Fromm, D., & MacWhinney, B. (2023). Discourse databases for use with clinical populations. In C. A. Coelho, L. R. Cherney, & B. B. Shadden (Eds.). *Discourse analysis in adults with and without communication disorders* (pp. 311-327). Plural Publishing, Inc.

17 Discourse Databases for Use With Clinical Populations

Davida Fromm and Brian MacWhinney

Introduction

To build a scientific discipline, one cannot overstate the advantages of shared databases and shared resources. The scientific process is built on verification and replication of empirical data. This is particularly true for the study of discourse in clinical populations. Even the best descriptions of language samples, coding systems, reliability procedures, and outcome measures pale in comparison with the ability to see and hear speakers, see coded transcripts, and replicate results. Terms that describe features of language samples (e.g., *nonfluent, fluent, paraphasic, circumlocutory, agrammatic, aprosodic, paragrammatic, tangential*) and severity of impairment (e.g., *mild, moderate*) may mean slightly different things to different people based on their training and experience. Likewise, within a given research or clinical program there may be good reliability on coding correct information units (CIUs) or word-finding problems, but does that reliability extend to other researchers and clinicians? Do we have the raw materials and conditions necessary not only to verify and replicate data, but also to establish psychometric properties such as test-retest metrics so we can confidently credit treatment for changes in discourse?

The Open Science and Open Data initiatives have stressed the importance of making language data widely available (Chiarcos & Pareja-Lora, 2019), citing both the immense effort of creating resources and the potential gains of sharing and reusing data for purposes of replication, new applications, or novel experiments. Given the time invested in transcription, it makes sense to extend the results beyond single clinics or research laboratories. Data sharing also makes it possible to have online collaborative commentaries (MacWhinney, 2007) and conduct systematic, data-based comparisons to ascertain best practices for discourse measurement and analysis. Large databases allow for more robust statistical treatment of data, avoiding the limitation mentioned so often at the end of research articles about small sample sizes, insufficient power, and risks of Type II errors. More specifically, large normative databases, particularly for aging populations, are necessary to fully appreciate typical performance profiles in otherwise healthy or unimpaired populations. Finally, at the most fundamental level, scientific research, especially that funded by public funding, comes with the responsibility of opening data and results to the public in a fully transparent fashion.

In this chapter, we describe the primary discourse databases currently available for adult populations with and without communication disorders. The meaning of "currently available" is not a trivial matter. Li et al. (2019) published a systematic review of worldwide resources for speech databases and found 10 databases that met these criteria: targeting individuals with neurological disorders, recording audio or video samples, and making their resources available for other researchers. However, some of the database links provided are no longer active, and the review did not specify the actual steps involved in obtaining access to the database. In this chapter, we focus on the TalkBank databases, which are both currently available and readily accessible to authorized professionals. We then examine a few of the other major databases currently being used in published research, although access to other data sets often requires more extensive application, review, and approval procedures.

Discourse Resources Available Through TalkBank Clinical Banks

The goal of the TalkBank project (https://talkbank.org/), funded by the National Institutes of Health and the National Science Foundation, is to support data sharing and direct, communitywide access to naturalistic recordings and transcripts of human and animal communication. TalkBank is the world's largest open-access, integrated repository of data on spoken language, containing shared databases of multimedia interactions for the study of child language, aphasia, traumatic brain injury (TBI), right hemisphere disorder (RHD), fluency, autism spectrum disorder, and more. This larger project grew from the model created by the original CHild Language Data Exchange System (CHILDES) project (MacWhinney, 2014), which began in 1984 and focused on first-language acquisition in young children. This section will present information about four TalkBank clinical language banks: AphasiaBank, DementiaBank, RHDBank, and TBIBank. Before describing each of the banks, we will highlight some shared features of the TalkBank clinical banks and explain how the Computerized Language ANalysis (CLAN) program (https://talkbank.org/manuals/CLAN.pdf) expedites and improves the process of transcription and discourse analysis.

Shared Features of TalkBank Clinical Banks: Index to Corpora

Each of the clinical language banks in TalkBank has its own webpage with links to valuable resources for researchers, educators, and clinicians. These starting pages are all accessible from https://talkbank.org. Each webpage has a link called "Index to Corpora," which shows a list of all the corpora in that database with relevant information (e.g., age, number of participants, type of discourse tasks, type of media). Clicking on any corpus name in that list brings up a page with more specific information about the corpus, the contributors, and the project from which the discourse data were collected. From those individual corpus pages, there are also links for downloading the language transcripts and media files or going to the Browsable Database, where you can listen to and watch (for video files) the language sample while also following along with the transcript. It should be noted here that these clinical data are password protected and available only to licensed clinicians, educators, and researchers who agree to abide by the TalkBank Ground Rules for data sharing (https://talkbank.org/share/).[1]

Shared Features of TalkBank Clinical Banks: CHAT and CLAN

All transcripts in the TalkBank databases use a single consistent format, called Codes for the Human Analysis of Talk (CHAT) (https://talkbank.org/manuals/CHAT.pdf). This format has been developed over many years to accommodate the needs of a wide range of research communities and disciplines. Transcripts in CHAT format can then make use of an extensive set of analysis programs, called CLAN (https://talkbank.org/manuals/CLAN.pdf). The CLAN program is free to download from the main TalkBank website for Windows, Mac, and Unix. Complete CHAT and CLAN user manuals are available at the website, as is an *SLP's Guide to CLAN*, which provides an abridged and more user-friendly introduction to transcribing and analyzing samples for SLPs. In addition, the website has a link to *Tutorial Screencasts*, with more than 40 short video tutorials on how to transcribe and perform many different CLAN-related functions.

CHAT transcripts start with header tiers that give information about the language and the speakers. As you can see in the example below from DementiaBank, the transcript includes a participant and an investigator; the participant is a 66-year-old male from the Pitt corpus, who was diagnosed with probable Alzheimer's disease and scored 20 on the Mini-Mental Status Exam.

@Begin

@Languages: eng

@Participants: PAR Participant, INV Investigator

@ID: eng|Pitt|PAR|66;|male|ProbableAD||Participant|20||

@ID: eng|Pitt|INV||||Investigator|||

Most CHAT file transcripts at the shared databases will look like the example below, where the speaker's utterance (transcribed manually) is followed by two lines that get added to the file automatically by running the MOR command. The %mor tier has morphological tagging and part-of-speech categories; the %gra tier shows pairwise grammatical relations between words. The information on these tiers is used for many automatic discourse analysis commands in CLAN. The black dot next to the utterance holds the temporal information, linking that utterance to the audio or video file in milliseconds.

