AphasiaBank: A Resource for Clinicians

Margaret M. Forbes, M.A.,[1] **Davida Fromm, Ph.D.,**[1] **and Brian MacWhinney, Ph.D.**[1]

ABSTRACT

AphasiaBank is a shared, multimedia database containing videos and transcriptions of ~180 aphasic individuals and 140 non-aphasic controls performing a uniform set of discourse tasks. The language in the videos is transcribed in Codes for the Human Analysis of Transcripts (CHAT) format and coded for analysis with Computerized Language ANalysis (CLAN) programs, which can perform a wide variety of language analyses. The database and the CLAN programs are freely available to aphasia researchers and clinicians for educational, clinical, and scholarly uses. This article describes the database, suggests some ways in which clinicians and clinician researchers might find these materials useful, and introduces a new language analysis program, EVAL, designed to streamline the transcription and coding processes, while still producing an extensive and useful language profile.

KEYWORDS: Aphasia, discourse, CHAT, CLAN, Computerized Language ANalysis

Learning Outcomes: As a result of this activity, the reader will be able to (1) describe the general contents and uses of the AphasiaBank data, and (2) describe the functionality of the EVAL program.

AphasiaBank was designed to provide researchers and researcher-clinicians with a large shared multimedia database of uniform discourse samples from individuals with and without aphasia for the purpose of studying their communication. However, this database and the associated Computerized Language ANalysis (CLAN) programs can also serve as valuable resources for speech-language pathol-ogists (SLPs) whose major responsibilities are clinical. A new CLAN program, EVAL, has been designed for busy clinicians who wish to have an overview of their patients' language strengths and weaknesses and who need data to document the initial language status and the language changes of their patients.

This article will focus on: (1) describing the parts of the AphasiaBank database likely to be

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Address for correspondence and reprint requests: Margaret M. Forbes, M.A., Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 (e-mail: forbesmm@cmu.edu).

Technology and the Adult-Focused SLP; Guest Editor, Audrey L. Holland, Ph.D.

most useful to clinicians to support their various roles; (2) introducing CLAN, a versatile set of programs that perform multiple automatic language analyses on transcripts in CHAT format[1]; and (3) presenting a brief overview for using a new CLAN program, EVAL, designed with clinicians in mind. This program, created by Leonid Spektor, completes a basic set of language analyses on a transcribed language sample and displays the results in an Excel spreadsheet, side by side with the results of a comparison group from the AphasiaBank database. It can also display comparative results of a particular aphasic patient at two or more different times, allowing comparison of pre- and post-therapy performance.

We recognize that SLPs who work with communicatively and/or cognitively disordered adults in clinical settings are expected fulfill many roles. This article suggests some possible uses of materials from the AphasiaBank database that can be a useful adjunct to many of these responsibilities. These responsibilities may include, for example, clinical supervision and teaching, staff and family education, clinical research activities, and lecturing in classes, at Grand Rounds, and to community groups. We begin with a brief history of the AphasiaBank project, followed by a description of how to access and use the database.

HISTORY OF APHASIABANK

AphasiaBank was designed to extend the model established by the Child Language Data Exchange System (CHILDES) for the field of child language acquisition to include the study of adult language. The CHILDES project, originated and directed by Brian MacWhinney, is an international cooperative venture, involving some 800 active users and 4000 affiliated members located in over 30 countries. The establishment of CHILDES was initiated with a planning meeting of 20 major child language researchers in 1984 and has received National Institutes of Health-National Institute of Child and Human Development (NIH/NICHD) support since 1987. Users have access to an extensive database of child language, and to an array of CLAN programs that can automatically analyze language in multiple ways.

Many new empirical studies of child language production rely on the analysis of data from the CHILDES database, and the majority of theoretical papers on normal child language that include production data currently use the CHILDES database. A recent count shows that more than 3500 published articles are based on the use of CHILDES data or programs.

Work on establishing a system of this type for the study of aphasia started in 2005 with a planning meeting of 20 senior aphasia researchers. Meeting participants specified the shape of the AphasiaBank protocol and outlined methods for data sharing and possible computational analyses. The AphasiaBank grant, prepared by Brian MacWhinney and Audrey Holland, was funded by the National Institutes of Health-National Institute on Deafness and other Communication Discorders (NIH/NIDCD) in 2007. AphasiaBank initially focused solely on collecting discourse samples in English from individuals with aphasia and nonaphasic controls. More recently the database has been extended to add discourse samples from people with dementia (DementiaBank) and traumatic brain injury (TBIBank) and samples in other languages.

