AphasiaBank as BigData

Brian MacWhinney, Ph.D.[1] and Davida Fromm, Ph.D.[1]

ABSTRACT

AphasiaBank has used a standardized protocol to collect narrative, procedural, personal, and descriptive discourse from 290 persons with aphasia, as well as 190 control participants. These data have been transcribed in the Codes for the Human Analysis of Transcripts (CHAT) format for analysis by the Computerized Language Analysis (CLAN) programs. Here, we review results from 45 studies based on these data that investigate aphasic productions in terms of these eight areas: discourse, grammar, lexicon, gesture, fluency, syndrome classification, social factors, and treatment effects. For each area, we also indicate how use of the CLAN programs has facilitated the analysis. We conclude with an examination of ways in which the size of the database could be increased through on-site recordings and data from teletherapy.

KEYWORDS: Databases, production, gesture, lexicon, syntax

Learning Outcomes: As a result of this activity, the reader will be able to (1) summarize the ways in which AphasiaBank data have been used in the research literature and (2) summarize the ways in which AphasiaBank data can be analyzed using CLAN.

The explosion of the Internet has fueled interest in configuring large bioinformatics databases. Such databases will encode information about genetics, life history, prior illness, treatments, allergies, test results, and many other medically important facts. Linkages across these various data types will help facilitate diagnosis and treatment. Moreover, searches across records for large numbers of patients can reveal protective and risk factors for certain diseases, as well as probabilities for the success of a given treatment. Although the construction of such databases must deal with privacy considerations and incompatibilities across computer systems, it is likely that medical databases will grow in size along with many other of the components of the BigData of the future.

BigData has also become increasingly important in the study of language usage. For example, systems such as SketchEngine (sketchengine.co.uk) provide easy methods for

[1]Carnegie Mellon University, Pittsburgh, Pennsylvania.

Address for correspondence: Brian MacWhinney, Ph.D., Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 (e-mail: macw@cmu.edu).

A Look to the Future: Big Data in Neurorehabilitation; Guest Editor, Yasmeen Faroqi-Shah, Ph.D., CCC-SLP

Semin Speech Lang 2016;37:10–22.
DOI: http://dx.doi.org/10.1055/s-0036-1571357.
ISSN 0734-0478.

searching specific linguistic patterns across written corpora composed of at least 10 billion words for major languages such as English, French, Chinese, Japanese, Russian, and Spanish. Search engines such as Google or Bing cast an even bigger net, although not in as much linguistic detail. For spoken language, as opposed to written language, corpora are much more modest in size, and many important corpora are not made publicly available, often because of Institutional Review Board and proprietary restrictions.

Although aphasia affects over a million people in the United States and many more people internationally, until recently there have been no publicly available corpora for the study of language in aphasia. It had seemed as if the rising tide of BigData would wash over the field of aphasiology with no impact. However, beginning in 2007, the AphasiaBank project at Carnegie Mellon University began to change this picture.[1] This project has now collected hour-long spoken language protocols from nearly 300 persons with aphasia (PWA) and nearly 200 age-matched normal controls. These videotaped protocols include standardized tests such as the Western Aphasia Battery–Revised (WAB-R) Aphasia Quotient (AQ) subtests and the Boston Naming Test (BNT),[2,3] along with picture description tasks, procedural discourse, storytelling, personal narrative, word and sentence repetition, and sentence comprehension. All of the protocols have been accurately transcribed in the CHAT system and linked temporally to the video data on the sentence level.[4] Using this linkage and password protections, these interactions can either be viewed in real time through a network browser or downloaded to a user's machine for later study and playback. For detailed transcript analysis, researchers rely primarily on the CLAN programs that are designed to operate on data transcribed in the CHAT format.[4]

Although AphasiaBank will never become BigData at the level of the corpora collected by Google or the National Security Agency, it certainly provides much bigger data than had been available prior to 2007, thereby allowing researchers, instructors, students, and speech-language pathologists to explore a wide range of new group-based comparisons that had not been possible earlier. In this article, we will summarize the results of 45 studies that used AphasiaBank data to explore eight research topics: discourse, grammar, gesture, lexicon, fluency, classification, social factors, and treatment effects. (A complete bibliography, available at http://talkbank.org/AphasiaBank/publications/, lists all published studies, book chapters, conference presentations and posters, and Master's and Doctoral theses that use or make reference to AphasiaBank materials and methods.) For each of these topic areas, we will first explain how CLAN analysis programs can be used to extract relevant findings from the database, and then we will summarize the results of the studies relevant to each topic.

DISCOURSE

Methods

To characterize patterns of discourse structure in aphasia, researchers have constructed measures based on propositional density, content information units (CIUs), and core lexicon (CoreLex). Propositional density counts are based on the conceptual analysis provided by Kintsch and Van Dijk,[5] as implemented computationally by Brown and colleagues in the Computerized Propositional Idea Density Rater (CPDIR) program.[6] We have reimplemented that analysis inside CLAN and this reimplementation matches up nearly perfectly with the CPDIR of Brown et al. Another approach to discourse structure looks at whether or not picture descriptions and storytellings include a set of core narrative targets. This analysis is computed by using the CLAN FREQ programs to search for these core words. Finally, CLAN can assess discourse structure through automatic computation of mean length of utterance (MLU) and mean length of turn. Additional discourse measures are provided by CLAN's KEYMAP, CHAINS, and CHIP programs. For all analyses, the GEM programs can be used to isolate and analyze previously tagged selections or "gems" (e.g., the procedural discourse task, a particular picture description task) within the transcript.