*PAR: I would grab two slices of bread . •

%mor: pro:sub|I mod|will&COND v|grab det:num|two n|slice-PL prep|of n|bread .

%gra: 1|3|SUBJ 2|3|AUX 3|0|ROOT 4|5|QUANT 5|3|OBJ 6|5|NJCT 7|6|POBJ 8|3|PUNCT

[1]Passwords are provided to all licensed SLPs, educators, and researchers. They should send an email request for TalkBank membership to Brian MacWhinney (macw@cmu.edu) with contact information, affiliation, and a brief general statement about how they envision using the resources. Students interested in accessing the data should ask their faculty advisors to join as members.

Advantages of transcribing in CHAT and analyzing a transcript with CLAN include:

(a) Smooth transcription. Transcribers can use normal English orthography (e.g., *can't, shoes, girl's*), and the morphological structure of these forms will be analyzed automatically by CLAN's MOR program.
(b) Sound playback. It is simple to time-link utterances in a CHAT file to the audio or video file and then replay the utterances individually to transcribe, add coding, check accuracy, add gestures, measure pauses, and so on.
(c) Faster analysis. Once transcripts are prepared, it takes seconds to run commands on a single transcript or hundreds of transcripts in a single step.
(d) Less demand for expertise and better reliability. Automatic analyses and computations are thoroughly replicable because repeated runs of the same command will always produce the same results. Results will not be dependent on the varying training and linguistic expertise of research and or clinical staff.

Shared Features of TalkBank Clinical Banks: Browsable Database

All TalkBank shared databases include links to the Browsable Database, which opens to a page of simple instructions on how to watch and hear videos (or audio files) while following along with the language transcript. A directory of corpora and files appears in the top left corner of the page in the Browsable Database. To watch, click on the file of interest and then press the play arrow next to the line in the transcript where you want to begin or on the video that appears in the lower left corner.

AphasiaBank

AphasiaBank (https://aphasia.talkbank.org/) is a shared, multimedia database for the study of communication in aphasia (Fromm, Forbes, et al., 2020; MacWhinney et al., 2011). Adults with aphasia often have impairments that affect their ability to communicate successfully at the discourse level. Analyses of discourse in aphasia have received a great deal of attention, with clinicians and researchers working to identify psychometrically sound approaches to discourse assessment and measurement. Since its inception in 2007, this database has grown to contain well over 1,000 videos and transcripts of people with aphasia (PWAs) and controls doing a variety of discourse tasks. A unique feature of this language bank is its standard discourse protocol and elicitation script for gathering language samples, which include free speech, picture descriptions, the *Cinderella* story narrative, and a procedural discourse task. The tasks were selected by a group of aphasiologists to capture a variety of discourse genres that were relevant to the population and used in the existing literature. For example, the protocol includes a free speech stroke story, asking individuals to tell what they remember about when they had their stroke and then about their recovery. Control participants were asked to tell what they remember about any illness or injury they had. Another protocol task is the *Cinderella* story, which has been used in aphasia literature for well over 30 years and is the second-most frequently reported language sampling technique for eliciting narratives in aphasia (Bryant et al., 2016).

The standard discourse protocol is augmented by comprehensive demographic data collection on all participants and a standard test battery that includes the Western Aphasia Battery–Revised (WAB-R) Aphasia Quotient subtests (Kertesz, 2007), the Boston Naming Test (Kaplan et al., 2001), the Verb Naming Test (Cho-Reyes & Thompson, 2012), and a repetition test developed to assess word-level and sentence-level repetition skills. Two tests of comprehension were added to the battery after initial data collection started: Sentence Comprehension, from the Philadelphia Comprehension Battery (Saffran et al., 1988); and Complex Ideational Material–Short Form, from the Boston Diagnostic Aphasia Exam (Goodglass et al., 2001). All this material (e.g., scripted discourse protocol, stimulus pictures, test materials) is available at the AphasiaBank website. The protocol collection includes more than 450 videos and transcripts of PWAs and more than 250 for controls in English. The protocol has been translated into other languages (e.g., Cantonese, Croatian, French, Italian, Mandarin, Spanish) and those corpora have been contributed to the database as well. Researchers and clinicians have also contributed videos and transcripts with media for nonprotocol-based discourse samples such as conversations, story retells, assessment tasks (e.g., Quick Aphasia Battery [Wilson et al., 2018], WAB-R picture descriptions, Grey Oral Reading Test [Wiederhold & Bryant, 2012]), and various types of treatment sessions (e.g., script training, group therapy).

Currently, password-protected access to the AphasiaBank corpora, demographic data, and test results has been granted to more than 1,250 faculty and licensed clinicians from more than 55 countries, all of whom have agreed to abide by the TalkBank data sharing ground rules. They work in a range of fields (e.g., speech-language pathology, linguistics, psychology, neurology, English, computer science, engineering) and use the database for research, teaching, and clinical purposes, examples of which are highlighted below. The database has been used in hundreds of publications, conference presentations, and theses, which can be accessed (without a password) from links at the AphasiaBank website (*Publications, Posters, and Presentations*). Given the amount of work that has been done, a review of what has been learned is beyond the scope of this chapter. However, the main areas of research include the development of new discourse outcome measures and norms, psychometric properties of discourse measures, new automatic discourse analysis tools, comparisons between manual and automatic analyses, comparisons of discourse outcome measures across genres, comparisons of different lexical diversity measures, informativeness and naming, coherence and cohesion, grammar, verb types and verb argument structure, gesture, fluency, paraphasia classification and error analysis, syndrome classification, demographic factors (e.g., race, sex, gender), change over time, and treatment effects.

Teaching Resources

The *Teaching* section of the webpage contains links to material gathered from various database resources. The goal for this section was to curate the vast resources to present cases, examples, and exercises that could be used for a wide range of educational purposes. A particularly complete teaching component is the *Grand Rounds* guided tutorial on aphasia, which focuses on how language differs across aphasia types (e.g., anomic, Broca's, conduction, global, transcortical motor, transcortical sensory, Wernicke's) and language tasks (e.g., free speech, naming, repetition, picture description).

The *Grand Rounds* pages include short case histories and discussion questions built around 40 captioned video segments from dozens of PWAs. Additional "Treatment Focus" questions stimulate thinking and discussions about ways to approach intervention to improve communication for each case presentation.

The *Examples* webpage provides very short, captioned video clips of common features from the connected speech of PWAs at the word level, sentence level, and discourse level. This page is organized by features (e.g., phonemic paraphasia, circumlocution, agrammatism, empty speech), with a description of the feature and links to video examples. Each video link includes basic information about the speaker (e.g., WAB-R AQ, aphasia type).