APHASIABANK DATABASE

The AphasiaBank database includes a variety of language samples in English, Spanish, German, Italian, Hungarian, Mandarin, and Chinese. It grows as new data are contributed. The database is password protected to protect the confidentiality of the aphasic participants, but is freely available to aphasia clinicians and researchers who contact Brian MacWhinney (macw@cmu.edu) to request membership.

This article focuses on the English section of the AphasiaBank database at TalkBank.org/AphasiaBank and on the recently developed TBIBank and DementiaBank. The English section of the AphasiaBank database is divided into three sections: Aphasia, Controls and NonProtocol, and Script. Script is a specialized database that will not be described here.

Aphasia

The Aphasia section of the database currently contains ~180 video recordings of people with

aphasia performing the AphasiaBank protocol, a uniform set of discourse tasks and aphasia tests. It also contains videos of ~140 nonaphasic control participants performing the AphasiaBank discourse tasks.

The protocol consists of four different discourse genres: personal narratives, picture descriptions, story retelling, and procedural discourse. Investigators use a script designed to keep the prompts consistent across investigators. The script includes a second level prompt with simplified questions to use with participants who do not respond within 10 seconds. The protocol is administered in a single session, which is recorded on video and generally takes an hour or less to complete.

The personal narratives are elicited by asking participants about their speech, their stroke or other neurological incident, their recovery, and an important event in their lives. For the picture descriptions, participants are shown three black and white drawings and asked to tell the stories they depict with a beginning, middle, and an end. The first picture stimulus is a four-paneled drawing of a child kicking a soccer ball and breaking a window, the second is a six-paneled drawing of a child refusing an umbrella and getting caught in the rain, and the third is the Nicholas and Brookshire picture showing the rescue of a cat from a tree.[2]

For the story telling task, participants are shown a picture book of Cinderella, with the words covered. They are asked to look through the book to remind them of the story. The book is then removed and they are asked to tell as much of the story as they can.

Finally, for procedural discourse, participants are asked to describe how they would make a peanut butter and jelly sandwich. (Test sites outside the United States may substitute another simple food preparation.) A photograph of peanut butter, bread, and jelly is available for participants who need further help.

In addition to the discourse tasks, the following tests are administered to aphasic participants: the Aphasia Quotient subtests of the Western Aphasia Battery—Revised[3]; the short form of the Boston Naming Test—Second Edition[4]; the Verb Naming Test from the Northwestern Assessment of Verbs and Sentences—Revised[5]; and the AphasiaBank Repetition test. The results of these tests are posted in an Excel spreadsheet at the AphasiaBank Web site (TalkBank.org/Aphasia-Bank, click Test Results Excel Sheet in the Protocol—Results column). Demographic data are also collected from participants, and the results are posted in an Excel spreadsheet at the AphasiaBank Web site (http://talkbank.org/AphasiaBank/demographics/, click on Current Demographic Database under the Aphasia column).

Control

The control database contains ~140 video recordings of nonaphasic people performing the AphasiaBank discourse tasks. In place of the stroke questions, these participants are asked about an illness or injury, their recovery from that illness or injury, and any experience they have had with people who have trouble communicating.

They are also given the Mini-Mental State Examination[6] and the Geriatric Depression Scale,[7] in addition to screening tests for hearing and vision. The results of these tests are posted at the Web site along with extensive demographic data (http://talkbank.org/AphasiaBank/demographics/, click on Current Demographic Database under the Controls column).

The discourse tasks for both aphasic and nonaphasic participants have been transcribed verbatim in CHAT format, and each transcript has been linked with its video, so that the videos can be viewed and heard, while the corresponding transcript is displayed, and each line is highlighted as it is spoken in the video. The language is also coded, using a coding system developed to capture characteristics of aphasic language (available from http://talkbank.org/AphasiaBank/, click error coding under the transcription column).

NonProtocol

The NonProtocol section of the English/AphasiaBank database contains samples of people with aphasia engaged in a variety of communicative tasks and interactions. Some are linked with video files, some with audio files only, and some are transcripts with no associated media.

This part of the database bears exploration by SLPs, as it contains a wide variety of materials illustrative of aphasic communication. A sampling of the materials in the NonProtocol section of the database will be described here. The database guide posted at the Web site (http://talkbank.org/AphasiaBank/ click on Database Guide in the Database column) provides additional information.