Studies

Fergadiotis and Wright analyzed data from 98 monolingual PWA from the AphasiaBank database to study the relationship between confrontation naming test scores and connected speech.[7] The former is a reliable, valid, and commonly used tool in aphasia assessment. However, discourse is more challenging to measure and analyze in a systematic, reliable, and time efficient manner. Yet, it is what PWA and their families regard as most important. For confrontation naming, the authors used scores from the WAB-R naming subtest,[2] the BNT, and the Verb Naming Test (VNT).[8] For discourse, the authors used discourse from free speech tasks, three picture description tasks, and the Cinderella storytelling task. They used structural equation modeling to predict the number of paraphasias produced and the amount of information conveyed (in CIUs) for the Cinderella storytelling. Results showed that confrontation-naming test scores were not good predictors of the proportion of paraphasias to expect in connected speech. Though the naming scores were found to be strongly related to informativeness on the Cinderella task, about half of the variance in the CIU measurement was unaccounted for, suggesting that to obtain important information about discourse requires evaluation of connected speech.

Dalton and Richardson measured informativeness in discourse by analyzing CoreLex and main concepts (MCs) in transcripts from 166 controls and 235 PWA doing one of the AphasiaBank picture description tasks (Broken Window).[9] Stems (or lemmas) produced by 50% or more of the controls were included in the CoreLex. Results indicated that PWA had significantly lower CoreLex scores and MCs than controls. CoreLex scores were significantly different for controls and each of the aphasia subtypes; MCs scores were significantly different for controls versus all aphasia subtypes except for those who scored above the WAB-R AQ cutoff but were still aphasic (NABW, or not aphasic by WAB). Both CoreLex and MCs scores were able to distinguish between the Broca group and all other aphasia subtypes (anomic, conduction, Wernicke, NABW). CoreLex and MCs scores were significantly and strongly correlated for all groups. The techniques used in this study demonstrate how lexical and concept-level discourse can be measured to evaluate and characterize language in aphasia. PWA used less typical language than did controls and the language was more variable across participants, probably reflecting the discourse impairments that result from lexical access and/or retrieval difficulties. These authors also urge multilevel discourse analysis for more functional and ecologically valid clinical intervention.

Cinderella stories from the AphasiaBank database were analyzed by Richardson et al to study story grammar and discourse measures specifically in anomic aphasia.[10] They matched 14 PWA and 14 nonaphasic participants. Interestingly, though all PWA had performed at or near ceiling on the WAB-R AQ, their Cinderella stories had a significantly reduced number and type of story components. In comparison with the nonaphasic participants, the stories from the PWA were shorter and less coherent. For the PWA group, the correlation between production of the story's CoreLex and the story length was significant, suggesting that word-finding deficits contribute to the observed reduction in story components.

Procedural discourse is assessed in the AphasiaBank protocol by asking participants to tell the investigator how they would make a peanut butter and jelly sandwich. In an analysis of responses from 141 PWA and 144 nonaphasic participants, Fromm et al reported large group differences in quantitative measures but minor group differences in qualitative and functional measures.[11] Compared with nonaphasic participants, the PWAs had a lower average MLU and produced significantly fewer words, fewer utterances, and shorter utterances. However, the top 10 nouns and verbs used by both groups were nearly identical.

Using AphasiaBank data, Capilouto and colleagues compared discourse in older and younger neurologically intact adults and found that the older adults had less elaborate discourse.[12] Four picture stimuli were used to elicit storytellings. Results indicated that age has an influence on discourse abilities in that the narratives of older healthy adults contained a significantly lower percentage of CIUs. The groups did not differ on words per minute, CIUs per minute, or proportion of main events mentioned. However, additional age-related differences were found in correlation analyses between CIU measures and proportion of main events. This study serves to highlight important considerations about the effect of age on discourse measures and different types of structured discourse tasks.

GRAMMAR

Methods

The automatic computation of grammar patterns in CLAN relies on the fact that transcripts correctly transcribed in CHAT can be automatically tagged for morphological structure using CLAN's MOR command and for syntactic structure using CLAN's MEGRASP command.[13] Once this is done, CLAN can also provide automatic computation of developmental sentence score (DSS) and the index of productive syntax.[14,15] In addition, the EVAL program can compare individual participants with larger groups in terms of *z*-scores across dozens of grammatical structures.[16] EVAL is a CLAN program developed specifically for clinicians and clinical researchers to use for simple and efficient analysis of connected speech. The outcome measures include duration, total utterances, MLU, types, tokens, type–token ratio (TTR), proposition density, errors, total words, total or percentage of nine parts of speech, and number of revisions and repetitions. Data are summarized in a spreadsheet and can be used to track changes over time or compare an individual's performance with a selected reference group from the AphasiaBank database. EVAL allows for automated computer analysis of streamlined transcriptions and provides a valuable snapshot of language use.

Studies

Llinàs-Grau and Martínez-Ferreiro conducted a focused syntactic analysis of the complementizer *that* in language samples of 100 PWA and 100 nonaphasic participants from the AphasiaBank database.[17] They identified cues (occurrences of *say*, *think*, and *know*) in the transcripts as triggers for *that.* Results showed that PWA produced significantly fewer subordinate clauses than nonaphasic participants and those with fluent aphasias produced subordinate structures significantly more than did those with nonfluent aphasias. However, both PWA and nonaphasic groups had similar patterns in the presence and absence of *that* across the three verb cues. The authors contend that the presence of *that* is not directly related to syntax and therefore would not necessarily be different in aphasia.