The *Classroom Activities* link downloads a Word file with ideas for exercises such as clinical assessment and treatment planning, measuring different aspects of discourse (e.g., CIUs, mean length of utterance [MLU]), using the EVAL command from the CLAN program (described below), coding speech errors (e.g., phonemic and semantic paraphasias), examining main concepts in narratives, and comparing across aphasia types as well as across other disorders (e.g., right hemisphere disorder). Several of these exercises were contributed by AphasiaBank members, and we appreciate and encourage this type of resource sharing. Surveys have repeatedly shown that knowledge about discourse analysis (e.g., measurement, sampling methods) is a major barrier to its use (Bryant et al., 2019; Bryant et al., 2017; Stark et al., 2021), making this an important area for continued development.

Discourse Analysis

The AphasiaBank homepage has a link to a webpage that is regularly updated with information about approaches to discourse analysis. The analyses currently featured include coherence (Wright et al., 2013; Wright et al., 2014), core lexicon (Dalton et al., 2020; Kim & Wright, 2020), CIUs (Nicholas & Brookshire, 1993), main concepts (Dalton & Richardson, 2019; Richardson & Dalton, 2020), story grammar (Greenslade et al., 2020; Richardson et al., 2021), and systemic functional linguistics (Groenewold & Armstrong, 2018). The page also describes methods for automated computation of the Northwestern Narrative Language Analysis (Fromm, MacWhinney, et al., 2020; Thompson et al., 1995) and the Quantitative Production Analysis (Fromm et al., 2021; Rochon et al., 2000; Saffran et al., 1989). Clicking on any of those topics brings up a page with descriptions of the analysis, links to relevant references, and tools and instructions for conducting the analyses. Along with the classroom activities, this is another important resource to continue developing and updating to address barriers to discourse analysis use. In conjunction with work underway by the Methodological and Data Quality task force from FOQUSAphasia (http://www.foqusaphasia.com), this *Discourse Analysis* section will add essential information concerning psychometric properties of commonly used discourse metrics to guide best practices in clinical and research settings (Stark et al., 2021).

CLAN Profile Analyses

Although it is beyond the scope of this chapter to discuss in detail, the AphasiaBank shared database has allowed for the development of several automated measurement tools specific to aphasia as well as the specific tasks in the protocol (Fromm, Forbes, et al., 2020). These commands can be used to get fast, reliable, and informative summaries of language performance. For example, the EVAL command produces a language profile in spreadsheet format that includes 34 output measures such as number of utterances, number of words, MLU, type-token ratio, vocD (Malvern et al., 2004), the moving average type-token ratio (MATTR: Covington & McFall, 2010), words per minute, percent or raw number of various parts of speech, noun-verb ratio, and open-class versus closed-class word ratio (Forbes et al., 2012). This command can show how a given individual compares to controls or others in the AphasiaBank database; it can also compare an individual's performance on the same task done at different points in time to evaluate changes following treatment. The CORELEX command uses core lexicon lists for the five AphasiaBank discourse protocol tasks (Dalton et al., 2020) to compute the number of core lexicon words an individual (or group of individuals) used for each task. FLUCALC can be used to analyze fluency, which is both a critical and yet somewhat elusive concept in the classification of aphasia, as it is associated with both grammatical factors and naming impairments, manifesting in a variety of behaviors such as hesitations, fillers, revisions, sound and sentence fragments, limited output, slower speech rate, and agrammatism (Gordon, 1998; Gordon & Clough, 2020). These tools provide automatic and efficient methods for providing key discourse outcome measures for both clinicians and researchers.

DementiaBank

DementiaBank (https://dementia.talkbank.org/) includes transcripts and media from individuals with various types of dementia as well as individuals with primary progressive aphasia and controls. The largest dementia corpus in the collection is the Pitt corpus (Becker et al., 1984), which contains longitudinal data for four language tasks (Cookie Theft picture descriptions, sentence construction, word fluency, and story retell) from more than 300 individuals with probable Alzheimer's disease (AD) and other types of dementia as well as over 200 elderly controls (Cookie Theft picture descriptions only). Another large database that was recently added is a subset of data from the Wisconsin Longitudinal Study (discussed in more depth below), which also contains Cookie Theft picture descriptions (more than 1,300) but no professional dementia diagnostic data (Herd et al., 2014). Other corpora in DementiaBank include conversations, semistructured interviews, and other language tasks from individuals with AD. The database also includes a corpus of discourse data from individuals with primary progressive aphasia, some seen multiple times, who completed *Cinderella* story narratives and other language tasks. In addition to the English data, DementiaBank corpora are available for German, Mandarin, Spanish, and Taiwanese. Pilot work has recently been initiated to create and use a standard discourse protocol to collect data from individuals with neurotypical and mild cognitive impairment and dementia. The DementiaBank webpage includes a *Protocol* section with details on the tasks and tests, selected specifically for these populations; data will be forthcoming (Lanzi, 2021).

Password-protected access to DementiaBank has been granted to more than 600 researchers, clinicians, and educators from over 50 countries, many of whom are computer scientists using DementiaBank data, primarily from the Pitt corpus (Becker et al., 1994), as the main testbed for the construction and benchmarking of language-based predictors of the onset of cognitive impairment. These data have been of particular interest to researchers who are using machine learning and linguistic analyses to automatically identify AD from short narrative samples as well as researchers who are working to improve speech-recognition skills in personal assistive robots trained to work with older adults with AD (e.g., Rudzicz, Wang, et al, 2014). The data set has been used recently by research groups all over the world in the context of the ADReSS Challenge for Interspeech in 2020 and again in 2021 to train classifiers for early automatic detection of dementia (Luz et al., 2020, 2021). A *Bibliography* link at the DementiaBank website includes more than 200 references to work that used DementiaBank data; a *Posters* link includes copies of posters that were presented at conferences.

RDHBank

RHDBank (https://rhd.talkbank.org/) focuses on the study of communication in adults with RHD resulting from brain damage to the right hemisphere (Minga et al., 2021; see Chapter 11). Symptoms of RHD include cognitive-communication deficits that impair pragmatic skills, resulting in difficulties producing and comprehending discourse. Specifically, difficulties with topic maintenance, discourse coherence and cohesion, inference generation, turn taking, question use, and the integration of contextual nuance are commonly seen in people with RHD (Blake, 2018). The cognitive-communication disorder associated with RHD has a negative impact on quality of life and social integration (Hewetson et al., 2021), leading to calls for better assessment and treatment approaches. Existing research about RHD discourse is limited in quantity and difficult to synthesize due to the use of different discourse tasks and outcome measures as well as methodological issues, such as limited descriptions of participants (Minga et al., 2021).

