# files in database: 182
EVAL

Language profile would take many hours to produce by hand.

Many clinical uses including:

• Initial language assessment

• Identify language problems in context

• Track and document progress in therapy

• Compare to a group you select from the large AphasiaBank database of aphasic and non-aphasic adults

• Guide treatment planning
Go to TalkBank.org/AphasiaBank, look for Programs and Manuals
THE FAMOUS PEOPLE PROTOCOL:

The most famous person in American History, according to the Atlantic magazine
Famous People

• Because the AphasiaBank protocol emphasizes spoken DISCOURSE, individuals with more severe aphasias and/or with severe AOS are under-represented

• To minimize this problem, we developed the Famous People Protocol (FPP)
Famous People

The Famous People Protocol consists of:

• 23 photos of famous people (entertainers, world figures, sports stars, U.S. Presidents)
• questions, mostly yes/no, designed to assess knowledge of these people and the events associated with them
Famous People

The goal is to assess what PWA **KNOW**, as opposed to what they can actually **SAY**, that is…

- To assess **KNOWLEDGE** -- how PWA express what they know - singing or humming, part word responses, gestures, answering yes/no questions about the famous people, whole or part word writing, etc.

- Differentiate “**I DON’T KNOW**” meaning, “I can’t say it” from when it is actual **lack of knowledge**
Clinical Uses

• Provide families and clinicians with some clues about techniques that can help them to differentiate what their PWA KNOW but can’t say, from what he or she REALLY doesn’t know

• Clinicians can use this information to determine ways to support the communication of their PWA and communicate this info to families
Tool for Conversation Coaching

• Over 60 PWA have completed the protocol, including moderately impaired people

• If we videotape it and then watch and discuss it with the PWA and significant other(s), it can be an important approach to teaching *conversational coaching*
Famous People

Go to TalkBank.org/AphasiaBank, and look in Protocols
AphasiaBank

Educational Applications

Research Applications

Davida Fromm
AphasiaBank

as of November 2014 ...

• English -- transcripts & media – standardized discourse protocol
  o 371 participants with aphasia
  o 190 non-aphasic controls
• Non-protocol material, aphasia groups, etc.
• Cantonese, French, German, Greek, Hungarian, Italian, Mandarin, and Spanish

• 367 professional members of AphasiaBank (from 34 countries)
  o educators, researchers, clinicians
Educational Applications

Example #1
• Create 2-minute videos of speakers with Broca aphasia and Anomia
• Save them as .wav files for students to access on Blackboard
• Students complete a language analysis on one of each aphasia type

Example #2
• Give students transcripts of language data
• Students analyze transcripts according to different linguistic models

Example #3
• Students prepare case studies
• Use transcripts, demographic data, test result data
Educational Applications, cont.

SHOW stuff
• what aphasia looks like and sounds like
• search database by aphasia type, AQ score, errors, tasks, etc.

LEARN how to do stuff
• transcription of aphasic language
• coding/annotation procedures
• linguistic analyses – macro, micro
• phonological analyses
• acoustic analyses
• discourse analyses
• conversation analyses
• differential diagnosis
Educational Applications, cont.

- Transcript Database
- Media Database
- Database Guide
- Browsable Database

Browse all transcripts
Simultaneously watch video and follow transcript
Search database
TalkBank Transcript Browser

Web browser support:
The TalkBank Transcript Browser has been tested to work with current versions of Firefox, Safari, Opera, and Chrome with Javascript enabled. It does not work with Internet Explorer. If you are experiencing any problems, please ensure that your software is up to date.

Fixing your QuickTime streaming settings:
If you have updated software and a good connection, but are still having difficulty playing media, your QuickTime streaming settings may have become corrupted. Fixing them is a bit complex, but often works.

Click here for directions on fixing your QT streaming settings.

To browse for materials to play:
Use the left menu to navigate through the TalkBank corpora. Click once on the folder you want to explore, and it will expand to show any available subfolders and transcripts. Once you have found the transcript you want to review, click its filename. Your browser will load the transcript in this space and any associated media below the navigation.

Next to each transcript name is an icon displaying whether the file has linked audio, video, or no media.

Controlling playback:
For transcripts that are linked to media, you can control playback in three ways:

1. Use the controls that display below the media to play, pause, and rewind/fast forward.
2. Use your mouse:
   ○ Play from the selected line until the end of the media: press the › at the end of the line.
   ○ Pause playback: press the ■ at the end of the playing line.
3. Use keyboard shortcuts:
   ○ Start playback: while the media is stopped, press s.
   ○ Pause playback: while the media is playing, press s.
   ○ Play the next line: press CTRL + Right arrow.
   ○ Play the previous line: press CTRL + Left arrow.
   ○ Play the selected line only: click the line once, then press p.