A study by den Ouden et al explored verb argument structure complexity effects in verb choice in the Cinderella storytelling task.[18,19] Analyzing samples from 173 PWA and 159 nonaphasic participants, the authors found no differences in verb argument structure properties in verbs used by the two groups but they reported that PWA used more frequent verbs and shorter verbs. This study is reported in its entirety in this issue of *Seminars in Speech and Language*.

In another study of the production of nouns and verbs in connected speech with performance on confrontation naming tests, Johnson et al analyzed AphasiaBank data from 142 nonaphasic participants.[20] The number of nouns correctly produced by all participants in the PWA group across all tasks was strongly correlated with BNT scores. Correct verbs, on the other hand, correlated with VNT scores for all participants on the Cinderella task and for those with Broca's aphasia on two of the picture description tasks. The high incidence of weak verbs and modals/auxiliaries in storytelling do not correspond as well with the verbs elicited in formal testing.

A cross-linguistic study of grammatical aspects of aphasia was conducted by Sung and coworkers.[21] They used the AphasiaBank database and protocol to compare English-speaking and Korean-speaking participants with Broca's and anomic aphasia on a picture description task. Specifically, their aim was to determine if the use of verbs and nouns differs across language groups and aphasia types. Language samples from the Cat Rescue picture stimulus from 10 nonaphasic speakers and 21 PWA from each language. The Korean- and English-speaking PWA were matched by aphasia type and severity. Results showed that Korean PWA produced more verbs and English PWA produced more nouns and pronouns. The English-speaking PWA had significantly higher noun-to-verb ratios than did the Korean-speaking PWA. No significant differences in noun-to-verb ratio was found on the basis of aphasia type, though the difference between anomic and Broca's aphasia on the verbs per utterance outcome measure was greater in PWA who spoke Korean than in those who spoke English. The nonaphasic English speakers generated significantly more utterances and more nouns than did the nonaphasic Korean speakers, and the Korean speakers generated significantly more verbs per utterance than did the English speakers. Further cross-linguistic investigations of this sort are encouraged.

Finally, Thorne and Faroqi-Shah (this issue) examined verbs in 164 PWA and 166 nonaphasic participants from the AphasiaBank database.[22] They found that the proportion of light verbs used was not significantly different between groups. However, in the aphasia group, a higher proportion of light verbs was associated with poorer semantic scores (idea density and VNT score), better syntactic scores (DSS), and a negative difference score (standardized idea density minus standardized DSS, indicating greater semantic impairment). The authors recommend considering syntactic ability as a continuous variable rather than two distinct groups (agrammatic versus nonagrammatic).

GESTURE

Methods

For the study of gesture, researchers have typically processed AphasiaBank files using the ELAN program developed by the Max-Planck Institute for Psycholinguistics in Nijmegen. This type of analysis can be facilitated by the use of the CHAT2ELAN and ELAN2-CHAT commands in CLAN that work to convert to and from ELAN. ELAN is an excellent program for gesture analysis. However, there are also powerful methods for gesture analysis in CLAN, based on linkage between the main transcript and subfiles for deeper analysis. We hope that future work with gesture makes wider use of some of these facilities.

Studies

Using the transcripts and videos from 98 PWA and 64 nonaphasic participants from the AphasiaBank database, Sekine and Rose analyzed the use of gesture in Cinderella storytellings.[23] Results revealed that in comparison to non-aphasic participants, more PWAs gestured and the predominant PWAs gestured more type was iconic. Furthermore, aphasia type was found to influence the types of gestures used. For example, significantly higher proportions of participants with Broca's and conduction aphasia used iconic character viewpoint gestures, and a significantly higher proportion of participants with Broca's aphasia used pantomime and number gestures. The authors recommend targeting particular types of gestures in aphasia therapy to improve communication. These findings corroborated and added to those from a previous study done with fewer participants by Sekine and colleagues.[24]

Jenkins et al also analyzed the Cinderella storytellings of 29 PWA and 29 age- and gender-matched nonaphasic participants from the AphasiaBank database to catalogue gesture types determine if gesture frequency was associated with more complex and better organized narratives.[25] They found that PWA produced significantly shorter narratives and more gestures than the nonaphasic group, but no significant correlations were found between gesture frequency and sentence complexity or narrative organization for both groups. In-depth analysis of a smaller subgroup of the participants revealed that over 40% of the PWA gestures were classified as lexical retrieval (those that accompanied a speech dysfluency or word-recall attempt) or other (shape/function not clear). The authors concluded that gesture frequency was not associated with better discourse production, using number of subordinated clauses within all matrix clauses as the outcome measure for sentence complexity and number of complete episodes and the measure for narrative organization. Both this study and those conducted by Sekine and colleagues mention the need to add gesture categories to previously defined coding systems and to modify some gesture category definitions to account for all of the gestures made by those in the aphasia group.[23,24]

Kong et al introduced a coding system called Database of Speech and Gesture for coding and quantifying gestures during connected speech.[26,27] They collected data from 119 nonaphasic native speakers of Cantonese in Hong Kong in three age groups: young, middle-aged, and senior. The AphasiaBank protocol was followed with adaptation for Chinese culture. Language transcripts and videos were linked using ELAN. The authors stress the importance of normative data and clearly differentiated coding systems for form and function of gestures.