One useful teaching tool is a transcript in the Menn corpus, in which the aphasic person S.K. is engaged in a conversation intended to demonstrate that someone with language impairment can still be cognitively intact. The Mackie corpus contains language samples of a person in two different conversational settings and then during formal language testing, providing a good illustration of how aphasic language can vary according to the situation. The Fridriksson corpus includes videos of several aphasic people describing the picnic picture from the Western Aphasia Battery.

The Goodwin corpus consists of videos of conversations with aphasic people, transcribed and coded according to the conventions of conversation analysis, illustrating a way of capturing communication when actual speech is limited. Holland1 and Holland2 contain interviews with stroke victims in the early acute stage following their strokes and interviews with people past the acute stage who describe the experience of having a stroke, becoming aphasic, and adapting to a new and different life with aphasia.

Materials in the NonProtocol database are also useful for demonstrating skills such as effective interviewing or techniques of communicating with aphasic people with little or no spoken language. The language on these videos, like that on the aphasia and control videos, is transcribed in CHAT format and linked line by line with the video. Further uses for these materials will likely become evident to those who explore the database. Like the Aphasia section, this part of the database is rich with materials that SLPs can use in their various roles. This part of the database is expected to continue to grow in the future to include instructional videos concerning aphasia types, what to listen and look for in eliciting aphasic language, and other useful information related to aphasia.

TBIBANK AND DEMENTIABANK

Two other parts of the database are noteworthy. TBIBank and DementiaBank are relatively recent additions to the TalkBank database. They are accessible at talkbank.org/TBIBank and talkbank.org/DementiaBank, respectively. TBIBank is a funded project and is well underway, and plans to extend DementiaBank are currently in progress.

TBIBank (Principal Investigator, Leanne Togher) was funded in 2010 by the National Health and Medical Research Council in Australia, and is based in Sydney. This is a longitudinal study in which brain injured people are videoed at six different time points postinjury performing a uniform set of tasks similar to those AphasiaBank is collecting for aphasic people, with the goal of identifying recovery patterns. Once they are fully transcribed and coded, these data will be posted at TBIBank database. Currently, 12 brain injured individuals and their communication partners have completed their initial 3-month assessment, and four have completed the 6-month assessment.

For DementiaBank, pilot data are being collected for submission in a grant proposal for a full-scale project to collect discourse data from people with dementia. Again, the plan is to video participants performing a uniform set of tasks, and post the videos at the DementiaBank Web site, linked to their coded transcriptions. Currently, more than 500 cookie theft picture descriptions (transcripts and audio files) from people with dementia and nondemented control participants are available at the DementiaBank Web site.

APHASIABANK ACCESS

AphasiaBank members can access these data in a variety of ways. The simplest way is to open the TalkBank Web site, and click on Browsable Database. Detailed instructions will be displayed describing how to see and hear the videos, or hear the samples with audio only, while reading the linked transcripts. Video/audio samples from all of these materials can also be downloaded from the web via the TalkBank.org Web site to be used for education of students, colleagues, and patients' families.

Members can also use the TalkBank database for their clinical work and for clinical research. CLAN tools can be used to assess a patient's language, and to track progress in therapy. CLAN's language analysis programs can also be used to analyze multiple transcripts in a wide variety of ways. The various programs are described in full in the CLAN manual, which can be downloaded from the TalkBank.org/AphasiaBank Web site. In addition, an error coding system developed specifically to capture the kinds of errors seen in aphasia is available from the TalkBank.org/AphasiaBank Web site. Look for the error coding link in the column under transcription. In a recent publication, Macwhinney et al[8] illustrate the use of CLAN programs to study phonological, lexical, semantic, morphological, syntactic, temporal, prosodic, gestural, and discourse features of language. The CLAN glossary provides step-by-step instructions for running several useful CLAN programs. This document is also available from the TalkBank.org/AphasiaBank Web site.

EVAL

We turn now to EVAL, a new CLAN program designed to streamline the process of collecting and analyzing a language sample as much as possible, while still providing an extensive and clinically useful language profile. Detailed instructions for the various uses of this program are posted at TalkBank.org, click on EVAL manual.

When this program is run on a transcript in CHAT format, EVAL produces a language profile that includes number of utterances, mean length of utterance (MLU) in words and morphemes, type/token ratio (TTR), average number of clauses per utterance, number of word-level errors, number of utterance-level errors, number of repetitions, number of retracing/self-corrections, duration of sample, parts of speech (nouns, verbs, pronouns, prepositions, adverbs, adjectives, conjunctions, and determiners), verb tenses (third person singular present, past, progressive, and perfect), and plurals.

