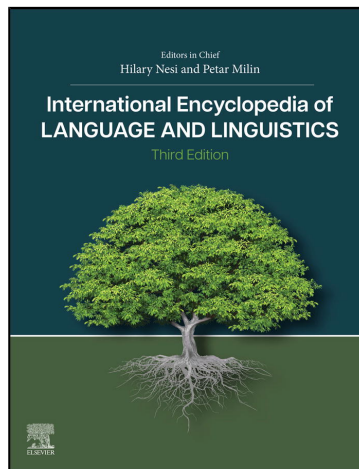


Provided for non-commercial research and educational use.  
Not for reproduction, distribution or commercial use.

This chapter was originally published in *International Encyclopedia of Language and Linguistics*, 3e (LAL3), published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use, including without limitation, use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation, commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<https://www.elsevier.com/about/policies/copyright/permissions>

Fromm, D., & MacWhinney, B. (2026). AphasiaBank: Resources for Academic and Clinical Teaching and Research. In: Nesi, H., Milin, P. (Eds.), *International Encyclopedia of Language and Linguistics*, 3e. Vol 12., pp. 362–366. UK: Elsevier. <https://dx.doi.org/10.1016/B978-0-323-95504-1.00030-2>. ISBN: 9780323955041

Copyright © 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

## AphasiaBank: Resources for Academic and Clinical Teaching and Research

Davida Fromm and Brian MacWhinney, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, United States

© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

<b>Introduction</b>	<b>362</b>
<b>History of AphasiaBank</b>	<b>363</b>
<b>Database</b>	<b>363</b>
<b>Transcripts and Analysis</b>	<b>363</b>
<b>Research Applications</b>	<b>364</b>
<b>Teaching Resources</b>	<b>364</b>
<b>Future Directions</b>	<b>365</b>
<b>Conclusion</b>	<b>365</b>
<b>Acknowledgments</b>	<b>365</b>
<b>References</b>	<b>365</b>
<b>List of Relevant Websites</b>	<b>366</b>

### Key Points

- AphasiaBank is part of the TalkBank project, focused on creating a shared database to advance the study and understanding of communication in aphasia.
- The freely downloadable CLAN program is used for transcribing and analyzing discourse samples across all TalkBank databases.
- The AphasiaBank database contains thousands of password-protected multimedia communicative interactions with people with aphasia that can be used for research and teaching purposes.
- The AphasiaBank website also has many resources and tools that can be used for teaching about aphasia, collecting and analyzing discourse samples, collaborative commentary and coding in language transcripts, and accessing hundreds of research articles and conference presentations that made use of the AphasiaBank database.

### Abstract

AphasiaBank is a shared, multimedia database for the study of communication in aphasia, which results from damage to the language areas of the brain. The database includes thousands of media files and transcripts from people with aphasia, mostly in English, but also in other languages. Approximately half of the collection uses a standard discourse protocol that was also administered to people without aphasia for comparison. AphasiaBank members can access the password protected transcripts and media files for a variety of research and teaching purposes. The website also includes information on research findings and methods for discourse analysis in aphasia.

### Introduction

Aphasia is a language impairment that results from damage (usually from a stroke) to the language areas of the brain (usually in the left hemisphere). Depending on factors such as the site and size of the brain damage, the language impairment may involve speaking, comprehending, reading, and/or writing. The speaking difficulty can manifest in different ways but often involves word finding problems, word errors (e.g., “foot” for “shoe”), sound errors (e.g., “duther” for “mother”), grammatical errors (e.g., omission of function words and morphemes), and/or unintelligible segments. The speech output may be fluent or nonfluent, due to filled pauses (e.g., uh, um), silent pauses, false starts, attempts at self-corrections, revisions, and repetitions. Some people with aphasia may have limb weakness or paralysis; some may also have accompanying apraxia of speech, which is a disruption in the sensorimotor speech programming that leads to an impairment in the ability to coordinate articulatory movements. People with aphasia benefit from treatment by speech-language pathologists and continued involvement in stroke groups that focus on improving language skills as well as enhancing participation in life activities, relationships, and self-esteem.

AphasiaBank is a shared database of multimedia communicative interactions with people with aphasia (MacWhinney et al., 2011). It is the only openly available data source for spoken language and communication in aphasia. The focus on spoken language reflects its central role in people’s lives, its importance in maintaining relationships and navigating activities of daily living.

The overarching goal of AphasiaBank is to improve the assessment and treatment of people with aphasia. Building an empirical database to advance the study and understanding of communication in aphasia provides resources and tools to help achieve this goal.