Like AphasiaBank, the RHDBank database includes a standard discourse protocol, demographic data collection, and a set of assessment procedures. The discourse protocol includes free speech, picture descriptions, the *Cinderella* storytelling task, a procedural discourse task, a question production task, and a first-encounter conversation (Kennedy et al., 1994). Overlaps with the AphasiaBank discourse protocol were intentional to allow for cross-disorder comparisons; other tasks (e.g., conversing to get to know a stranger) were included to assess the specific pragmatic and social discourse aspects of language that commonly affect individuals with RHD. The test battery assesses cognitive-linguistic functioning and visuospatial neglect and includes the Cognitive Linguistic Quick Test (Helm-Estabrooks, 2017), the Apples Test (Bickerton et al., 2011), the General Short Form of the Communicative Participation Item Bank (Baylor et al., 2013), and the Indented Paragraph Test (Caplan, 1987). A small number of other nonprotocol corpora (e.g., Cookie Theft picture descriptions from individuals following right hemisphere strokes) have been contributed to the database as well.

The RHDBank website contains links to a *Bibliography* showing references for published or presented work based on RHDBank data; another link also allows users to see conference *Posters*. So far, discourse analyses have focused predominantly on measures of global coherence, main concepts, question use (type and frequency) in conversation and structured tasks, and the relationship between discourse behaviors and cognitive functioning.

As in AphasiaBank, the teaching resources in RHDBank include a *Grand Rounds* tutorial and a *Classroom Activities* page. The RHD *Grand Rounds* is an educational platform that explains and illustrates the communication behaviors typically seen in individuals with RHD through case presentations, video clips, and discussion questions. The extralinguistic impairments (e.g., anosognosia, aprosodia, neglect) that can occur with RHD are explained as well, and key research articles on these topics are linked and briefly summarized. The *Classroom Activities* link downloads a document with ideas for using RHDBank resources to compare language disorders following left and right hemisphere stroke, plan assessment and treatment for two of the RHD cases from the *Grand Rounds*, code an RHD transcript for global coherence and main concepts, compute CIUs in RHD transcripts and compare them with aphasia and control transcripts, and more. These educational resources are important tools to use to increase exposure to and understanding of the subtle but debilitating cognitive-communication disorders that occur in RHD. These individuals often slip through the cracks both in referrals to SLPs and in research on effective diagnostic and treatment approaches.

TBIBank

TBIBank (https://tbi.talkbank.org/) is a shared database of multimedia interactions for the study of communication in people with TBI. TBI can result in cognitive-communication disorders that may affect all aspects of language (e.g., speaking, listening, reading, writing, pragmatics) as well as attention, reasoning, memory, and executive function (see Chapters 9–10). Discourse in TBI has been described as disorganized, inappropriate, tangential, unclear, redundant, and self-focused. This repository includes two sizeable corpora that each used a standard discourse protocol. One has longitudinal data that allow for the study of communication recovery after TBI (Stubbs et al., 2018; Togher et al., 2014). The complete set of media files in the Togher corpus have been only partially transcribed, but a wealth of data is available for the same 58 individuals at six measurement times ranging from 3 months to 3 years postonset. Discourse tasks in this corpus were similar to those used in the AphasiaBank protocol with two minor changes: substituting what the individual remembers about their brain injury instead of stroke in the free speech segment, and substituting a Vegemite and cheese sandwich (culturally appropriate for the Australian population) instead of peanut butter and jelly for the procedural discourse segment. Test data were also specific to the population and included a variety of TBI-relevant measures such as the Frenchay Dysarthria Assessment (Enderby & Palmer, 2008) and the Functional Assessment of Verbal Reasoning and Executive Functioning (MacDonald, 2005). The other corpus has 55 speakers with closed head injuries and 52 controls who did a variety of discourse tasks including story retelling, story generation, and informal conversation (Coelho et al., 2003). Demographic data and test results are available for these corpora. Several other smaller corpora include samples of conversations and various other discourse tasks.

Two new corpora were collected with the goal of understanding how typed discourse relates to individual differences in temperament and cognitive-linguistic performance among teens and adults with a recent or chronic history of concussion. One includes 231 written narrative and expository samples from 91 individuals (Stockbridge & Newman, 2019). The other includes multiple written expository samples from a separate group of 487 English-speaking, international roller derby players with significant histories of concussion and subconcussive exposure (Stockbridge et al., 2022).

Currently, there are more than 225 TBIBank members from more than 15 countries.

The webpage has links to a *Bibliography* and *Posters and Presentations* that have used the shared database. Examples of the research topics that have been investigated include discourse recovery in the first year following severe TBI (Elbourn et al., 2019), social communication assessment (Steel & Togher, 2019), discourse processing (Peach & Hanna, 2021), and conversational topics discussed by individuals with severe TBI and their communication partners (Brassel et al., 2016).

TBIBank *Grand Rounds* is a teaching resource designed to educate users on characteristics of discourse impairments typically seen in this population (Elbourn et al., 2020). Specifically, it includes modules with text, videos, and question-and-answer sections that cover the cognitive-communication impairment typically seen in TBI as well as its variability, assessment, treatment, recovery patterns, and comorbidities. As with RHDBank, tools that promote a more thorough understanding of the impairment and clinical best practices can translate into better overall outcomes for individuals with TBI.

Other Adult Discourse Database Resources

The literature refers to other adult language corpora, but a variety of factors often prevents their use in advancing research and education. For example, some are not publicly available, and some do not include enough demographic or other relevant metadata for meaningful analysis. However, a few of them have been made available to researchers, who must apply for access and demonstrate that their projects meet ethical criteria of institutional review boards (IRBs) and who are able to have their institutions sign data use agreements (DUAs). These databases ensure confidentiality, security, and privacy protection for participants through anonymization and deidentification. We highlight three resources below, with examples of research projects that have made use of their data. For a broader perspective, we refer readers to a recent systematic review of speech databases that include individuals with aphasia, dementia, stuttering, and other neurocognitive disorders (Li et al., 2019).

Carolina Conversations Collection

The Carolina Conversations Collection (CCC; Pope & Davis, 2011) is a corpus of older speakers from diverse ethnic and language groups, talking about health, illness, and their daily lives. The corpus is divided into two cohorts: unimpaired speakers and those with dementia. It includes audio and video files and time-aligned transcripts (in LaBB-CAT format) that can be downloaded by approved users. Access to this database requires a three-step process of developing a proposal explaining why the data are needed, getting IRB approval for the proposal from one's home institution, and then having the Medical University of South Carolina (MUSC) cosponsor the proposal and getting the MUSC's IRB approval for it as well.

