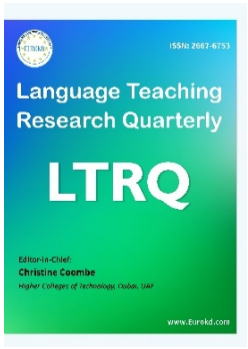


Language Teaching Research Quarterly

2024, Vol. 44, 31–40

# Augmenting Clinical Insights with Computing: How TalkBank has Impacted Assessment and Treatment of Speech and Language Disorders

**Nan Bernstein Ratner**

University of Maryland, College Park, United States

***Received*** *15 December 2023* ***Accepted*** *19 June 2024*

**Abstract**

Our purpose is to highlight the contributions of TalkBank initiatives to improved understanding of clinical impairments in adult and child speakers and examine remaining challenges and proposed solutions.We review the origins and development of TalkBank initiatives that have targeted a wide array of typical and atypical child and adult populations. In particular, we discuss how such sets of data have given rise to evaluation and validation of traditional measures used to appraise spoken language performance. The durable contributions of AphasiaBank and CHILDES archives are already evident in a body of published research that has re-evaluated, refined and reconceptualized how we evaluate and set therapeutic goals for speakers with expressive speech and language impairments. More recent archival initiatives, such as PhonBank and FluencyBank, are also making impacts. Beyond improvements in basic and applied science in communication development and disorders, archival data are also being used to test and improve accessibility for communicatively impaired speakers. TalkBank has transformed how research in communication disorders is conducted. It no longer relies on small, unshared research ventures that enable limited clinical impact or follow-up research inquiries. Rather, it has enabled large-scale, more generalizable research more likely to spur further research and enable more rapid translation to clinical practice.

**Keywords:** *TalkBank, Language Disorders, Aphasia, Fluency, Dementia*

**How to cite this article (APA 7th Edition):**

Bernstein Ratner, N. (2024). Augmenting clinical insights with computing: How TalkBank has impacted assessment and treatment of speech and language disorders *Language Teaching Research Quarterly*, *44*, 31-40. https://doi.org/10.32038/ltrq.2024.44.05

* Corresponding author.
E-mail address: nratner@umd.edu
https://doi.org/10.32038/ltrq.2024.44.05

## [1]Introduction

Imagine a set of Jeopardy questions of the following sort: *When was the Internet invented, and by whom? Who invented the first set of data analytics for personal computers and when*? And finally, *who invented the world's first public library*?[2]

Now we can phrase the questions a little differently. What was the first large scale system for sharing language research data across sites around the world and when was it started? Who guided the development of the first freely available language analysis software? And who established the first "public library" for the preservation and sharing of language data for basic research and clinical use?

The answers to these questions will all involve the work of Brian MacWhinney, and the scope of his efforts (now collected under the umbrella of the TalkBank initiative) across these diverse but related areas have indelibly changed the ways in which we approach collection and analysis of language data. In turn, these changes contributed to improved understanding not only of typical communication development in children learning a wide array of languages, but also of individuals with numerous profiles of communication disorders, including language loss due to stroke and traumatic brain injury (AphasiaBank and TBIBank), speakers with phonological disorders (PhonBank), and those with fluency disorders (FluencyBank).

We provide a short history to ground these accomplishments. In 1984, only a year following the "birth" of the World Wide Web, Brian MacWhinney and Catherine Snow anticipated the need for data preservation, sharing and standardization to advance knowledge in the study of child language development. Funded by the John D. and Catherine T. MacArthur Foundation, the founding project that would evolve into today's TalkBank initiative (TalkBank.org) was launched. In its early days, users needed to use pre-Internet protocols such as Telnet or physically mailed disks to utilize its resources. However, the Child Language Date Exchange System (or CHILDES) quickly revolutionized a field of study that had been limited by the resource-intensive nature of language transcription and analysis. From an initial set of 14 corpora, mostly from English language learners (MacWhinney & Snow, 1985; MacWhinney & Snow, 1990; MacWhinney, 1996; MacWhinney, 2014), CHILDES and the ensuing specialty focus "banks" discussed later in this article have grown to much more than 1.4TB of transcript data, with more than 5TB of accompanying media data (Bernstein Ratner & MacWhinney, 2020). Analysis of the collected data was facilitated by the development of the open-access, free software program Computerized Language Analysis (CLAN; MacWhinney, 2000) and subsequent improvements to automatically tag morphosyntax and grammar in English and additional languages (Sagae et al., 2010; MacWhinney et al., 2020),

[1]This paper is part of a special issue (2024, 44) entitled: In Honour of Brian MacWhinney's Five-Decade Contributions to Language and Psychology Research (edited by Zhisheng (Edward) Wen and Hassan Mohebbi).