The Database of Speech and Gesture system was used by Kong et al in a study of coverbal gestures in Cantonese speakers using the AphasiaBank protocol with 48 PWA and 48 nonaphasic participants.[28] As in previously reviewed studies of native English speakers, the aphasia group used significantly more gestures per word. Compared with the nonaphasic participants whose gestures were mostly for enhancing their language, the PWA used a higher proportion of content-carrying gestures. Aphasia severity was negatively correlated with frequency of gesture use, indicating that more severely impaired participants used more gestures. Again, detailed descriptions of gesture forms and functions provide important information for clinical assessment and remediation.

LEXICON

Methods

CLAN provides tools for three types of lexical analyses. The first analysis relies on standard frequency counting using the FREQ program. This counting can be performed either on whole words as transcribed on the main line or on morphological stems as analyzed on the %mor tier. A second type of analysis examines lexical diversity as measured through type-token ratio (TTR), vocabulary diversity (vocD),[29] or moving average type-token ration (MATTR).[30]

Studies

The effect of discourse type on lexical diversity and productive vocabulary was investigated by Fergadotis and Wright using 27 PWA and 27 age- and education-matched nonaphasic participants from the AphasiaBank database.[31,32] Lexical diversity was quantified using a lemma-based analysis of content words as well as a computational tool that reduced the impact of sample length and genre. Results showed that lexical diversity was significantly lower in PWA for all discourse tasks. Furthermore, a significant group by task interaction was found. In the nonaphasic group, lexical diversity scores were significantly different across all three discourse tasks but in the aphasia group, lexical diversity scores were only significantly different for two of the tasks (single pictures and sequential pictures, single pictures and storytelling). For the nonaphasic group, the highest lexical diversity scores occurred in the storytelling task and the lowest in the single picture description task. For the PWA group, the highest scores occurred with sequential pictures and the lowest with single pictures.

To assess validity evidence for lexical diversity measurement, Fergadiotis et al compared four measures to determine how effective they were at measuring lexical diversity in PWA using AphasiaBank data from 101 participants' Cinderella storytellings.[33] Results from a variety of statistical techniques and modeling with confirmatory factor analyses suggested that the measure of textual lexical diversity (MTLD) and the MATTR were the best measures for capturing lexical diversity in aphasia.[30,34,35] No difference was found between vocD and hypergeometric distribution,[34] but these measures require samples of at least 50 tokens and may introduce bias under certain conditions. This kind of careful analysis of measurement tools is invaluable to the study of discourse in aphasia. As a result of this study, MATTR was added to the CLAN program.

In an investigation of language from participants with posttraumatic stress disorder (PTSD), Caglar used 10 free speech samples of Broca's aphasia from the AphasiaBank database for comparison purposes based on the fact that regional cerebral blood flow to Broca's area is reduced in individuals with PTSD when talking about their trauma.[36] The Zipfian distribution (in which the frequency of a word is inversely proportional to its frequency ranking) was used to compare the groups. All participants' language samples conformed to Zipf's law. However, the slope of the Zipfian distribution was significantly different for the aphasia group. The PWA used fewer words than the PTSD group and used them more frequently. Several possible explanations for the results were posited but no clear conclusions could be reached.

Law et al also used AphasiaBank data to explore the issue of lexical retrieval of nouns and verbs in connected speech versus confrontation naming type tasks.[37] In this study of 19 Chinese speakers with anomic aphasia and 19 nonaphasic Chinese speakers, the authors matched the target stimuli between word classes (nouns and verbs) and tasks (picture naming and narrative) on potentially confounding psycholinguistic variables (age of acquisition and familiarity). Not surprisingly, PWA had more word-finding problems than did the nonaphasic group. Their performance was affected by both word class (better with nouns) and task (better with picture naming) when age of acquisition and familiarity were balanced. When imageability was also controlled, the only significant effect was task. The authors of this study echo the importance of analyzing connected speech in language assessment and treatment evaluation.

For assessment of communication in severe aphasia, the AphasiaBank discourse protocol was not an appropriate or productive tool. In response to that need, Holland et al developed the Famous People Protocol.[38,39] The protocol includes photographs of 24 famous people from sports, entertainment, and politics. It allows individuals with severe aphasia opportunities to demonstrate their competency in conveying real-world knowledge. In addition, clinicians and families can learn what strategies aphasic individuals use to communicate. Seventy-five PWA and 28 nonaphasic participants have been tested with the Famous People Protocol. The instrument does not seem to show age or education bias, and it has yielded a range of scores from 20 to 99 (maximum = 100) in the PWA group and 89 to 100 in the nonaphasic comparison group. Transcriptions of these evaluations have yet to be done to mine these videos for what can be learned about natural communication in severe aphasia.

FLUENCY

Methods

CHAT transcription provides a variety of special markers designed to encode and analyze dysfluencies,[40] including fillers, initial repetition, blocking, internal pausing, drawling, word repetition, and retracing. Utterance internal and external pause time can be computed using the TIMEDUR program. Currently, pause length must be marked manually. However, we are developing word-level alignment methods that will allow for automatic computation of pause duration. For acoustic analysis of pauses and dysfluencies, transcripts in CHAT can be automatically converted to the format needed for the Phon program,[41] which links tightly with acoustic analysis in Praat.[42]

Studies

Basilakos et al used AphasiaBank data to evaluate a fluency measure in a study on the association between regional white matter damage and poststroke fluency.[43] The authors wanted to use the WAB-R fluency scale score but needed to confirm its relationship to words per minute, which is the measure used by other related studies. A significant correlation ($r = 0.546$) was found between the WAB-R fluency subtest score and words per minute calculated from 247 language samples in the AphasiaBank database, showing that words per minute and WAB-R were related but not completely redundant.