Continuous playback:
By default, media playback is continuous. However, some transcripts contain linked utterances separated by long periods of silence or untranscribed audio. To skip these periods and only play the linked utterances, set the "Continuous playback" option to "Off", and then click "Set options".

To run CLAN commands:
CLAN commands may be entered in the text box below Command line: (on left, below the directory listing), and run entirely in the browser. File names can be entered manually, or by pressing the [*] symbol next to the file name. Displayed next to Command line: is your current working directory.

Hiding dependent tiers:
Codes, comments, events, and descriptions of interest to the researcher are often typed below the main tier. If a file contains such dependent tiers, a listing of these tiers will display below the navigation on the left. For legibility, the user has the option of only displaying select tiers. To hide a tier, deselect its checkbox, and then click "Set options".
Search the Browsable Database

Find all participants with Wernicke’s aphasia

Command line: AphasiaBank/English/Aphasia/

freq +s"*Wernickel*" +t*PAR +t@ID *.cha +u +re > Run

28 Total number of different item types used
56 Total number of items (tokens)
0.500 Type/Token ratio
Browsable Database

2 videos – personal narratives about stroke

- Wernicke  AQ = 43.1  tucson15a
- Broca  AQ = 59.7  wright205a

Note: Transcripts will show speaker tiers only for ease of following.
But the transcripts also include other tiers such as:
1. a morphological tier under each speaker tier
2. a grammatical relations tier
@Loc: AphasiaBank/English/Aphasia/Tucson/tucson15a.cha
@Begin: AphasiaBank/English/Aphasia/Tucson/tucson15a.cha
@Languages: eng
@Participants: PAR tucson15a Participant, INV Investigator
@ID: eng:Tucson|PAR|74|1.1|male|Wernicke|tucson15a|Participant|||43.1|
@ID: eng:Tucson|INV|...|tucson15a|Investigator||
@Media: tucson15a, video
@G: Speech
*INV: I'm going to be asking you to do some talking.
*PAR: how do you think your speech is these days?
*PAR: it's working but it's working hard.
*PAR: yeah.
@G: Stroke
*INV: did you have a stroke?
*PAR: yes.
*INV: do you remember when you had your stroke?
*PAR: it's <two years> /[ ] &=ges:two two years ago.
*INV: uhhuh.
*INV: can you tell me about it?
*PAR: I just <heard the> /[ ] heard the saber *[ *:s:uk] <when I> /[ ] &uh when I just would stop.[+ jar]
*PAR: had a bid *[ *:s:uk] . [+ jar]
*PAR: <and then> /[ ] and then it went or> /[ ] &i it stayed for a long time . [+ es]
*PAR: &=ges:mouth build *[ *:s:uk] for the long time . [+ jar]
*PAR: and then finally I xxx and walk . [+ jar]
*PAR: then here again in xxx I hit the wheels too . [+ jar]
*PAR: it's gotta do with coffee *[ *:s:uk] . [+ jar]
*PAR: but that wouldn't break his arm he just &j break . [+ jar]
*PAR: or the mother *[ *:s:uk] . [+ jar]
*PAR: and I lived through it, yeah.
*PAR: it's getting better though.
*PAR: but it's &uh hard.
*INV: it's been hard.
*PAR: yeah.
*INV: I bet.
*INV: tell me about your recovery.
*INV: what kinds of things have you done since your stroke to try to get better?
*PAR: well different groups. [+ gram]
*PAR: one two three . [+ gram]
*PAR: tapping is baby bə-buزمة [: x@m] *[ *:n:uk] . [+ jar]
*PAR: and &uh they help me .
*PAR: and they hɛv@u [: x@m] *[ *:n:uk] me better and think about things to do . [+ jar]
*PAR: and find the people &uh watching you know . [+ es]
*PAR: I [: ] I wanna [: want to] read when I can .
*PAR: <you know it> /[ ] you know it .
*PAR: I can see it but I have think about it but I have think about it bɛtə-o@u [: x@m] *[ *:n:uk] .
*PAR: and you know that same thing that you walk . [+ gram]
*PAR: I could do that but I think about that .
*PAR: better and better . [+ gram]
*PAR: or sometimes I say just +/.
@Loc: AphasiaBank/English/Aphasia/Wright/wright205a.cha
@Begin
@Languages: eng
@Participants: PAR wright205a Participant, INV Investigator
@ID: eng|Wright|PAR|55.10.|male|Broca|wright205a|Participant||59.7|
      eng|Wright|INV||wright205a|Investigator||
@Media: wright205a, video
*INV: www.
  @G: Speech
  *INV: I'm going to be asking you to do some talking.
  *INV: how do you think your speech is these days?
  *PAR: small & ges:little.
  *INV: small?
  *PAR: yeah & ges:yes.
  *INV: not enough speech?
  *PAR: hmmmm & ges:no.
  *INV: hmmmm.
  *PAR: nope.
  @G: Stroke
  *INV: do you remember when you had your stroke?
  *PAR: & ges:yes yeah.
  *PAR: can you tell me about it?
  *PAR: &uh: ges:fall fell down. [+ gram]
  *PAR: &uh pool. [+ gram]
  *PAR: &uh &uh son & points me & points: self. [+ gram]
  *PAR: &um &uh fell down. [+ gram]
  *PAR: & ges:fall & s fell down. [+ gram]
  *PAR: & ges: black out. [+ gram]
  *INV: what are your first memories after your stroke?
  *PAR: mm Dawn &uh wife. [+ gram]
  *PAR: (..) & ges:yes &uh ex wife yeah. [+ gram]
  *INV: oh.
  *INV: tell me about your recovery.
  *INV: what kinds of things have you done to try to get better since your stroke?
  *PAR: & uh & grabs:right leg left [↓ right] [* s:r-ret] [↓↓] & uh & right side. [+ gram]
  *PAR: & um [x 3] speech & ges:mouth & ges: no longer. [+ gram]
  *PAR: & um left [↓ right] [* s:r-ret] [↓] & points:right leg & th
  & points: right hand & ges: right hand < left [↓ right]
  [* s:r-ret] & points:right leg no> [↓↓] right this
  & points: right hand & ges: right side. [+ gram]
Educational Applications, cont.