[2]According to Wikipedia and other sources: January 1, 1983 is considered the official birthday of the Internet, when Bob Kahn and Vint Cerf developed the Transfer Control Protocol/Internetwork Protocol (TCP/IP) that enabled computer-to-computer communication. The first version of Microsoft office suite was distributed in October 1983 and is credited to former Xerox programmers, Charles Simonyi and Richard Brodie, who were hired in 1981 by Microsoft founders Bill Gates and Paul Allen. The world's oldest known library was founded sometime in the 7th century B.C. for the "royal contemplation" of the Assyrian ruler Ashurbanipal. The first public library was established by Asinius Pollio in Rome some time before 4 AD.

link media to transcripts (MacWhinney & Fromm, 2022), and enable automatic generation of transcripts using automatic speech recognition (ASR; Liu, et al., 2023).

The CHILDES initiative was followed by the creation of AphasiaBank (Forbes, et al., 2012; MacWhinney, et al., 2010; MacWhinney, et al., 2011; MacWhinney & Fromm, 2016). This in turn was followed by the establishment of PhonBank (Rose, et al., 2013; Rose & Stoel-Gammon, 2015), and means to create compatibility between Phon-annotated corpora and CHILDES data (Rose, et al., 2006). The success of focused repositories and accompanying software tailored for discipline-specific analysis spurred the construction of ASDBank for autism spectrum disorder (MacWhinney, 2019), FluencyBank (Bernstein Ratner & MacWhinney, 2018) for typical fluency development, stuttering and cluttering, TBIBank for traumatic brain injury (Elbourn, et al., 2023; Power, et al., 2020; Togher, et al., 2023), RHDBank for right hemisphere disorder (Minga, et al., 2021, 2022), and DementiaBank (Lanzi, et al., 2023). Today, all of these combine under the umbrella of TalkBank.org.

## CHILDES: From Typical Development to Clinical Implementation

First, it should be noted that CHILDES has become the *de facto* repository for language acquisition data from children learning over 40 different languages (not counting numerous distinct varieties of English and children learning numerous combinations of more than one language). These children vary from typically-developing monolingual and bilingual children to those with varying developmental disorders, such as autism spectrum disorder (ASD), hearing loss, Down Syndrome, focal brain injury and developmental language disorder of unknown etiology (MacWhinney, 1994). These repositories have, in turn, enabled identification of developmentally appropriate growth in child speakers across a wide range of communities. They have also informed the understanding of functional deficits that arise when this process goes awry or is slowed by inequalities in the experiences of children growing up in socioeconomically disadvantaged environments. It is beyond the scope of this article to list all of the seminal work that has been achieved across these areas, but as of the time that this manuscript was written in late 2023, the number of citations to CHILDES resources used in published research reports (as tallied by Google Scholar, using the term “Child Language Data Exchange System”) was roughly 5,500 articles, books and chapters.

CHILDES data have also been used to validate and refine measures for clinical assessment of children’s language for both diagnostic and goal setting purposes. The obvious utility of CLAN software for generating clinically-relevant indices of performance and progress has led to the incorporation of historical appraisal algorithms for child language development such as Developmental Sentence Scoring and the Index of Productive Syntax within “bundled” analyses [e.g., KidEval] for speech-language pathologists’ (SLPs’) use (Bernstein Ratner & MacWhinney, 2016, 2020, 2023; Garbarino, et al., 2020; Yang, et al., 2022). It has also been used to develop metrics for child language assessment that reduce the bias against speakers of non-mainstream versions of North American English inherent in older measures that are traditionally relied on by practicing SLPs (Overton, et al., 2021).

### *AphasiaBank: Closing a Loop between Assessment, Progress Monitoring and Best Practices*

Historically, the transition to clinically relevant corpora, protocols and computer-assisted analyses was made with the founding of AphasiaBank in 2007 (MacWhinney et al., 2011).