## History of AphasiaBank

AphasiaBank is part of the TalkBank system, which is currently the largest repository of shared databases for spoken language, with over one dozen language banks devoted to specific areas of study. Initial funding for AphasiaBank was awarded to Brian MacWhinney and Audrey Holland at Carnegie Mellon University in 2007, following a meeting two years earlier of 20 senior aphasia researchers who agreed on the need for a shared protocol, a shared database, more discourse analysis tools, and increased availability of computational tools for the aphasia research community. AphasiaBank members (currently numbering 1500) can freely access the password-protected media files (audio and video) and language transcripts in the database for educational and research purposes. Other resources at the AphasiaBank webpage, such as posters presented at conferences and discourse analysis approaches are openly available without registration. Researchers, educators, and licensed clinicians are eligible for membership and must agree to abide by the TalkBank data sharing ground rules.

## Database

The AphasiaBank database contains thousands of media files and language transcripts of people with different types and severities of aphasia. A unique feature of this language bank, that has since been adopted by some of the other adult language banks, is its use of a standard discourse protocol and elicitation script for gathering language samples. The protocol includes free speech tasks, three picture descriptions, the Cinderella story narrative, and a procedural discourse task. These protocol samples are mostly in English, but smaller corpora are available for French, Cantonese, Croatian, German, Greek, Italian, Japanese, Mandarin, Romanian, Spanish, and some bilingual participants. The standard discourse protocol is augmented by a standard test battery (e.g., Western Aphasia Battery-Revised [WAB-R; Kertész, 2007], Boston Naming Test [BNT; Kaplan et al., 2001], Verb Naming Test [VNT; Cho-Reyes & Thompson, 2012]) and comprehensive demographic data on participants. The advantage of having a standard discourse protocol and test battery is that it allows for the collection of a large and robust database with maximal comparability across tasks and participants.

In addition to the protocol samples, aphasia researchers and clinicians have contributed many other collections of media files and transcripts using other language tasks (e.g., conversations, story retells, Cookie Theft picture descriptions) as well as assessments (e.g., Quick Aphasia Battery (Wilson et al., 2018), Gray Oral Reading Test (Wiederholt & Bryant, 2012)) and treatment approaches (e.g., script training, group therapy). These non-protocol samples also number in the thousands and contain corpora in multiple languages, though the majority are in English. All the protocol and non-protocol corpora are listed in an index and described in more detail on individual corpus pages. Each individual corpus page has links for accessing the media files and transcripts, which can be downloaded or viewed through a browser at the Browsible Database. The AphasiaBank website provides links to all these resources and materials.

Two interesting things happened during the COVID pandemic years, beginning in March of 2020. One was that AphasiaBank videos became vital to the education of students who were no longer able to be in their classrooms or clinical placements. The second was that researchers developed and refined tools for administering the AphasiaBank discourse protocol via teletesting. The website now includes links and materials for collecting data from people with aphasia and controls using a variety of formats. This has proven to be an effective and efficient way to increase participation.

## Transcripts and Analysis

Language samples in AphasiaBank and the entire TalkBank system are transcribed in CHAT format and linked to the media files. Recent advances have allowed for the transcription process to be done automatically so that CHAT transcripts are created from audio files with each word and utterance linked (temporally aligned) to the media file (Liu et al., 2023). The automatic transcripts still require human review to check accuracy and code behaviors such as revisions and word errors, but the task is dramatically less time intensive than transcribing and linking the sample from scratch. These CHAT transcripts can then be analyzed with a set of programs called CLAN (Computerized Language Analysis), which enables a wide range of linguistic and discourse analyses. All AphasiaBank transcripts include three tiers of information under each utterance in the transcript: (1) a %mor tier with information about part-of-speech (e.g., noun, preposition) and morphology (e.g., plural, past); (2) a %gra tier with information about grammatical relations between words; and (3) a %wor tier with timestamps for each word spoken. Using this information, CLAN programs can be used to do analyses such as mean length of utterance (in words or morphemes), words per minute, proposition density, lexical diversity, part-of-speech analyses, and grammatical complexity. Several CLAN commands have been developed to allow for composite profiles of transcript files, making language sample analysis fast, efficient, and reliable (Fromm et al., 2020). EVAL, for example, allows users to compare an individual participant's expressive language data to means and standard deviations