Hird et al used 28 Cinderella storytellings from AphasiaBank participants to demonstrate the sensitivity of an automated fluency measure called the fluency profiling system designed to quantify fluency in natural spoken discourse.[44] Specifically, the fluency profiling system determines the locations of speech–pause transitions and estimates the temporal characteristics of speech and pause distributions in connected speech. The authors selected 18 participants with Broca's aphasia and 10 nonaphasic participants whose sample met an empirically defined signal-to-noise ratio. Results showed that all Broca participants had shorter mean speech segment durations than nonaphasic participants, and half of the Broca participants had significantly longer mean short pause durations. They conclude that the fluency profiling system is a sensitive tool for profiling cognitive and motor processes in aphasia.

In a study of Cantonese speakers using the AphasiaBank protocol, Lee et al measured the duration of content and function words in discourse in 17 participants with fluent aphasia and 17 nonaphasic speakers.[45] They used automatic time alignment on the audio signals to measure the beginning and end time of all subsyllable units in the utterance. Average syllable duration and average duration per word were greater for those with aphasia. Syllable and word durations were longer for content words than function words but the difference was smaller in the aphasia group than in the control group. The authors explain the longer word duration in aphasic speech as a possible result of impaired lexical retrieval.

CLASSIFICATION

Methods

Since the seminal studies of Broca and Wernicke,[46,47] aphasiologists have sought to link patterns of brain injury to patterns of language deficit. Because we do not currently have access to magnetic resonance imaging brain scans of our participants from the chronic period, we cannot work directly on this issue. However, we are able to use statistical methods to characterize the ways in which aphasic symptoms cluster behaviorally. These analyses rely on the indices computed by the EVAL command in CLAN, which provides output in spreadsheet format for further analysis through the R programming language (www.r-project.org).

Studies

MacWhinney applied k-means clustering to data from 161 individuals in the AphasiaBank database.[48] The highest loading variables in this analysis were for repetition, naming, comprehension, and fluency. Clusters were designed to maximize values on the adjusted Rand index and Fowlkes-Mallows index. Using these criteria, five clusters emerged:

1. Fluent with mild anomia (24 anomic, 5 nonaphasic, 5 transcortical motor)
2. Fluent with severe anomia (25 anomic, 3 Broca, 16 conduction, 5 Wernicke)
3. Fluent with severe anomia and poor repetition (3 Broca, 6 conduction, 1 transcortical motor, 6 Wernicke)
4. Dysfluent (11 Broca, 9 conduction, 1 transcortical sensory)
5. Dysfluent with comprehension deficits (20 Broca, 1 conduction, 3 global, 1 Wernicke)

Together, these results indicate a division of Broca into a group without serious comprehension deficits and another group with serious deficits.

SOCIAL FACTORS

Methods

Studies of the role of social factors on language in aphasia can make use of all the tools we have discussed for studying lexicon, grammar, discourse, gesture, and fluency.

Studies

To investigate the influence of educational background and occupational status on the spoken language of persons with Broca's aphasia, Johansson conducted morphological, semantic, phonological, and lexical analyses on the discourse samples from six AphasiaBank participants, three of whom were university graduates and three of whom were high school graduates.[49] This research project, completed as part of a master's degree program in linguistics, showed that university graduates used a greater number of grammatical categories, more words, and longer average sentences than did the high school graduates. High school graduates made at least five times more semantic errors and at least three times more phonological errors and university graduates made 30% more neologistic errors. Johansson concluded that spoken language in persons with Broca's aphasia is different based on educational and occupational background.

Fromm et al sought to address the concern that gender and or age may be associated with participants' responses to the Cinderella storytelling task.[50] Linguistic analyses were conducted on the stories produced by 98 PWA and 98 nonaphasic participants. Within the PWA group, males and females did not differ on total number of utterances and total words. Females, however, showed greater lexical diversity than males. Age was weakly and inversely associated with total utterances. In the nonaphasic group, males and females did not differ on any of the outcome measures, and age was weakly and inversely associated with total words. The authors concluded that the Cinderella remains a useful tool for evaluation of discourse in aphasia.

Campbell et al asked experienced and naive observers to watch nine AphasiaBank video clips of speakers with different types and severities of aphasia in a project that related discourse measures to observer perceptions of typicality and comfort.[51] After watching, they rated their perceptions about how normal the speech seemed and how comfortable they would be listening to the person, having a conversation with the person, or socializing with the person. Results suggested that MLU, CoreLex, MCs, and WAB-R AQ scores were all significantly and positively correlated with the observers' responses; TTR was not significantly correlated to either. In addition, experienced observers were more sensitive to deviations from normal in the PWA but they also had higher comfort ratings. Finding relationships between discourse outcome measures in aphasia and perceptions of those who interact with PWA can have an influence on functional treatment strategies that will help PWA participate more fully in a variety of life roles.