Designed from the start to collect a large body of language data from typical and language-impaired adults based on a standard discourse protocol, AphasiaBank rapidly gained both research and clinical prominence. The initiative established the utility of tasks such as telling the Cinderella story (MacWhinney, et al., 2010) and the Famous Faces Protocol (Holland, et al., 2019) for assessment (Fromm, et al., 2020; Fromm, MacWhinney & Thompson, 2020; Stark, et al., 2021), aphasia subtyping (Fromm, et al., 2022) and progress monitoring of clinical participants (Holland, et al., 2017). The establishment of online tutorials (Grand Rounds) for education of SLPs in training, automated discourse analysis tools such as EVAL (Forbes et al., 2012) and C-QPA (Fromm, et al., 2021), and creation of the Collaborative Commentary tool (MacWhinney & Fromm, 2023, see Fromm & Kowalski, this issue) have made the use of TalkBank resources in the evaluation, treatment monitoring and practitioner education a standard component of research, clinical practice and education activities involving adults with acquired language loss. Most recently, the AphasiaBank initiative has led to working group activities that have engaged in protocol development for tracking the outcomes of aphasia rehabilitation treatments (Brady, et al., 2020; Stark, et al., 2021; Kristinsson, et al., 2023).

*Old Data Repurposed to Answer New Questions*

Even from the beginning, the premise that old data could inform new questions was robustly confirmed. Beyond mere replication and extensions of results, entirely new research initiatives made use of existing corpora that had been gathered with entirely different research goals in mind. As a personally relevant case study, the Bernstein corpus had originally been compiled as part of a descriptive study to examine acoustic features of infant-addressed speech (IDS; Bernstein Ratner, 1984), and was one of the earlier donations to CHILDES. A decade later, the data were utilized, by Brent and Cartwright (1996), to demonstrate that infants could plausibly segment spoken language input to identify word boundaries in running speech, using phonotactic distributional regularities. The corpus has now become “the de-facto standard for evaluating segmentation models” (Goriely, et al., 2023) that seek to understand how infants manage to identify words in the earliest stages on language learning. The corpus has been further used to test models of unsupervised induction of grammar in machine language learning (Glushchenko, et al., 2019), a prospect not remotely envisioned during the original study, when data were collected on reel-to-reel analog tapes, and acoustically analyzed using a dedicated mainframe computer that had to be booted with punched paper tape.

Some of Dementia Bank’s roots reflect a different, but equally serendipitous history (Fromm, personal communication), and involve the Pitt corpus. Its cassette tapes from the 1980s had been stored in mushroom mines. The TalkBank initiative oversaw having the materials excavated, digitized, and transcribed around 2010, and they have since become a major source of work by researchers all over the world on automatic detection of mild cognitive impairment and dementia in adults, with almost 500 articles, conference presentations, and theses to date.

*FluencyBank: Basic Research and Clinical Education in Stuttering and Cluttering*

In 2018, the National Institutes of Health and National Science Foundation contributed to the establishment of FluencyBank, specifically developed to preserve annotated, media-linked data on typical and atypical speech fluency profiles in adults and children (Bernstein Ratner &

MacWhinney, 2018). In addition to preservation of historically invaluable longitudinal data from ground-breaking studies of the onset and outcomes of childhood onset stuttering, the initiative created cross-linguistically applicable codes for marking fluency within transcripts across language communities, and developed a research and clinical tool (FluCalc) that could generate fluency profiles for clinical and research participants in seconds. Its Voices of Adults/Children who Stutter/Clutter are being used in the preparation of clinicians to work with this historically underserved community. Both its Voices and research corpora are being used to update and revise traditional descriptions of how stuttering is distributed within speakers' conversational efforts (Warner, et al., 2023). Critical to decades of advisement to parents of young children who stutter, its holdings are now being used to re-evaluate the evidence base and effectiveness of such therapeutic guidance, some of it apparently well-founded (e.g., LaSalle & Wolk, 2023), while others appear to do little to alter children's short- or long-term fluency profiles (e.g., Burns & Bernstein Ratner, 2022; Garbarino & Bernstein Ratner, 2022; Godsey & Bernstein Ratner, 2022). FluencyBank holdings have also been deployed, as we describe further in this article, to foster automated assessment of fluency profiles, as well as remove barriers to use of voice assistants by people who stutter.