If you want to download a transcript or media file …

The Database Guide – annotated index about every corpus
Research Applications

Research Applications, cont.

~ 100 entries in AphasiaBank bibliography

- journal articles
- book chapters
- conference papers and posters
- doctoral dissertations
- master’s theses
- senior theses
- student research projects


Presentations and Posters link …

The following are a selection of posters and presentations that have documented various AphasiaBank methods and applications:

- **Academy of Aphasia 2011** Is the Cinderella task biased for age or gender? -- Fromm, MacWhinney, Forbes, & Holland
- **Academy of Aphasia 2011** Computational analysis of AphasiaBank transcripts and video -- MacWhinney, Fromm, Forbes, & Holland
- **Academy of Aphasia 2012** Comprehensive statistical analysis of aphasia types -- MacWhinney, Fromm, Forbes, Holland, Makris, Todd, & Greenhouse
- **Academy of Aphasia 2013** Developing a new procedure for assessing communication in severe aphasia: The Famous People Protocol -- Holland, Fromm, Forbes, & MacWhinney
- **Academy of Aphasia 2014** Measuring the coherence of healthy and aphasic discourse production in Chinese using Rhetorical Structure Theory (RST) -- Kong, Linnik, Law, Shum
- **Academy of Aphasia 2014** Duration of content and function words in oral discourse by speakers with fluent aphasia: Preliminary data -- Lee, Kong, Wang
- **ASHA 2013** Discourse measures in aphasia and observer ratings of comfort and typicality -- Campbell, Williams, Hudspeth, Franklin, & Richardson
IS THE CINDERELLA TASK BIASED FOR AGE OR GENDER?

Davida Fromm, Brian MacWhinney, Margaret Forbes
Carnegie Mellon University, Pittsburgh, PA
Audrey Holland, University of Arizona, Tucson AZ
University of Arizona, Tucson, AZ
Research Applications, cont.

- linguistic analyses
- discourse analyses
- formulaic and figurative language
- gesture use and body movement
- disfluencies
- use of different verb types in spoken discourse
- use of pronouns/pronoun errors
- prosody/perception of prosody
- create machines with automated classifiers to distinguish among disorders
- regression analyses
- cross-linguistic analyses
Develop new measures
• a clinician-friendly discourse analysis tool using main concept analysis – Hudspeth et al.
• test the validity of new measures for global coherence – Fergadiotis et al.