for comparison databases from AphasiaBank (Forbes et al., 2012). It could also be used to compare a participant's change over time following treatment. Another program, CORELEX, automatically measures the typicality of words used in response to a specific task. Based on core lexicon lists (words produced by 50% or more of a normative sample) from the AphasiaBank discourse protocol tasks as well as the Cookie Theft picture description task, it identifies and tallies the core lexicon words produced by an individual or group of individuals. Two additional examples, C-NNLA and C-QPA, automatically compute outcome measures from two well-established grammatical analysis systems, the Northwestern Narrative Language Analysis (Thompson et al., 1995) and the Quantitative Production Analysis (Rochon et al., 2000; Saffran et al., 1989). These systems provide detailed analyses of morphological content (e.g., number of regular and irregular plurals, possessives), general language measures (e.g., mean length of utterance, number of words and utterances), lexical variables (e.g., number of nouns, verbs, and pronouns), and structural analysis (e.g., embeddings, verb phrases, subject noun phrases) that quantify and characterize grammatical features of language in aphasia. Both systems require considerable training, linguistic expertise, and time for manual scoring. The automated commands allow for efficient and reliable analyses of large numbers of discourse samples and produce results in spreadsheet format. All these analyses run in a matter of seconds and produce results that can be viewed on the computer screen or in spreadsheet format. The website provides manuals and tutorial screencasts to help guide new CLAN users in these various activities.

## Research Applications

AphasiaBank data and methods have been used in over 500 publications, presentations, and theses. Normative data from the standard discourse protocol have been used to develop clinician-friendly discourse evaluation tools such as core lexicons and main concept checklists. Researchers have used the AphasiaBank database to examine a wide variety of relevant topics such as comparisons of different lexical diversity measures, the interaction of task and language, crosslinguistic differences, psychometric properties of microlinguistic measures, copula production, proposition density differences across aphasia types, factors contributing to fluency judgments, formulaic language use, word retrieval, word error analyses, semantic aspects of verb production, agrammatism, syntax, temporal overlap between gestures and speech, temporal and episodic organization of discourse, macrostructural analyses of narrative production, story grammar, linguistic mechanisms of coherence, prosody, recovery, crosslinguistic differences, treatment outcomes, and the effect of demographic and other clinical variables on language performance. The database has also been used in the machine learning and AI communities for a variety of purposes such as automatic classification of speech errors, automatic prediction of aphasia severity, and training a chatbot to assist with word finding problems. A complete bibliography (with links to articles) is available at the website.

A new assessment tool, the Famous People Protocol (FPP), was developed to elicit more discourse data from people with severe aphasia when it became clear that the standard discourse protocol was too difficult for the more severely impaired participants (Holland et al., 2019). This instrument uses a detailed interactive protocol and pictures of famous entertainers, world figures, sports figures, and former United States presidents. People with aphasia are encouraged to use any modality (e.g., gesture, pantomime, drawing, singing) or compensatory strategies (e.g., circumlocuting, self-cueing) to communicate. The FPP videos show how the examiner can identify individually appropriate strategies people with severe aphasia can use to communicate when speech is otherwise limited.

## Teaching Resources

The AphasiaBank teaching materials were developed to tap the rich resources available in the database and meet needs for academic and clinical instruction. *Grand Rounds* provides a guided tutorial on aphasia with short case histories for 16 individuals, 40 captioned video segments, descriptions of classic aphasia types, and questions to stimulate discussion about the language samples and potential treatment approaches. *Examples* offers very short, captioned video clips of common features from the connected speech of people with aphasia. These clips show features at the word level (e.g., anomia, phonemic paraphasia, semantic paraphasia, circumlocution, neologism, jargon, conduit d'approche), the sentence level (e.g., agrammatism, empty speech), and at the discourse level. *Classroom Activities* is a collection of ideas for exercises, several of which were contributed by AphasiaBank members. Examples of exercises include clinical assessment and treatment planning, measuring aspects of discourse (e.g., correct information units, mean length of utterances), using CLAN commands like EVAL, coding speech errors, examining main concepts in narratives, and comparing across aphasia types as well as across other disorders (e.g., traumatic brain injury, right hemisphere disorder).

Collaborative Commentary (CC) is a relatively new tool that has direct applications for research and teaching across all TalkBank databases (MacWhinney & Fromm, 2023). Briefly, it allows a researcher, instructor, or clinician to form a commentary group composed of multiple group members. Members could be students in a class, research staff in a lab, or colleagues. Group members can watch specific videos in the Browsable Database, follow along in the language transcript, and insert analytic comments or codes directly into the transcript tagged to a specific utterance or group of utterances. Those insertions are visible only to members of the group, or even just to the group leader, depending on the settings chosen by the group leader. Examples of how an instructor can use CC with the AphasiaBank database include having students: (1) identify behaviors typical of aphasia (e.g., paraphasias, circumlocutions, anomia); (2) differentiate between fluent and nonfluent behaviors; (3) look for the presence or absence of self-monitoring and self-correcting by speakers with aphasia; and (4) practice scoring responses to tests (e.g., BNT). All these skills are important for

developing competence in assessing, diagnosing, and planning treatment. Clinical researchers can use CC to measure and establish reliability with coding systems (e.g., gestures, agrammatism, paragrammatism, coherence) or scoring systems (e.g., fluency scoring for the picture description task on the WAB-R). CC has several different search options that allow users to search by specific codes or by specific group members and click to open the relevant segments in the Browsable Database. Once a relevant segment is open, feedback can be sent directly by email to the group member who entered that particular comment or code. This feature can be extremely useful for classroom exercises. Results of the CC searches also can be downloaded to spreadsheets, which can facilitate statistical analyses of coding or scoring data.