*Corpora Deployment to Perform Automatic Clinical Appraisal*

While considered the Gold Standard for clinical appraisal, language sample analysis (LSA) faces significant obstacles to routine use in most speech-language pathology settings (MacWhinney & Fromm, 2022). TalkBank efforts have been focused on the facilitation of in-depth, software augmented analyses for clinical work with both children and adults with communicative impairments, particularly those that can guide informed intervention goal-setting (see Garbarino, et al., 2020; Guo, et al., 2018).

In the process of doing this work, we have discovered that some traditional measures are quite sensitive to detection of language delays or differences, such as Mean Length of Utterance (MLU), while others, such as Type-Token Ratios (TTR) are not (Yang, et al., 2022; Bernstein Ratner, et al., 2024). Further, we have distinguished diagnostically sensitive measures from those that are superior in informing goal setting, a much more challenging prospect for therapists and teachers (Overton, et al., 2021). One of these is the Index of Productive Syntax (IPSyn; Yang, et al., 2022). We were able to refine this traditionally time-consuming measure to use shorter samples (saving time in the assessment process), and delineate which subscales have the highest value in diagnosis and therapy planning in the clinical process.

For adults with a range of clinical impairments, large databases gathered using standardized discourse protocols that include control subjects allow for the development of discourse measurement tools for main concept analysis and core lexicons. Such protocols enable more precise evaluation of both linguistic and conceptual impairments that may follow brain damage producing aphasia or other expressive communication limitations that accompany brain injury (Dalton & Richardson, 2015; 2019; Richardson & Dalton, 2020).

In the same vein, once archived and standardized in format, large corpora from TalkBank are now increasingly deployed to test and validate automated algorithms for early detection of communication disorders. These uses were not originally anticipated. For example, it now appears to be possible to recognize signs of mild cognitive impairment and dementia in spontaneous speech samples archived at DementiaBank ((e.g., Haider, et al., 2019; Luz, et al.,

2021; Liang, et al., 2022; Ye, et al., 2021). This advance could enable remote screening of subtle emerging symptomology, as well as further inform how components of communicative competence decline with the onset of dementia and related disorders.

*Corpora as a Mechanism to Remove Barriers to Communication*

From the outset, TalkBank was committed to free and open access to data and computing resources in the service of knowledge building (MacWhinney, et al., 2018; MacWhinney & Bernstein Ratner, 2022). Beyond this mission, however, TalkBank corpora have recently been used to test algorithms for automatic speech recognition (ASR) by speakers who often find themselves disenfranchised from the use of digital assistants because of speech disorders that impede accurate mapping of their intended messages. A major case in point has been the frustration of People who Stutter in their efforts to use speech assistant technology, such as Alexa (Robinson, 2022). Independent of the motivations for seeking NIH support to establish FluencyBank, tracking of citations to its use shows an overwhelming focus on software refinement to improve accurate tracking of stuttered speech by ASR algorithms (e.g., Al-Banna, et al., 2022; Mohapatra, et al., 2022). Tang et al. (2023) used AphasiaBank and DementiaBank corpora to both identify and reduce word error recognition rates for speakers with these significant handicapping conditions. We see this work progressing to other speaker communities, such as dysarthria and apraxia. In a similar vein, the notorious difficulty of transcribing typically-developing and articulation impaired child speech, which stands as an impediment to conducting analysis of many pediatric populations, may eventually be rectified by use of CHILDES and PhonBank repositories as grist for algorithm development and testing.

*TalkBank as a Paradigm Shift with Enduring Impacts*

As MacWhinney and Snow noted in 1985, it was once perfectly natural for researchers to follow small numbers of children, expend enormous amounts of resources analyzing their language behaviors. As a case in point, Brown (1973), broadly credited with one of the first efforts to create a metric for typical preschool language development, Mean Length of Utterance [MLU]) followed only three children, although his findings have been surprisingly durable across decades of further study, many of which have used CHILDES data (e.g., Yang, et al., 2022).