Researchers have used this database to study several topics. Boyd Davis, one of the cocreators of the CCC, has published several articles and book chapters on dementia discourse, much of it focused on social-interactional analyses (Davis & Maclagan, 2009, 2010). Stickle and Wanner (2019, 2020) examined syntactic structures, looking at both grammatical accuracy and the range of linguistic complexity in the conversations of persons with dementia. Nasreen et al. (2019a, 2019b) examined questions, responses, and misunderstandings in conversations with people with AD. Guinn and Habash (2012) conducted a discriminative analysis of AD participants' disfluencies to create algorithms for automatic classification of speakers as AD versus non-AD (Guinn et al., 2014). Abdalla et al. (2018) used automatic extraction methods on this corpus as well as the DementiaBank Pitt corpus to study discourse relations according to Rhetorical Structure Theory. Also working with both DementiaBank and CCC data, Rudzicz, Chang Curry, et al. (2014) extracted more than 200 lexical/syntactic and acoustic features to determine which features were indicative of *trouble-indicating behaviors* in the speech of individuals with AD. Green et al. (2012) used the CCC corpus to inform the development of a computational model for a natural language-processing system that can listen to conversation between someone with AD and an unimpaired partner and make suggestions to the partner that would aid in maintaining the conversational flow. Luz et al. (2018) used logistic regression on content-free features (e.g., dialogue duration, turn duration, number of words, speech rate information) extracted from subsets of these CCC interviews to build a predictive model that could differentiate AD and non-AD speech with 86.5% accuracy (de la Fuente Garcia et al., 2020).

The Wisconsin Longitudinal Study

The Wisconsin Longitudinal Study (WLS; Herd et al., 2014) is a large-scale, long-term longitudinal study of a random sample of more than 1,300 men and women who graduated from Wisconsin high schools in 1957. Data on these participants were collected at five additional time points through 2011 and covered a wide range of areas. Most relevant to this chapter are the Cookie Theft picture descriptions, which were added to the test protocol in 2011. Measures of category and letter fluency data, immediate and delayed word recall, and similarities data allow for some related measures of cognitive abilities. As mentioned earlier, this subset of WLS data has been shared with TalkBank and is available at the DementiaBank website.

Researchers are beginning to use this large data set as an additional resource for discourse studies in aging. Guo et al. (2021) recently reported on machine learning models to improve the performance of deep learning-based methods. The challenge in using this rich resource is that the WLS metadata do not contain dementia diagnoses. However, these authors used results from cognitive tests to establish groups for automatic dementia detection. Noorian et al. (2017) have also tapped this resource to increase the size of normative data for advancing this work on automated detection. The WLS website is clear and informative, the methodology is well-documented, and the staff is responsive to requests for data sharing and collaboration.

Wisconsin Registry for Alzheimer's Prevention (WRAP)

WRAP is a large, longitudinal study that was established in 2001 and designed to identify cognitive features and biomarkers that may predict AD risk. At time of enrollment, participants are late-middle-aged adults with a parental family history of probable AD (Johnson et al., 2018). Detailed visits occur approximately every 2 years and involve a variety of cognitive assessments, anthropometric measures, laboratory tests, and questionnaire ratings. The data include Cookie

Theft picture descriptions that have been used to analyze connected language for early features of cognitive decline (Evans et al., 2021; Mueller et al., 2016; Mueller et al., 2018). Outside researchers can apply for access to this database by completing a detailed request (e.g., NIH formatted biosketch, project summary, specific aims, significance, approach, deliverables) with proof of IRB approval.

Conclusions

Most of the research that has utilized these shared databases could not have been accomplished without access to the amount and type of data available from these resources. We hope this chapter has provided encouragement to those who collect valuable discourse data to further this effort by contributing their corpora to shared databases for the benefit of the community. To facilitate this process, it is important to obtain full informed consent for data sharing (https://talkbank.org/share). The DementiaBank Pitt corpus is a prime example of a data set that has served a purpose beyond anything the original investigators could have envisioned and, as a result, has had a major impact on the push to develop clinical tools for automatic screening and detection of dementia.

The advantages of shared databases are many. These shared databases have facilitated the development of new discourse evaluation tools for clinicians and researchers in the field, using automated analyses as well as transcription-based and nontranscription-based analyses. Norms and benchmarks have been established for comparing participants' discourse performance to that of controls or others based their age, sex, and diagnosis. Many smaller academic programs are unlikely to afford students the opportunity to see a full range of disorders. Additionally, the recent COVID-19 pandemic necessitated remote learning, which further limited students' clinical experiences. The rich data sets have provided material for online tutorials and other teaching resources that fill important gaps in access and breadth of clinical exposure and allow students to learn about state-of-the art discourse analysis techniques.

Acknowledgments

We are indebted to the many colleagues who have collaborated with us and contributed data, and especially to the thousands of individuals who have participated and consented to share their data. This work was supported by the National Institute on Deafness and other Communication Disorders [Grant R01-DC008524] (2007–2022, awarded to MacWhinney).

References

Abdalla, M., Rudzicz, F., & Hirst, G. (2018). Rhetorical structure and Alzheimer's disease. *Aphasiology*, *32*(1), 41–60. https://doi.org/10.1080/02687038.2017.1355439

Baylor, C., Yorkston, K., Eadie, T., Kim, J., Chung, H., & Amtmann, D. (2013). The Communicative Participation Item Bank (CPIB): Item bank calibration and development of a disorder-generic short form. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1190–1208. https://doi.org/10.1044/1092-4388(2012/12-0140)

Becker, J. T., Boller, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis. *Archives of Neurology, 51*(6), 585–594. https://doi.org/10.1001/archneur.1994.00540180063015

Bickerton, W. L., Samson, D., Williamson, J., & Humphreys, G. W. (2011). Separating forms of neglect using the Apples Test: Validation and functional prediction in chronic and acute stroke. *Neuropsychology, 25*(5), 567–580. https://doi.org/10.1037/a0023501

Blake, M. L. (2018). *The right hemisphere and disorders of cognition and communication: Theory and clinical practice.* Plural Publishing. https://www.ch-stquentin.fr/sites/stquentin-ch.local/files/webform/pdf-the-right-hemisphere-and-disorders-of-cognition-and-communication-margaret-lehman-blake-pdf-download-free-book-4836432.pdf