Hancock et al used 30 nonaphasic language samples from the AphasiaBank database to study perceptions of gender and femininity.[52] Written samples of their descriptions of an illness or injury were shown to 40 readers who were asked to decide if the speaker was male or female. Gender-indicative content (e.g., husband, wife) was removed from the samples beforehand. Results showed that gender prediction accuracy based solely on the language was near chance. No strong evidence was found for language differences between males and females. Fourteen linguistic variables were analyzed to compare the male and female language samples. Though some of the variables were found to be somewhat more predictive than others, none were found to be powerful predictors of how people perceive gender or femininity.

The AphasiaBank protocol was used to study cultural language variations in older (50 to 59 years) and younger (20 to 29 years) adults from the Appalachian region.[53] No significant differences were found in story grammar scores or global coherence scores between older and younger groups, though the younger group did demonstrate significantly higher local coherence scores than the older group. Story grammar scores were not correlated with local or global coherence for either group. The groups did not differ in total number of words but a significant correlation was found between story grammar and total number of words for both groups. This study emphasizes the importance of documenting linguistic variations across regions and cultures.

TREATMENT EFFECTS

Methods

Researchers are now paying increasing attention to using AphasiaBank methods for the study of treatment effects. Traditionally, these effects have been measured through performance on standardized tests such as the WAB. However, such measures fail to evaluate the functional use of language in conversation, narration, and description. Also, without detailed transcripts linked to audio, one cannot evaluate the impact of treatment on gesture and fluency. Current efforts to devise a core outcome set of measures for assessing treatment effects confront a variety of methodological challenges.[54,55] Because AphasiaBank methods focus on the functional use of language AphasiaBank methods, they can play a big role in this effort.

Studies

The AphasiaBank database includes other corpora in addition to those based on the standardized discourse protocol. For example, two corpora in the script collection include transcripts and media for PWA who have gone through treatment programs using script training. In response to interest expressed by researchers wanting to contribute these data to AphasiaBank and use CLAN for analysis, we developed a new CLAN program called SCRIPT. The program compares a participant's performance to the model script and measures duration, total words produced, number and percent words correct, number and percent words omitted, number of words added, number of recognizable errors, number of unrecognizable errors, number of utterances with unintelligible segments, and number of utterances with no output.

Fridriksson et al tested 13 individuals with Broca's aphasia using scripts in three conditions: audiovisual feedback, audio-only feedback, and spontaneous speech on the assigned topics.[56] PWA generated a greater variety of words in the audiovisual feedback condition. Functional magnetic resonance imaging demonstrated greater bilateral cortical activation for speech produced during the audiovisual condition. After a 6-week treatment program using the technique of speech entrainment, or mimicking audiovisual stimuli, participants were able to produce a greater variety of words (greater than 60% more) for trained and untrained scripts with and without the audiovisual stimuli at 1- and 6-weeks posttraining. Furthermore, neuroimaging results demonstrated treatment-related decreases in cortical activation in areas of the left posterior-inferior parietal lobe. The authors discuss the role this area of the brain may play in compensating for impaired gating in Broca's area. Further testing of treatment approaches with both automated linguistic analyses and neuroimaging provides valuable insights into the actual mechanisms underlying the behavioral changes.

Script training and CLAN's script analysis program were also used for 14 PWA at the Adler Aphasia Center. Their transcripts and media were also contributed to the AphasiaBank database and analyzed for a conference presentation by Szabo et al.[57] Clinicians customized the scripts, cues, and training hierarchy for each participant. The treatment was found to be successful for four participants, unsuccessful for three, and borderline for seven. Success was associated with increased verbal productions, increased confidence, increased independence, and opportunities to engage with others. Those who were unsuccessful had difficulty mastering their scripts.

Boyle examined test–retest stability of word retrieval errors in 10 PWA (5 with anomic aphasia, 3 with conduction aphasia, 2 with Broca's aphasia and mild apraxia of speech) from the AphasiaBank database.[58] The picture and storytelling discourse samples were elicited in two sessions separated by 2 to 7 days. She reports that when the discourse samples are combined, the following measures have good test–retest reliability: number of phonological errors per minute, semantic errors per minute, time fillers per minute, and repetitions per minute.

Holland et al (personal communication) analyzed data on 26 AphasiaBank participants who were tested on at least two occasions, separated by at least 1 year. All had completed formal speech therapy but were active participants in ongoing activities of aphasia centers. Surprisingly, 16 individuals demonstrated improved WAB-R AQ scores (positive AQ change scores greater than the WAB-R AQ standard error of the mean), seven were basically unchanged, and three worsened. Concurrent changes on several AphasiaBank tasks were also found, suggesting that the WAB-R improvements were noted in more natural discourse as well.

CONCLUSION AND PROSPECTS

Although AphasiaBank provides aphasiology with much bigger data than was available in the past, it has not yet provided truly BigData. However, there are efforts underway to bridge this gap. For the study of child language, the newly funded HomeBank project (homebank.talkbank.org) shows how researchers can begin to access 24-hour-a-day recordings of parents and their children in the home context. Similar methods could be developed for studying PWA. Also, the advent of teletherapy through systems such as Constant Therapy (Kiran, this issue) promise to provide still wider coverage of the ways in which PWA are able to manage successful recovery. Together, our current methods, along with those under development, will eventually provide us with truly BigData for the study of language in aphasia.