Single subject design in aphasia research was very common (Thompson, 2006). It was also perfectly natural for the data gathered in such investigations to sit in researchers' file cabinets until they retired, with little obvious place to archive or use such data further. We can contrast such early reports and the many similar bodies of data that have been lost to future generations with publications making use of TalkBank resources just since 2020: Yang, et al. (2022) revised and re-normed a clinical measure of preschool language development by using records from well over 1,000 children. Stark and Fukuyama (2021) were able to contrast typical adult microstructure during discourse from those seen in a variety of aphasic profiles using data from more than 500 speakers. Suarez-Rivera, et al. (2022) were able to identify age of acquisition of early core vocabulary, and conditions facilitating their learning in over 5,500 children whose data reside in CHILDES. Benway, et al. (2023) provide a PhonBank training corpus just for production and perception of [r] that contains over 100, 000 utterances.

Simply put, the world of language study can be easily demarcated, just as we do historically, by noting what we knew and how we learned it, BC (Before CHILDES) and AC (After CHILDES). An enduring debt is due to Brian MacWhinney, whose efforts have taken us this far; it's likely that our collective, collegial benefits from data sharing in language studies have only just begun to accrue.

**ORCID**

https://orcid.org/0000-0002-9947-0656

**Acknowledgements**

In addition to the invaluable contributions of Brian MacWhinney, both to my own work, and the global language science community, the author acknowledges the support of the TalkBank community over the years. The author wishes also to thank Davida Fromm for her always generous input and feedback.

**Funding**

Work cited in this article authored by Nan Bernstein Ratner and colleagues has been supported by the following organizations and grants: NIDCD R01DC015494 to Brian MacWhinney; National Science Foundation BCS1626300 to Brian MacWhinney and Nan Bernstein Ratner; and NIDCD R01DC017152 to Nan Bernstein Ratner.

**Ethics Declarations**

**Competing Interests**

No conflicting interests are declared; Nan Bernstein Ratner currently serves as Chair of the TalkBank Advisory Board.



## References

Al-Banna, A. K., Edirisinghe, E., Fang, H., & Hadi, W. (2022). Stuttering disfluency detection using machine learning approaches. *Journal of Information & Knowledge Management, 21*(02), 2250020. https://doi.org/10.1142/S0219649222500204

Benway, N. R., Preston, J. L., Hitchcock, E., Rose, Y., Salekin, A., Liang, W., & McAllister, T. (2023). Reproducible speech research with the artificial intelligence–ready PERCEPT Corpora. *Journal of Speech, Language, and Hearing Research*, 1-24. https://doi.org/10.1044/2023_JSLHR-22-00343

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother–child speech. *Journal of Child Language*, *11*(3), 557-578. https://doi.org/10.1017/S030500090000595X

Bernstein Ratner, N., Han, Y., & Yang, J. S. (2024). Should we stop using lexical diversity measures in children's language sample analysis? *American Journal of Speech-Language Pathology*, 1-16. https://doi.org/10.1044/2024_AJSLP-23-00457

Bernstein Ratner, N., & MacWhinney, B. (2016). Your laptop to the rescue: Using the Child Language Data Exchange System archive and CLAN utilities to improve child language sample analysis. *Seminars in Speech and Language*, 37, 2, 74-84. https://doi.org/10.1055/s-0036-1580742.

Bernstein Ratner, N., & MacWhinney, B. (2018). Fluency Bank: A new resource for fluency research and practice. *Journal of Fluency Disorders, 56*, 69-80. https://doi.org/10.1016/j.jfludis.2018.03.002.

Bernstein Ratner, N., & MacWhinney, B. (2020). TalkBank resources for psycholinguistic analysis and clinical practice. In *Development of Linguistic Linked Open Data Resources for Collaborative Data-Intensive Research in the Language Sciences*, 131.

Bernstein Ratner, N., & MacWhinney, B. (2023). Assessment and therapy goal planning using free computerized language analysis software. *Perspectives of the ASHA Special Interest Groups*, *8*(1), 19-31. https://doi.org/10.1044/2022_PERSP-22-00156

Brady, M. C., Ali, M., VandenBerg, K., Williams, L. J., Williams, L. R., Abo, M., ... & Wright, H. H. (2020). RELEASE: a protocol for a systematic review based, individual participant data, meta-and network meta-analysis, of complex speech-language therapy interventions for stroke-related aphasia. *Aphasiology*, *34*(2), 137-157. https://doi.org/10.1080/02687038.2019.1643003

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1-2), 93-125. https://doi.org/10.1016/S0010-0277(96)00719-6