Boyle, M. (2015). Stability of word-retrieval errors with the AphasiaBank stimuli. *American Journal of Speech-Language Pathology, 24*(4), S953–S960. https://doi.org/10.1044/2015_AJSLP-14-0152

Brassel, S., Kenny, B., Power, E., Elbourn, E., McDonald, S., Tate, R., . . . Togher, L. (2016). Conversational topics discussed by individuals with severe traumatic brain injury and their communication partners during subacute recovery. *Brain Injury, 30*(11), 1329–1342. https://doi.org/10.1080/02699052.2016.1187288

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics and Phonetics, 30*(7), 489–518. https://doi.org/10.3109/02699206.2016.1145740

Bryant, L., Ferguson, A., Valentine, M., & Spencer, E. (2019). Implementation of discourse analysis in aphasia: Investigating the feasibility of a knowledge-to-action intervention. *Aphasiology, 33*(1), 31–57. https://doi.org/10.1080/02687038.2018.1454886

Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology, 31*(10), 1105–1126. https://doi.org/10.1080/02687038.2016.1239013

Caplan, B. (1987). Assessment of unilateral neglect: A new reading test. *Journal of Clinical and Experimental Neuropsychology, 9*(4), 359–364. https://doi.org/10.1080/01688638708405056

Chiarcos, C., & Pareja-Lora, A. (2019). Open data—linked data—linked open data—linguistic linked open data (LLOD): A general introduction. In A. Pareja-Lora, M. Blume, B. C. Lust, & C. Chiarcos (Eds.), *Development of linguistic linked open data resources for collaborative data-intensive research in the language sciences.* The MIT Press. http://library.oapen.org/handle/20.500.12657/23502

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Coelho, C. A., Youse, K., Lê, K., & Feinn, R. (2003). Narrative and conversational discourse of adults with closed head injuries and non-brain-injured adults: A discriminant analysis. *Aphasiology, 17*(5), 499–510. https://doi.org/10.1080/02687030344000111

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics, 17*(2), 94–100. https://doi.org/10.1080/02687038.2012.693584

Dalton, S. G. H., Kim, H., Richardson, J. D., & Wright, H. H. (2020). A compendium of core lexicon checklists. *Seminars in Speech and Language, 41*(1), 45–60. https://doi.org/10.1055/s-0039-3400972

Dalton, S. G. H., & Richardson, J. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. *American Journal of Speech-Language Pathology, 28*(Suppl. 1), 293–320. https://doi.org/10.1044/2018_AJSLP-17-0166

Davis, B. H., & Maclagan, M. (2009). Examining pauses in Alzheimer's discourse. *American Journal of Alzheimer's Disease and Other Dementias, 24*(2), 141–154. https://doi.org/10.1177/1533317508328138

Davis, B., & Maclagan, M. (2010). Pauses, fillers, placeholders and formulaicity in Alzheimer's discourse. *Fillers, Pause,s and Placeholders, 93*, 189-216. https://doi.org/10.1075/tsl.93

de la Fuente Garcia, S., Haider, F., & Luz, S. (2020). Cross-corpus feature learning between spontaneous monologue and dialogue for automatic classification of Alzheimer's dementia speech. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5851–5855). IEEE. https://doi.org/10.1109/EMBC44109.2020.9176305

Elbourn, E., Kenny, B., Power, E., Honan, C., McDonald, S., Tate, R., . . . Togher, L. (2019). Discourse recovery after severe traumatic brain injury: Exploring the first year. *Brain Injury, 33*(2), 143–159. https://doi.org/10.1080/02699052.2018.1539246

Elbourn, E., Togher, L., Steel, J., Power, E., Fromm, D., & MacWhinney, B. (2020). *TBIBank Grand Rounds: Online education platform about cognitive-communication disorders resulting from traumatic brain injury.* https://tbi.talkbank.org/education/class-tbi/

Enderby, P., & Palmer, R. (2008). *Frenchay dysarthria assessment (FDA-2)* (2nd ed.). Pro-Ed.

Evans, E., Coley, S. L., Gooding, C., Norris, N., Ramsey, C. M., Green-Harris, G., & Mueller, K. D. (2021). Preliminary assessment of connected speech and language as marker for cognitive change in late-middle-aged Black/African American adults at risk for Alzheimer's disease. *Aphasiology*, 1–24. https://doi.org/ *(FDA-2)* 10.1080/02687038.2021.1931801

Forbes, M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language, 30*(3), 217–222. https://doi.org/10.1055/s-0032-1320041

Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2020). Using AphasiaBank for discourse assessment. *Seminars in Speech and Language, 41*(41), 10–19. https://doi.org/10.1055/s-0039-3399499

Fromm, D., Katta, S., Paccione, M., Hecht, S., Greenhouse, J., MacWhinney, B., & Schnur, T. T. (2021). A comparison of manual versus automated Quantitative Production Analysis of connected speech. *Journal of Speech, Language, and Hearing Research, 64*(4), 1271–1282. https://doi.org/10.1044/2020_JSLHR-20-00561

Fromm, D., MacWhinney, B., & Thompson, C. K. (2020). Automation of the Northwestern Narrative Language Analysis System. *Journal of Speech, Language, and Hearing Research, 63*(6), 1835–1844. https://doi.org/10.1044/2020_JSLHR-19-00267

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston Diagnostic Aphasia Examination* (3rd ed.). Lippincott Williams & Wilkins.

Gordon, J. K. (1998). The fluency dimension in aphasia. *Aphasiology, 12*(7–8), 673–688. https://doi.org/10.1080/02687039808249565

Gordon, J. K., & Clough, S. (2020). How fluent? Part B. Underlying contributors to continuous measures of fluency in aphasia. *Aphasiology, 34*(5), 643–663. https://doi.org/10.1080/02687038.2020.1712586

Green, N., Guinn, C. I., & Smith, R. W. (2012). Assisting social conversation between persons with Alzheimer's disease and their conversational partners. In J. Alexandersson, P. Ljunglöf, K. F. McCoy, B. Roark, & A. Waller (Eds.), *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies* (pp. 37–46). https://aclanthology.org/W12-2906.pdf

Greenslade, K. J., Stuart, J. E., Richardson, J. D., Dalton, S. G., & Ramage, A. E. (2020). Macrostructural analyses of Cinderella narratives in a large nonclinical sample. *American Journal of Speech-Language Pathology, 29*(4), 1923–1936. https://doi.org/10.1044/2020_AJSLP-19-00151

Groenewold, R., & Armstrong, E. (2018). The effects of enactment on communicative competence in aphasic casual conversation: A functional linguistic perspective. *International Journal of Language and Communication Disorders, 53*(4), 836–851. https://doi.org/10.1111/1460-6984.12392

Guinn, C. I., & Habash, A. (2012). Language analysis of speakers with dementia of the Alzheimer's type. In *2012 AAAI Fall Symposium: Artificial Intelligence for Gerontechnology* (pp. 8–13). Arlington.