## Future Directions

There are many promising directions for new materials and uses in AphasiaBank. Here are some of the top priorities:

- Given the fact that so many of the transcripts in AphasiaBank are already linked directly to video, we have a rich database for analyzing the role of gesture in communication in aphasia. Linking up gesture coding to the Collaborative Commentary system would be particularly instructive, allowing us to see most clearly how gesture can supplement language in people with aphasia.
- Some sites have MRI scans for participants in AphasiaBank. With proper informed consent, it would be interesting to link imaging data to language analyses.
- Currently, AphasiaBank data is overwhelmingly from English. Adding more data from other languages could give us a clearer idea of the ways in which different languages are impacted by aphasia. To do this, we can rely on our new methods for applying Universal Dependency tagging to a wide variety of languages, using the Batchalign program (Liu et al., 2023).
- AphasiaBank can profit from further automation of the process of data collection and analysis. Web-based methods for data collection can allow us to reach a wider variety of participants and we are already using tools for automatic speech recognition to speed up the transcription process. Making these methods available to clinicians would greatly expand our ability to grow the database and improve its coverage.

## Conclusion

AphasiaBank started as a new shared database in the TalkBank system in 2007 with a small collection of videos and transcripts contributed by a small group of aphasiologists. The collection has continued to grow, currently into the thousands, with an almost equal number of files that use the standard discourse protocol and files that use other materials. The project has allowed for the development of new discourse measurement tools and a better understanding of communication in aphasia. Importantly, AphasiaBank has served as a model for the development of other shared databases for the study of communication in adult language: TBIBank, DementiaBank, and RHDBank. In total, these clinical banks have over 3000 members from around the world and from various fields (e.g., speech-language pathology, linguistics, psychology, neurology, computer science, information engineering, artificial intelligence, robotics, biomedical engineering). AphasiaBank and the related repositories have made critical contributions by providing freely available resources and tools for education and research to advance their respective fields and, ultimately, improve the quality of services to people who need them.

## Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grants R01-DC008524 (2022-2027 awarded to MacWhinney). The authors thank the people with aphasia who generously consented to and participated in the collection of data for the AphasiaBank database. The authors also thank the university faculty members and aphasia center clinicians who contributed data or allowed AphasiaBank researchers to seek participants from their programs.

## References

- 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.
- Forbes, M. M., Fromm, D., & MacWhinney, B. (2012, August). AphasiaBank: A resource for clinicians. In *Seminars in speech and language* (Vol. 33, No. 03, pp. 217–222). Thieme Medical Publishers.
- Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2020, January). Using AphasiaBank for discourse assessment. In *Seminars in speech and language* (Vol. 41, No. 01, pp. 010–019). Thieme Medical Publishers.
- Holland, A., Forbes, M., Fromm, D., & MacWhinney, B. (2019). Communicative strengths in severe aphasia: The famous people protocol and its value in planning treatment. *American Journal of Speech-Language Pathology*, 28(3), 1010–1018.
- Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston naming test* (2nd ed.). Pro-Ed.
- Kertész, A. (2007). *Western aphasia battery-revised*. PsychCorp.
- Liu, H., MacWhinney, B., Fromm, D., & Lanzi, A. (2023). Automation of language sample analysis. *Journal of Speech, Language, and Hearing Research*, 66(7), 2421–2433.

- MacWhinney, B., & Fromm, D. (2023). Collaborative Commentary for understanding communication disorders. *American Journal of Speech-Language Pathology*, 32(5S), 2580–2588.
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25(11), 1286–1307.
- Rochon, E., Saffran, E., Berndt, R., & Schwartz, M. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, 72(3), 193–218.
- Saffran, E., Berndt, R., & Schwartz, M. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, 37(3), 440–479.
- Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B. J., Schneider, S. L., & Ballard, K. J. (1995). A system for the linguistic analysis of agrammatic language production. *Brain and Language*, 51, 124–129.
- Wiederholt, J. L., & Bryant, B. R. (2012). *Gray oral reading test* (5th ed.). 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.

### List of Relevant Websites

- AphasiaBank. Retrieved from <http://aphasia.talkbank.org>.
- CHAT manual. Retrieved from <https://talkbank.org/manuals/CHAT.pdf>.
- CLAN manual. Retrieved from <https://talkbank.org/manuals/CLAN.pdf>.
- CLAN program. Retrieved from <https://dali.talkbank.org/clan/>.
- TalkBank. Retrieved from <https://talkbank.org/>.