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Burns, M., & Bernstein Ratner,N. (2022) It's just that simple: Parental language complexity in early childhood stuttering. In Wagovich, S. (ed.) Proceedings of the 2022 Joint World Congress on Stuttering and Cluttering Montreal, Canada. https://theifa.org/ifa-congresses-2/ifa-congress-proceedings/2022-jwc-proceedings.html

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology*, *24*(4), S923-S938. https://doi.org/10.1044/2015_AJSLP-14-0161

Dalton, S. G. H., & Richardson, J. D. (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*(1S), 293-320. https://doi.org/10.1044/2018_AJSLP-17-0166

Elbourn, E., MacWhinney, B., Fromm, D., Power, E., Steel, J., & Togher, L. (2023). TBIBank: An international shared database to enhance research, teaching and automated language analysis for traumatic brain injury populations. *Archives of Physical Medicine and Rehabilitation*, *104*(5), 824-829. https://doi.org/10.1016/j.apmr.2022.12.192

Forbes, M. M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language*, 33, 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, 1, 10-19). https://doi.org/10.1055/s-0039-3399499

Fromm, D., Greenhouse, J., Pudil, M., Shi, Y., & MacWhinney, B. (2022). Enhancing the classification of aphasia: a statistical analysis using connected speech. *Aphasiology*, *36*(12), 1492-1519. https://doi.org/10.1080/02687038.2021.1975636

Fromm, D., Greenhouse, J., Hou, K., Russell, G. A., Cai, X., Forbes, M., ... & MacWhinney, B. (2016). Automated proposition density analysis for discourse in aphasia. *Journal of Speech, Language, and Hearing Research*, *59*(5), 1123-1132. https://doi.org/10.1044/2016_JSLHR-L-15-0401

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

Garbarino, J., Bernstein Ratner, N., & MacWhinney, B. (2020). Use of computerized language analysis to assess child language. *Language, Speech, and Hearing Services in Schools*, *51*(2), 504-506. https://doi.org/10.1044/2020_LSHSS-19-00118

Garbarino, J., & Bernstein Ratner, N. (2022). What is the role of questioning in young children's fluency?. *American Journal of Speech-Language Pathology*, *31*(5), 2061-2077. https://doi.org/10.1044/2022_AJSLP-21-00209

Glushchenko, A., Suarez, A., Kolonin, A., Goertzel, B., & Baskov, O. (2019). Programmatic link grammar induction for unsupervised language learning. In *Artificial General Intelligence: 12th International Conference, AGI 2019, Shenzhen, China, August 6–9, 2019, Proceedings 12* (pp. 111-120). Springer International Publishing. https://doi.org/10.1007/978-3-030-27005-6_11

Godsey, A., & Bernstein Ratner, N. (2022). All in good time: A preliminary study of parent-child turn taking in early stuttering. In Wagovich, S. (ed.) *Proceedings of the 2022 Joint World Congress on Stuttering and Cluttering Montreal, Canada*. https://theifa.org/ifa-congresses-2/ifa-congress-proceedings/2022-jwc-proceedings.html

Goriely, Z., Caines, A., & Buttery, P. (2023). Word segmentation from transcriptions of child-directed speech using lexical and sub-lexical cues. *Journal of Child Language*, 1-41. https://doi.org/10.1017/S0305000923000491

Guo, L. Y., Eisenberg, S., Bernstein Ratner, N., & MacWhinney, B. (2018). Is putting SUGAR (Sampling Utterances of Grammatical Analysis Revised) into language sample analysis a good thing? A response to Pavelko and Owens (2017). *Language, Speech, and Hearing Services in Schools*, *49*(3), 622-627. https://doi.org/10.1044/2018_LSHSS-17-0084

Haider, F., De La Fuente, S., & Luz, S. (2019). An assessment of paralinguistic acoustic features for detection of Alzheimer's dementia in spontaneous speech. *IEEE Journal of Selected Topics in Signal Processing*, *14*(2), 272-281. https://doi.org/10.1109/JSTSP.2019.2955022.