This program is a particularly valuable tool for SLPs who are under increasing pressure to demonstrate evidence-based treatment. EVAL automatically produces a language profile for one or more transcripts, and displays the profile in an Excel spreadsheet, side by side with the profile of a comparison group from the AphasiaBank database. Two or more transcripts of the same person at different times can also be compared, to assess progress.

To illustrate one clinical application of EVAL, we will use the example of an anomic patient admitted for therapy. We will compare his pre- and post-therapy performance of the procedural discourse task of describing how to make a peanut better and jelly sandwich. The

Table 1 EVAL Pre-therapy and Post-therapy Spreadsheet

File	Speaker ID
Post—eval01b.cha	eng|EVAL|PAR|60;1.|male|Anomic|eval01b|Participant||
Pre—eval01a.cha	eng|EVAL|PAR|59;11.|male|Anomic|eval01a|Participant||
	Database gems: Sandwich
	eval @ + t*PAR: + gSandwich + u

File	Duration	# Utts	MLU Words	MLU Morphemes	# Different Types	# Item Tokens
Post	0:00	3	11	11.667	21	33
Pre	0:00	5	4.4	4.4	17	22

File	TTR	Clause/Utt Ratio	Word Errors	Utt Errors	Nouns	Plurals	Verbs
Post	0.636	1.333	0	0	10	0	4
Pre	0.773	0.4	3	3	4	0	2

results of this comparison are displayed in Table 1.

The first line (eval01b) of the spreadsheet depicted in Table 1 represents the patient's post-therapy results, and the second line (eval01a) represents the results from his initial assessment before therapy. The spreadsheet is divided into three tiers in this table, but EVAL displays the analysis on the computer screen as a single spreadsheet, with the data for each person arrayed across a single line.

In this case, the participant produced five utterances when first evaluated (eval01a), and three at his second evaluation (eval01b) 2 months later. Since he produced more utterances at his initial evaluation, it at first appears that this participant did better at the first assessment than at the post-therapy assessment. But looking further on the spreadsheet, one sees that the participant has actually improved on all of the other language measures, including MLU in words and morphemes, TTR, and clauses per utterance, apparently producing more grammatically complex utterances than he had initially.

Other improvements in performance are evident: word and utterance errors have decreased, and the number of nouns and verbs has increased. Plurals in this sample remained unchanged.

As this example illustrates, EVAL can be a valuable tool for demonstrating change in therapy. While transcribing a language sample can be somewhat time-consuming, once the sample is transcribed it can provide a great deal of concrete information about language, and it can be analyzed in a variety of ways.

Only the EVAL program has been described here, and it is important to remember that more than 25 CLAN programs and an error coding system for aphasia are freely available to AphasiaBank users. In addition, users are welcome to devise their own codes, and to use CLAN programs to analyze them.

We have tried to indicate some of the many ways that practicing SLPs can use the AphasiaBank data and its language analysis tools. Recognizing the multiple roles that SLPs are expected to fulfill: clinician, supervisor, teacher, researcher, as well as the pressing demand for evidence-based practice, we hope that AphasiaBank can serve as a source of useful tools for language analysis and resources to support these varied activities.

REFERENCES

1. MacWhinney B. The CHILDES Project: Tools for Analyzing Talk. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2000
2. Nicholas LE, Brookshire RH. Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. J Speech Hear Res 1995;38 (1):145–156
3. Kertész A. Western Aphasia Battery revised. San Antonio: PsychCorp; 2007
4. Kaplan E, Goodglass H, Weintraub S. Boston Naming Test. 2nd ed. Austin, TX Pro-Ed.; 2001
5. Thompson CK. Northwestern Assessment of Verbs and Sentences—Revised. Evanston, IL: Northwestern University Press; 2011
6. Folstein M, Folstein S, Fanjiang G. Mini-mental State Examination. Lutz, FL: Psychological Assessment Resources, Inc.; 2002
7. Brink TL, Yesavage JA, Lum O, Heersema P, Adey MB, Rose TL. Screening tests for geriatric depression. Clin Gerontol 1982;1:37–44
8. MacWhinney B, Fromm D, Forbes M, Holland A. AphasiaBank: Methods for studying discourse. Aphasiology 2011;25:1286–1307