Guinn, C., Singer, B., & Habash, A. (2014). A comparison of syntax, semantics, and pragmatics in spoken language among residents with Alzheimer's disease in managed-care facili-

& Thompson, C.
f the Northwest-
Analysis System.
ıge, and Hearing
4. https://doi.org/
0267
3arresi, B. (2001).
Examination (3rd
: Wilkins.
ıency dimension
2(7–8), 673–688.
·87039808249565
)20). How fluent?
ıtors to continu-
phasia. *Aphasiol-*
'doi.org/10.1080/
th, R. W. (2012).
ɔn between per-
se and their con-
Jexandersson, P.
ıark, & A. Waller
ird Workshop on
sing for Assistive
ɔs://aclanthology
Richardson, J. D.,
E. (2020). Mac-
lerella narratives
. *American Jour-*
ıthology, 29(4),
0.1044/2020_AJ
E. (2018). The
communicative
al conversation:
ective. *Interna-*
ıd Communica-
351. https://doi
'2
)12). Language
ementia of the
AI Fall Sympo-
r Gerontechnol-
(2014). A com-
and pragmatics
residents with
ged-care facili-

ties. In *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-Health (CICARE)* (pp. 98–103). IEEE. https://doi.org/10.1109/CICARE.2014.7007840

Guo, Y., Li, C., Roan, C., Pakhomov, S., & Cohen, T. (2021). Crossing the "Cookie Theft" corpus chasm: Applying what BERT learns from outside data to the ADReSS challenge dementia detection task. *Frontiers in Computer Science, 3*, 26. https://doi.org/10.3389/fcomp.2021.642517

Helm-Estabrooks, N. (2017). *Cognitive linguistic quick test-plus*. Pearson.

Herd, P., Carr, D., & Roan, C. (2014). Cohort profile: Wisconsin Longitudinal (WLS). *International Journal of Epidemiology, 43*(1), 34–41. https://doi.org/10.1093/ije/dys194

Hewetson, R., Cornwell, P., & Shum, D. H. (2021). Relationship and social network change in people with impaired social cognition post right hemisphere stroke. *American Journal of Speech-Language Pathology, 30*(2S), 962–973. https://doi.org/10.1044/2020_AJSLP-20-00047

Johnson, S. C., Koscik, R. L., Jonaitis, E. M., Clark, L. R., Mueller, K. D., Berman, S. E., . . . Sager, M. A. (2018). The Wisconsin Registry for Alzheimer's Prevention: A review of findings and current directions. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring, 10*, 130–142. https://doi.org/10.1016/j.dadm.2017.11.007

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd ed.). Pro-Ed.

Kennedy, M., Strand, E., Burton, W., & Peterson, C. (1994). Analysis of first-encounter conversations of right-hemisphere damaged participants. *Clinical Aphasiology, 22*, 67–80. http://aphasiology.pitt.edu/159/1/22-05.pdf

Kertesz, A. (2007). *Western Aphasia Battery–Revised.* Pearson.

Kim, H., & Wright, H. H. (2020). A tutorial on core lexicon: Development, use, and application. *Seminars in Speech and Language, 41*, 20–31. https://doi.org/10.1055/s-0039-3400973

Lanzi, M. A. (2021). *DementiaBank: Methods for studying discourse* [Manuscript in preparation]. Department of Communication Sciences and Disorders, University of Delaware.

Li, Y., Lin, Y., Ding, H., & Li, C. (2019). Speech databases for mental disorders: A systematic review. *General Psychiatry, 32*(3) e100022. https://doi.org/10.1136/gpsych-2018-100022

Luz, S., de la Fuente, S., & Albert, P. (2018). A method for analysis of patient speech in dialogue for dementia detection. https://arXiv.org/abs/1811.09919

Luz, S., Haider, F., de la Fuente, S., Fromm, D., & MacWhinney, B. (2020). *Alzheimer's dementia recognition through spontaneous speech: The ADReSS challenge.* https://arXiv.org/abs/2004.06833

Luz, S., Haider, F., de la Fuente, S., Fromm, D., & MacWhinney, B. (2021). *Detecting cognitive decline using speech only: The ADReSS challenge.* https://arXiv.org/abs/2104.09356

MacDonald, S. (2005). *Functional assessment of verbal reasoning and executive strategies.* CCD Publishing. http://www.ccdpublishing.com/favres.aspx

MacWhinney, B. (2007). The TalkBank project. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and digitizing language corpora* (Vol. 1, pp. 163–180). Palgrave-Macmillan. https://psyling.talkbank.org/years/2007/palgrave.pdf

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs*. Psychology Press.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*, 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Malvern, D., Richards, B., Chipere, N., & Purán, P. (2004). *Lexical diversity and language development.* Palgrave Macmillan. https://doi.org/10.1057/9780230511804

Minga, J., Johnson, M., Blake, M. L., Fromm, D., & MacWhinney, B. (2021). Making sense of right hemisphere discourse using RHDBank. *Topics in Language Disorders, 41*(1), 99–122. https://doi.org/10.1097/TLD.0000000000000244

Mueller, K. D., Hermann, B., Mecollari, J., & Turkstra, L. S. (2018). Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology, 40*(9), 917–939. https://doi.org/10.1080/13803395.2018.1446513

Mueller, K. D., Koscik, R. L., Turkstra, L. S., Riedeman, S. K., LaRue, A., Clark, L. R., . . . Johnson, S. C. (2016). Connected language in late-middle-aged adults at risk for Alzheimer's disease. *Journal of Alzheimer's Disease, 54*(4), 1539–1550. https://doi.org/10.3233/JAD-160252

Nasreen, S., Purver, M., & Hough, J. (2019a). A corpus study on questions, responses and misunderstanding signals in conversations with Alzheimer's patients. *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue—Full Papers (SEMDIAL)* (Vol. 13). London, United Kingdom. http://semdial.org/anthology/Z19-Nasreen_semdial_0013.pdf

Nasreen, S., Purver, M., & Hough, J. (2019b). Interaction patterns in conversations with Alzheimer's patients [Poster presentation and abstract]. In C. Martin-Vide, M. Purver, & S. Pollak (Eds.), *Proceedings: Seventh International Conference on Statistical Language and Speech Processing.* Ljubljana, Slovenia. http://www.eecs.qmul.ac.uk/~jhough/papers/NasreenEtAl2019SLSP.pdf