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. https://doi.org/10.1044/2019_AJSLP-18-0283

Holland, A., Fromm, D., Forbes, M., & MacWhinney, B. (2017). Long-term recovery in stroke accompanied by aphasia: A reconsideration. *Aphasiology*, *31*(2), 152-165. https://doi.org/10.1080/02687038.2016.1184221

Kristinsson, S., Basilakos, A., Den Ouden, D. B., Cassarly, C., Spell, L. A., Bonilha, L., ... & Fridriksson, J. (2023). Predicting outcomes of language rehabilitation: Prognostic factors for immediate and long-term outcomes after aphasia therapy. *Journal of Speech, Language, and Hearing Research*, *66*(3), 1068-1084. https://doi.org/10.1044/2022_JSLHR-22-00347

Lanzi, A. M., Saylor, A. K., Fromm, D., Liu, H., MacWhinney, B., & Cohen, M. L. (2023). DementiaBank: Theoretical rationale, protocol, and illustrative analyses. *American Journal of Speech-Language Pathology*, *32*(2), 426-438. https://doi.org/10.1044/2022_AJSLP-22-00281

LaSalle, L., & Wolk, L. (2023). Adult recasts as fluency-facilitators in preschoolers who stutter: Evidence from FluencyBank. *Journal of Fluency Disorders*, *76*, 105971. https://doi.org/10.1016/j.jfludis.2023.105971

Liang, X., Batsis, J. A., Zhu, Y., Driesse, T. M., Roth, R. M., Kotz, D., & MacWhinney, B. (2022). Evaluating voice-assistant commands for dementia detection. *Computer Speech & Language*, *72*, 101297. https://doi.org/10.1016/j.csl.2021.101297

Liu, H., MacWhinney, B., Fromm, D., & Lanzi, A. (2023). Automation of Language Sample Analysis. *Journal of Speech, Language, and Hearing Research*, 1-13. https://doi.org/10.1044/2023_JSLHR-22-00642

Luz, S., Haider, F., de la Fuente Garcia, S., Fromm, D., & MacWhinney, B. (2021). Alzheimer's dementia recognition through spontaneous speech. *Frontiers in Computer Science*, *3*, 780169. https://doi.org/10.48550/arXiv.2004.06833

MacWhinney, B. (1996). The CHILDES system. *American Journal of Speech-Language Pathology*, *5*(1), 5-14.

MacWhinney, B. (1994). Using CHILDES to study language disorders. *Journal of Communication Disorders*, *27*(2), 67-70.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

MacWhinney, B. (2014). What we have learned. *Journal of Child Language*, *41*(S1), 124-131. https://doi.org/10.1017/S0305000914000142

MacWhinney, B. (2019). Understanding spoken language through TalkBank. *Behavior Research Methods*, *51*, 1919-1927. https://doi.org/10.3758/s13428-018-1174-9

MacWhinney, B., & Bernstein Ratner, N. (2022). Dynamic norming and open science. *Journal of Speech, Language, and Hearing Research*, *65*(3), 1183-1185. https://doi.org/10.1044/2022_JSLHR-22-00019.

MacWhinney, B., Feldman, H., Sacco, K., & Valdes-Perez, R. (2000). Online measures of basic language skills in children with early focal brain lesions. *Brain and Language*, *71*(3), 400-431. https://doi.org/10.1006/brln.1999.2273

MacWhinney, B., & Fromm, D. (2016). AphasiaBank as BigData. *Seminars in Speech and Language*, 37, 1, 10-22). https://doi.org/10.1055/s-0036-1571357

MacWhinney, B., & Fromm, D. (2016). Child Language Data Exchange System tools for clinical analysis. *Seminars in Speech and Language*, 37, 2, 3-73.

MacWhinney, B., & Fromm, D. (2022). Language Sample Analysis with TalkBank: An update and review. *Frontiers in Communication*, *7*, 865498. https://doi.org/10.1055/s-0036-1580743

MacWhinney, B., & Fromm, D. (2023). Collaborative Commentary for understanding communication disorders. *American Journal of Speech-Language Pathology*, *32*(5S), 2580-2588. https://doi.org/10.1044/2023_AJSLP-22-00385

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, *25*(11), 1286-1307. https://doi.org/10.1080/02687038.2011.589893

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology*, *24*(6-8), 856-868. https://doi.org/10.1080/02687030903452632

MacWhinney, B., Fromm, D., Rose, Y., & Bernstein Ratner, N. (2018). Fostering human rights through TalkBank. *International Journal of Speech-Language Pathology*, *20*(1), 115-119. https://doi.org/10.1080/17549507.2018.1392609

MacWhinney, B., Roberts, J. A., Altenberg, E. P., & Hunter, M. (2020). Improving automatic IPSyn coding. *Language, Speech, and Hearing Services in Schools*, *51*(4), 1187-1189. https://doi.org/10.1044/2020_LSHSS-20-00090