DISCLOSURES
Brian MacWhinney: no conflicts
Davida Fromm: no conflicts

ACKNOWLEDGMENTS
This work was supported by grant DC0008524 from the National Institutes of Health (National Institute on Deafness and Other Communication Disorders).

REFERENCES

1. MacWhinney B, Fromm D, Holland A, Forbes M. AphasiaBank: data and methods. In: Müller N, Ball M eds. Research Methods in Clinical Linguistics and Phonetics. New York, NY: Wiley; 2012: 268–287
2. Kertész A. Western Aphasia Battery. San Antonio, TX: PsychCorp; 2007
3. Kaplan E, Goodglass H, Weintraub S. Boston Naming Test, 2nd ed. Austin, TX: Pro-Ed; 2001
4. MacWhinney B. The CHILDES Project: Tools for Analyzing Talk, 3rd ed. Mahwah, NJ: Lawrence Erlbaum; 2000
5. Kintsch W, Van Dijk T. Toward a model of text comprehension and production. Psychol Rev 1978; 85:363–394
6. Brown C, Snodgrass T, Kemper SJ, Herman R, Covington MA. Automatic measurement of propositional idea density from part-of-speech tagging. Behav Res Methods 2008;40(2):540–54
7. Fergadiotis G, Wright HH, Green SB. Psychometric evaluation of lexical diversity indices: assessing length effects. J Speech Lang Hear Res 2015; 58(3):840–852
8. Cho-Reyes S, Thompson CK. Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). Aphasiology 2012;26(10):1250–1277
9. Dalton SG, Richardson JD. Core-lexicon and main-concept production during picture sequence description in non-brain-damaged adults and adults with aphasia. Am J Speech Lang Pathol 2015;24:923–938
10. Richardson J, Hudspeth SG, Saunders W, Payne V. Story narrative in anomic aphasia: analysis of story grammar components. Paper presented at: American Speech-Language-Hearing Association Convention; 2013; Chicago, IL
11. Fromm D, Forbes M, Holland A, MacWhinney B. PWAs and PBJs: Language for describing a simple procedure. Paper presented at: Clinical Aphasiology Conference; 2013; Tucson, AZ
12. Capilouto G, Wright HH, Wagovich SA. CIU and main event analyses of the structured discourse of older and younger adults. J Commun Disord 2005; 38(6):431–444
13. Sagae K, Davis E, Lavie A, MacWhinney B, Wintner S. High-accuracy annotation and parsing of CHILDES transcripts. In: Proceedings of the 45th Meeting of the Association for Computational Linguistics. Prague, Czech Republic: ACL; 2007:1044–1050
14. Lee LL, Canter SM. Developmental sentence scoring: a clinical procedure for estimating syntactic development in children's spontaneous speech. J Speech Hear Disord 1971;36(3): 315–340
15. Scarborough HS. Index of productive syntax. Appl Psycholinguist 1990;11:1–22
16. Forbes M, Fromm D, Holland A, MacWhinney B. EVAL: A tool for clinicians from AphasiaBank. Paper presented at: Clinical Aphasiology Conference; 2014; St. Simons Island, GA
17. Llinàs-Grau M, Martínez-Ferreiro S. On the presence and absence of "that" in aphasia. Aphasiology 2013;28:62–81
18. den Ouden D, Malyutina S, Richardson J. Verb Argument Structure in Narrative Speech: Mining the AphasiaBank. Tucson, AZ: Academy of Aphasia; 2015
19. Malyutina S, Richardson JD, den Ouden D. Verb Argument Structure Complexity in Discourse in Aphasia. Orlando, FL: ASHA; 2014
20. Johnson K, Kurland J, Parker J, Fromm D, MacWhinney B. Nouns and Verbs in Naming and Storytelling Tasks in Aphasia: Verbs Are Another Story. Atlanta, GA: ASHA; 2012
21. Sung JE, Lee SE, DeDe G. Crosslinguistic differences in a picture description task between Korean and English speaking individuals with aphasia. Paper presented at: Clinical Aphasiology Conference; 2015; Monterey, CA
22. Thorne J, Faroqi-Shah Y. Testing the Division of Labor Hypothesis of Aphasic Verb Production Using Big Data. Tucson, AZ: Academy of Aphasia; 2015
23. Sekine K, Rose ML. The relationship of aphasia type and gesture production in people with aphasia. Am J Speech Lang Pathol 2013;22(4):662–672
24. Sekine K, Rose M, Foster A, Attard M, Lanyon L. Gesture production patterns in aphasic discourse: in-depth description and preliminary predictions. Aphasiology 2013;27:1031–1049
25. Jenkins T, Coppola M, Coelho C. Gesture frequency and discourse quality in aphasia. Paper presented at: Clinical Aphasiology Conference; 2015; Monterey, CA
26. Kong A, Law SP, Kwan C, Lai C, Lam V. Database of speech and gesture (DOSAGE): a database of forms and functions of gesture use during verbal communication. J Nonverbal Behav 2015;39:93–111