Evaluate existing measures
• confirm the relationship between WPM and WAB fluency scale scores – Basilakos et al.
• use protocol tasks for test-retest reliability analyses – Boyle
Research Applications, cont.

Discourse comparisons
• determine if different types of discourse affect lexical diversity and productive vocabulary – Fergadiotis et al.
• determine the relationship between discourse production and confrontation naming in PWA – Fergadiotis et al.

Treatment
• use protocol tasks for treatment efficacy outcome measures – Kurland

Other
• determine if people can make accurate gender predictions based on reading samples from controls about illness and injury – Hancock et al.
To join AphasiaBank …

email Brian MacWhinney

macw@cmu.edu
Professor Leanne Togher

- **Relevant Financial Relationships:**
  - Salaried full professor in Communication Sciences and Disorders at the University of Sydney
  - Research funded by the Australian National Health and Medical Research Council
  - Recipient of a 6 year NHMRC Senior Research Fellowship
  - Speech Pathology Australia research funding
  - Technical support from TalkBank team at Carnegie Mellon University

- **Relevant Non-Financial Relationships:**
  - Serves as reviewer for several peer-reviewed journals
History of TBI Bank

- Expert consensus + Evidence: 1st internationally ratified and evidence-based communication protocol for use with TBI population
- Aphasia Bank tasks modified to include questions about recovery and coping after brain injury (rather than stroke)
- Added in extra communication and cognitive/executive functioning tasks for the purposes of a longitudinal communication recovery study
  - 10 minute casual conversation with a significant other
  - La Trobe Communication Questionnaire (Douglas et al 2000)
  - Neuropsychological screening battery
Communication after TBI

- Discourse requires integrated cognitive, linguistic and social skills
- Discourse analysis = Gold standard (Coelho 2007; Togher 2001)

Challenges:
- Time consuming
- Many choices for analysis

TBI Bank can help:
- Transcribe faster and more accurately
- Run multiple analyses with a click
The goal of TalkBank is to foster fundamental research in the study of human and animal communication. It will construct sample databases within each of the subfields studying communication. It will use these databases to advance the development of standards and tools for creating, sharing, searching, and commenting upon primary materials via networked computers.
TBI Bank

Database
- Browsable Database
- Transcript Database
- Media Database

Protocol
- Protocol List
- Protocol Instructions
- Full Protocol
- Troubleshooting

Protocol Results
- Coding Sheet for Investigator Administration
- Coding Sheet for Self Administration
- Coding Sheet to Supplement Self Administration
- Excel Spreadsheet for Data
- Excel Spreadsheet for Data (blank)
- Test Results Spreadsheet
- Test Results Spreadsheet (blank)
Monologic Discourse Tasks

› I. Free Speech Samples (Brain Injury Story & Important Event)
› II. Picture Descriptions (Broken Window; Refused Umbrella & Cat Rescue)
› III. Story Narrative (Cinderella)
› IV. Procedural Discourse (Peanut Butter & Jelly Sandwich or similar)
A. BRAIN INJURY STORY and COPING

1. “I’m going to be asking you to do some talking. How do you think your speech is these days?”
   If no response in approximately 10 seconds, prompt:
   “How's your talking?”
   Listen, encourage full response. If no response, use Troubleshooting questions.

2. Tell me what you remember about when you had your head injury.
   If no response in approximately 10 seconds, prompt:
   “What other things have people told you about when you had your head injury?”
   Listen, encourage full response.
   At a natural juncture add:
3. "Tell me about your recovery. What kinds of things have you done to try to get better since your head injury?"

If no response in approximately 10 seconds, prompt:

“Tell me about any changes you’ve needed to make in your daily life.”
Cultural differences – an Aussie adaptation

› Cheese and vegemite sandwich
Supplementary Tests

- Aphasia Bank Repetition Test (2007)
- Verb Naming Test (from the Northwestern Assessment of Verbs and Sentences-Revised, Field Test Version)
- Western Aphasia Battery-Revised (2007) -- AQ only
- Optional - Verbal Fluency (F, A, S).
- Repeatable Battery for the Assessment of Neuropsychological Status (1998)
Longitudinal communication recovery following severe TBI
(Togher, McDonald, Tate, Turkstra, Holland & MacWhinney)

Background and progress

› Major aims of this project are to identify predictive factors of communication outcomes at 2 years and to identify optimal time frames for recovery

› 5 year NHMRC Project Grant

› N=58 participants

› Data collected at 3, 6, 9, 12 and 24 months post injury

› Data collection will be completed by March 2015 (i.e. 2 year data)

› Protocol includes:

  - Standardised communication assessments including aphasia, dysarthria, cognitive communication

  - TBI Bank Protocol – range of standardised tests and discourse tasks including picture description, procedural description, illness narrative, conversation with significant other

  - Neuropsychological screening battery
Question: Is it feasible to conduct the TBI Bank protocol with TBI participants during the sub acute phases of recovery (i.e. at 3 months and 6 months post injury)?