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research, 36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Noorian, Z., Pou-Prom, C., & Rudzicz, F. (2017). On the importance of normative data in speech-based assessment. *arXiv preprint arXiv:1712.0069.* https://arxiv.org/pdf/1712.00069.pdf

Peach, R. K., & Hanna, L. E. (2021). Sentence-level processing predicts narrative coherence following traumatic brain injury: Evidence in support of a resource model of discourse processing. *Language, Cognition and Neuroscience, 36*(6), 1–17. https://doi.org/10.1080/23273798.2021.1894346

Pope, C., & Davis, B. (2011). Finding a balance: The CCC corpus. *Corpus Linguistics and Linguistic Theory*, 7(1),143–161. https://doi.org/10.1515/CLLT.2011.007

Richardson, J. D., & Dalton, S. G. H. (2020). Main concepts for two picture description tasks: An addition to Richardson and Dalton, 2016. *Aphasiology, 34*(1), 119–136. https://doi.org/10.1080/02687038.2018.1561417

Richardson, J. D., Dalton, S. G., Greenslade, K. J., Jacks, A., Haley, K. L., & Adams, J. (2021). Main concept, sequencing, and story grammar analyses of Cinderella narratives in a large sample of persons with aphasia. *Brain Sciences, 11*(1), 110. https://doi.org/10.3390/brainsci11010110

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language, 72*(3), 193–218. https://doi.org/10.1006/brln.1999.2285

Rudzicz, F., Chan Currie, L., Danks, A., Mehta, T., & Zhao, S. (2014). Automatically identifying trouble-indicating speech behaviors in Alzheimer's disease. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 241–242). http://dx.doi.org/10.1145/2661334.2661382

Rudzicz, F., Wang, R., Begum, M., & Mihailidis, A. (2014). Speech recognition in Alzheimer's disease with personal assistive robots. In J. Alexandersson, D. Anastasiou, C. Jian, A. Nenkova, R. Patel, F. Rudzicz, . . . D. Zhekova, (Eds.), *Proceedings of the Fifth Workshop on Speech and Language Processing for Assistive Technologies* (pp. 20–28). https://aclanthology.org/W14-1904.pdf

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37*(3), 440–479. https://doi.org/10.1016/0093-934X(89)90030-8

Saffran, E. M., Schwartz, M. F., Linebarger, M. C., Martin, N., & Bochetto, P. (1998). *Philadelphia Comprehension Battery* [Unpublished manuscript].

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., . . . Roberts, A. C. (2021). Spoken discourse assessment and analysis in aphasia: An international sur-

vey of current
Language, and
4366–4389. ht
JSLHR-20-007(
Steel, J., & Togher,
tion assessment
A narrative rev
matic and disc
Brain Injury, 33
.1080/02699052
Stickle, T., & War
patterns exhibit
in conversation
43–63. https://
001
Stickle, T., & Wann
syntactic error i
sons with deme
coparticipants.
from the talk of
85–109). Palgr
.org/10.1007/97
Stockbridge, M. D
S. (2022). Conc
roller derby. *Fr*
https://www.fro
fneur.2022.8099
Stockbridge, M. D
Enduring cogni
in individuals w
American Journa
ogy, 28(4), 1554
44/2019_AJSLP-
Stubbs, E., Togher,
Forbes, M., Mac
(2018). Procedur
adults with sever
and 6 months po

vey of current practices. *Journal of Speech, Language, and Hearing Research, 64*(11), 4366–4389. https://doi.org/10.1044/2021_JSLHR-20-00708

Steel, J., & Togher, L. (2019). Social communication assessment after traumatic brain injury: A narrative review of innovations in pragmatic and discourse assessment methods. *Brain Injury, 33*(1), 48–61. https://doi.org/10.1080/02699052.2018.1531304

Stickle, T., & Wanner, A. (2019). Transitivity patterns exhibited by persons with dementia in conversation. *Applied Linguistics, 40*(1), 43–63. https://doi.org/10.1093/applin/amx001

Stickle, T., & Wanner, A. (2020). Making sense of syntactic error in conversations between persons with dementia and their non-impaired coparticipants. In T. Stickle (Ed.), *Learning from the talk of persons with dementia* (pp. 85–109). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-43977-4_6

Stockbridge, M. D., Keser, Z., & Newman, R. S. (2022). Concussion in women's flat-track roller derby. *Frontiers in Neurology*, 65. https://www.frontiersin.org/articles/10.3389/fneur.2022.809939/full

Stockbridge, M. D., & Newman, R. (2019). Enduring cognitive and linguistic deficits in individuals with a history of concussion. *American Journal of Speech-Language Pathology, 28*(4), 1554–1570. https://doi.org/10.1044/2019_AJSLP-18-0196

Stubbs, E., Togher, L., Kenny, B., Fromm, D., Forbes, M., MacWhinney, B., . . . Power, E. (2018). Procedural discourse performance in adults with severe traumatic brain injury at 3 and 6 months postinjury. *Brain Injury, 32*(2), 167–181. https://doi.org/10.1080/02699052.2017.1291989

Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B., Schneider, S., & Ballard, K. (1995). A system for the linguistic analysis of agrammatic language production. *Brain and Language, 51*(1), 124–129. https://doi.org/10.1006/brln.1995.1057

Togher, L., Elbourn, E., Kenny, B., Power, E., McDonald, S., Tate, R., . . . MacWhinney, B. (2014). TBIBank is a feasible assessment protocol to evaluate the cognitive communication skills of with people with severe TBI during the subacute stage of recovery. *Brain Injury, 28*(5–6), 723–723. https://doi.org/10.3109/02699052.2014.892379

Wiederholt, J. L., & Bryant, B. R. (2012). *Gray Oral Reading Tests–Fifth Edition: Examiner's manual*. Pro-Ed.

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLOS ONE, 13*(2), e0192773. https://doi.org/10.1371/journal.pone.0199469

Wright, H. H., Capilouto, G. J., & Koutsoftas, A. (2013). Evaluating measures of global coherence ability in stories in adults. *International Journal of Language and Communication Disorders, 48*(3), 249–256. https://doi.org/10.1111/1460-6984.12000

Wright, H. H., Koutsoftas, A., Capilouto, G. J., & Fergadiotis, G. (2014). Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Aging, Neuropsychology, and Cognition, 21*(2), 174–196. https://doi.org/10.1080/13825585.2013.794894