27. Kong A, Law SP, Kwan CC, Lai C, Lam V. A coding system with independent annotations of gesture forms and functions during verbal communication: development of a database of speech and GEsture (DoSaGE). J Nonverbal Behav 2014
28. Kong AP, Law SP, Wat W, Lai C. Employment of gestures in spontaneous verbal discourse by speakers with aphasia. Procedia Soc Behav Sci 2013;94;
29. Malvern D, Richards B, Chipere N, Purán P. Lexical Diversity and Language Development. New York, NY: Palgrave Macmillan; 2004
30. Covington MA, McFall JD. Cutting the Gordian knot: the moving-average type–token ratio (MATTR). J Quant Linguist 2010;17(2):94–100
31. Fergadiotis G, Wright HH. Lexical diversity for adults with and without aphasia across discourse elicitation tasks. Aphasiology 2011;25(11): 1414–1430
32. Fergadiotis G, Wright H. Modelling confrontation naming and discourse performance in aphasia. Aphasiology, 1–17. Aphasiology 2015;29:1–17
33. Fergadiotis G, Wright HH, West TM. Measuring lexical diversity in narrative discourse of people with aphasia. Am J Speech Lang Pathol 2013; 22(2):S397–S408
34. McCarthy PM, Jarvis S. Voc-D: a theoretical and empirical evaluation. Lang Test 2007; 24:459–488
35. Covington M. MATTR User Manual (CASPR Research Report 2007–05). Athens, GA: University of Georgia; 2007
36. Caglar L. Traumatized Broca's Area: A Linguistic Analysis of Speech in Posttraumatic Stress Disorder. Utrecht, The Netherlands: Humanities, University of Utrecht; 2012
37. Law SP, Kong AP, Lai LW, Lai C. Effects of context and word class on lexical retrieval in Chinese speakers with anomic aphasia. Aphasiology 2015;29(1):81–100
38. Holland A, Fromm D, Forbes M, MacWhinney B. Developing a New Procedure for Assessing Communication in Severe Aphasia: The Famous People Protocol. Lucerne, Switzerland: Academy of Aphasia; 2013
39. Holland A, Fromm D, Forbes M, MacWhinney B. The Famous People Protocol. Paper presented at: International Aphasia Rehabilitation Conference; 2014; The Hague, The Netherlands
40. Bernstein Ratner N, Rooney B, MacWhinney B. Analysis of stuttering using CHILDES and CLAN. Clin Linguist Phon 1996;10(3): 169–188
41. Rose Y, MacWhinney B. The PhonBank Project: data and software-assisted methods for the study of phonology and phonological development. In: Durand J, Gut U, Kristoffersen G eds. The Oxford Handbook of Corpus Phonology. Oxford, UK: Oxford University Press; 2014:380–401
42. Boersma P, Weenink D. Praat, a System for Doing Phonetics by Computer. Amsterdam, The Netherlands: Institute of Phonetic Sciences of the University of Amsterdam; 1996
43. Basilakos A, Fillmore PT, Rorden C, Guo D, Bonilha L, Fridriksson J. Regional white matter damage predicts speech fluency in chronic post-stroke aphasia. Front Hum Neurosci 2014; 8:845–888
44. Hird K, Kirsner K, Little D, Oehmen R, Dunn J. Objective automated analysis of natural language: The Fluency Profiling System as a measure of the efficiency of dynamic language networks. Paper presented at: Frontiers in Human Neuroscience Conference Abstract: ACNS-2012 Australasian Cognitive Neuroscience Conference. Brisbane, Australia, 2012
45. Lee T, Kong A, Wang H. Duration of Content and Function Words in Oral Discourse by Speakers with Fluent Aphasia: Preliminary Data. Miami, FL: Academy of Aphasia; 2014
46. Broca P. Remarque sur le siege de la faculté du language articulé, suivie dune observation d'aphèmie (perte de la parole). Bull Soc Anat Paris 1861; 36:330–356
47. Wernicke C. Die Aphasische Symptomenkomplex. Breslau: Cohn & Weigart; 1874
48. MacWhinney B. Symptom Clustering for AphasiaBank Data. San Francisco, CA: Academy of Aphasia; 2012
49. Johansson B. The Influence of Educational Level and Occupational Status on the Spoken Language Production of Persons with Agrammatic Aphasia. Gothenburg, Sweden: Languages and Literatures, Gothenburg University; 2012
50. Fromm D, MacWhinney B, Forbes M, Holland A. Is the Cinderella Task Biased for Age or Gender? Montréal, Canada: Academy of Aphasia; 2011
51. Campbell S, Williams N, Hudspeth S, Franklin D, Richardson J. Discourse Measures in Aphasia and Observer Ratings of Comfort and Typicality. Chicago, IL: ASHA; 2013
52. Hancock AB, Stutts HW, Bass A. Perceptions of gender and femininity based on language: implications for transgender communication therapy. Lang Speech 2014
53. Ward K. Cultural Language Variations: An Examination of Appalachian Discourse. Huntington, WV: Communication Disorders, Marshall University; 2014
54. Wallace SJ, Worrall L, Rose T, Le Dorze G. Measuring outcomes in aphasia research: a review of current practice and an agenda for standardisation. Aphasiology 2014;28(11):1364–1384

55. MacWhinney B. Challenges facing COS development for aphasia. Aphasiology 2014; 28:1393–1395
56. Fridriksson J, Hubbard HI, Hudspeth SG, et al. Speech entrainment enables patients with Broca's aphasia to produce fluent speech. Brain 2012;135 (Pt 12):3815–3829
57. Szabo G, Fromm D, Heimlich T, Holland A. Script training and its application to everyday life in an aphasia center. Paper presented at: Clinical Aphasiology Conference; 2014; St. Simons Island, GA
58. Boyle M. Stability of word-retrieval errors with the AphasiaBank stimuli. Am J Speech Lang Pathol 2015;24(4):S953–S960