Participants:
- N=48 participants with severe TBI were assessed using the TBI Bank protocol at 3 months
- N=53 participants were assessed at 6 months
Compliance with protocol completion

At 3 months
› 45/48 completed the full protocol
› 3/48 partially completed the protocol

At 6 months
› 51/53 participants completed the full protocol
› 2/53 partially completed the protocol
Subtest completion

Participants were less likely to complete:

✗ The Aphasia Bank Repetition Test
✗ Describing an important event

All participants completed:

✓ Cinderella Story
✓ Recovery narrative
✓ Procedural discourse task (making a Cheese and Vegemite sandwich)
Studies in progress

• Comparison of 3 and 6 month data
  • conversational topics (Sophie Brassel – SP Honours)
  • global ratings of conversation (An An Chia – SP Honours)
  • procedural discourse (Elin Stubbs – Karolinska Institute Sweden Masters)
    • narrative discourse – picture description (Stephanie Weir – BHS Honours – collaborating with Jessica Richardson)

• Recovery during the first 12 months (Elise Elbourn - PhD)
  • Incidence of aphasia, dysarthria and cognitive communication disorders
  • Narrative discourse recovery
  • Important event
Conclusions

TBI Bank

› Diagnostic tool
› Feasible for use with participants in the subacute stages of recovery
› Objective discourse measures from computerised analysis
› Aims to reduce participant burden in long term
› Shared international database to foster collaborative research activity
› Interested parties can contribute discourse to the repository
› Tool for students and researchers to study spoken discourse
Contact: leanne.togher@sydney.edu.au
Michelle S. Bourgeois, PhD, CCC-SLP

Disclosures:
No Financial relationship with AphasiaBank
Employed by the University of South Florida
This guide provides documentation regarding the DementiaBank corpora in the TalkBank database. All of these data are available from http://talkbank.org/DementiaBank/. DementiaBank data are only available to members.

- TalkBank is an international system for the exchange of data on spoken language interactions. The majority of the corpora in TalkBank have either audio or video media linked to transcripts. All transcripts are formatted in the CHAT system and also are available as XML.

- To jump to the relevant section, click on the page number to the right of the corpus.

- 1. Holland 2
- 2. Kempler 2
- 3. Pitt (University of Pittsburgh School of Medicine) 3
## Index of /media/DementiaBank

<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holland/</td>
<td>26-Mar-2014 15:30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kempler/</td>
<td>04-May-2010 15:47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lu-mandarin/</td>
<td>04-May-2010 15:47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lu-taiwanese/</td>
<td>27-Nov-2012 14:50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSU/</td>
<td>04-Jun-2012 15:05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitt/</td>
<td>03-Oct-2014 12:00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Source of Pitt Data

This folder contains transcripts and audio files from several sections of the language assessment gathered as part of a larger protocol administered by the Alzheimer and Related Dementias Study at the University of Pittsburgh School of Medicine.

The study was funded in 1983 by the National Institute on Aging. The Alzheimer Disease Research Center was created in 1985 and funded by the National Institute on Aging as well.

- Participants included elderly controls and people with probable and possible Alzheimer’s Disease. Data were gathered longitudinally, on a yearly basis.

- Demographic data are available from the "Data Spreadsheet" link under the heading "Pitt Corpus". Notes about the spreadsheet are in a ReadMe file link under the same heading.

- Four separate folders contain transcripts and audio files from these tasks: Cookie Theft picture description, Sentence Construction, Story Recall, Verbal Fluency.

- Students from FSU, OSU, and USF are trained to clip specific files from the audio file, transcribe them using the CHAT procedures, link the transcript to the audio file, and upload to DementiaBank.

- Several students have conducted studies using CLAN to analyze specific subsets of the data.
Language Tasks

- **Cookie Theft Picture Description**
  
  How many...
  - days in a year?
  - inches in a foot?
  - eggs in a dozen?
  - senators to a state?
  - ounces in a pound?

- **Yes/No Questions (from Western Aphasia Battery)**
  
  Is your name Smith (Brown, Matthews)?
  Do you live in Toronto (Pittsburgh, Windsor)?
  Are you a woman (man)?
  Are the lights on?
  Is the door closed?
  Do you eat a banana before you peel it?

- **Repeat sentences after investigator**
  
  He is fishing with George.
  Why don’t you give them some candy?
  etc.

- **Word fluency**
  
  Name any thing you can find in a supermarket. (Timed)
  Name all the items I (investigator) am wearing.
  Things that begin with the letter “F”

- **Name the item that...**
  
  you use to drink coffee/tea.
  you use to fix your hair.
  you use to tell the time.

- **What do you do with...**
  
  a razor?
  soap?
  a pencil?
  etc.

- **Name the word that is being described**
  
  A prickly plant in the desert
  An item used to hit a tennis ball
  etc.

- **How are the following words alike?**
  
  Apple/banana (& Are these words names of fruits or animals?)
  Coat/suit (& Are these words names of clothing or fruit?)
  etc.

- **Name the word that does not belong**
  
  Dog, cat, car
  Fish, car, train
  Boy, door, man

- **Does this sentence make sense? (If not, tell me why or correct the sentence.)**
  
  He drank some coffee.
  Where's in the package?
  I’m not as tall am I am.

- **Sentence Construction: Tell me a short, simple sentence using the given word(s)**
  
  Pencil
  Child/hospital
  etc.

- **Story Recall**
  
  Maze
  Toboggan
  etc.
# Index of /media/DementiaBank/Pitt

<table>
<thead>
<tr>
<th>Name</th>
<th>Last_modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>cookie/</td>
<td>03-Oct-2014 12:00</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>fluency/</td>
<td>04-Nov-2014 12:41</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>recall/</td>
<td>04-Nov-2014 12:42</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>sentence/</td>
<td>04-Nov-2014 12:42</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Research to date


Plans for the Future

- Expand the Database
  - Invite others to contribute data
  - Collect new data using modified AphasiaBank protocol
  - Obtain grant funding to support the expansion of the database
Language Banking for Language Disorders

Automation

Brian MacWhinney, CMU
Automated Analyses

- EVAL
- KIDEVAL
- CPIDR
- IPSyn
- DSS
Looking under the Hood
CPI DR

- Kintsch & Van Dijk 1978 / Anderson & Bower 1973
- Covington – CPI DR 1, 2, 3
- Computerized Propositional Idea Density Rater
- Snowdon, Kemper, et al. 1996 – Nun Study
- Covington linked proposition counter to his tagger.
- PD ranges from 0.00 to 1.00
- Our implementation depends on having already computed an accurate % mor line
Two Examples

P conj | and
pro | it
cop | be\&PAST\&13S
P adv | so
P adj | surreal

pro:sub | I
P v | think\&PAST
pro | it
cop | be\&PAST\&13S
prep | like
P qn | some
+...
DSS

- Developmental Sentence Scoring (Lee, 1974)
- Used in perhaps 100 studies.
- Implemented in CP (Long, Fey, Channell, 2003)
- Reimplemented in CLAN with higher accuracy
- Relies on %mor line
DSS Rule Fragment

FOCUS: \((v|\ast)^+ (cop|be&3S)^+ (aux|be&3S^v|\ast)^-\text{PRESP}\)

POINTS: C1

FOCUS: \((v|\ast\text{-PAST})^+ (v|\ast\text{-3S})^+ (v|\ast&\text{PAST})^+ (cop|\ast)^+ (aux|be^*)\)

POINTS: C2

FOCUS: 
\((mod|\text{can}^v|\ast)^+ (mod|\text{will}^v|\ast)^+ (mod|\text{may}^v|\ast)^+ (mod|do^v|\ast)^+ (mod|do^\text{neg}|not^v|\ast)^+ (mod|do^v|\ast)^+\)

POINTS: C4
Testing Rules

• When compared against age in large normal corpora as the Gold Standard, which segments of DSS are most predictive?
• Are some even negatively correlated?
• Can we add others that are more predictive?
• These questions can be addressed through automatic statistical methods
IPSyn and QPA

• IPSyn has been used in 63 empirical studies, although it is tedious to compute.
• It is useful once MLU ceilings out.
• QPA is seldom used, because it is difficult to compute by hand.
• Same logic: based on %mor, declarative rules, testing through automated statistical comparisons of rule sections
In Progress

• Automatic QPA
• Statistical Typology
• FluencyBank measures:
  • Pause detection
  • Fluency profile
  • Prosodic analysis
Thanks for your attention