MacWhinney, B., & Snow, C. (1985). The Child Language Data Exchange System. *Journal of Child Language*, *12*(2), 271-295. https://doi.org/10.1017/S0305000900006449

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, *17*(2), 457-472. https://doi.org/10.1017/S0305000900013866

Minga, J., Fromm, D., Jacks, A., Stockbridge, M. D., Nelthropp, J., & MacWhinney, B. (2022). The effects of right hemisphere brain damage on question-asking in conversation. *Journal of Speech, Language, and Hearing Research*, *65*(2), 727-737. https://doi.org/10.1044/2021_JSLHR-21-00309

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. https://doi.org/10.1097%2Ftld.0000000000000244

Mohapatra, P., Pandey, A., Islam, B., & Zhu, Q. (2022, July). Speech disfluency detection with contextual representation and data distillation. In *Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications* (pp. 19-24). https://doi.org/10.1145/3539490.3539601

Overton, C., Baron, T., Pearson, B. Z., & Bernstein Ratner, N. (2021). Using free computer-assisted language sample analysis to evaluate and set treatment goals for children who speak African American English. *Language, Speech, and Hearing Services in Schools*, *52*(1), 31-50. https://doi.org/10.1044/2020_LSHSS-19-00107

Power, E., Weir, S., Richardson, J., Fromm, D., Forbes, M., MacWhinney, B., & Togher, L. (2020). Patterns of narrative discourse in early recovery following severe traumatic brain injury. *Brain Injury*, *34*(1), 98-109. https://doi.org/10.1080/02699052.2019.1682192

Richardson, J. & Dalton, S. (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

Robinson, J. (2022, August 23). *I stutter. This is what you're not hearing.* The New York Times.

Rose, Y., Hedlund, G. J., Byrne, R., Wareham, T., & MacWhinney, B. (2013). Phon: A computational basis for phonological database building and model testing. *Cognitive Aspects of Computational Language Acquisition*, 29-49.

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Wareham, T. (2006, April). Introducing Phon: A software solution for the study of phonological acquisition. *In Proceedings of the Annual Boston University Conference on Language Development. Boston University Conference on Language Development (Vol. 2006, p. 489).* PMID: 26933382; PMCID: PMC4769870.

Rose, Y., & Stoel-Gammon, C. (2015). Using PhonBank and Phon in studies of phonological development and disorders. *Clinical Linguistics & Phonetics*, *29*(8-10), 686-700. https://doi.org/10.3109/02699206.2015.1041609

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2010). Morphosyntactic annotation of CHILDES transcripts. *Journal of child language*, *37*(3), 705-729. https://doi.org/10.1017/S0305000909990407

Stark, B. C., Dutta, M., Murray, L. L., Bryant, L., Fromm, D., MacWhinney, B., ... & Sharma, S. (2021). Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. *American Journal of Speech-Language Pathology*, *30*(1S), 491-502. https://doi.org/10.1044/2020_AJSLP-19-00093

Stark, B. C., & Fukuyama, J. (2021). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. *Language, Cognition and Neuroscience*, *36*(5), 562-585. https://doi.org/10.1080/23273798.2020.1862258

Suarez-Rivera, C., Linn, E., & Tamis-LeMonda, C. S. (2022). From play to language: Infants' actions on objects cascade to word learning. *Language Learning*, *72*(4), 1092-1127. https://doi.org/10.1111/lang.12512

Tang, J., Chen, W., Chang, X., Watanabe, S., & MacWhinney, B. (2023). A new benchmark of aphasia speech recognition and detection based on E-Branchformer and multi-task learning. *arXiv preprint arXiv:2305.13331*. https://doi.org/10.48550/arXiv.2305.13331

Thompson, C. K. (2006). Single subject controlled experiments in aphasia: The science and the state of the science. Journal of communication disorders, 39(4), 266-291. https://doi.org/10.1016/j.jcomdis.2006.02.003

Togher, L., Elbourn, E., Kenny, B., Honan, C., Power, E., Tate, R., ... & MacWhinney, B. (2023). Communication and psychosocial outcomes 2-years after severe traumatic brain injury: development of a prognostic model. *Archives of Physical Medicine and Rehabilitation*. https://doi.org/10.1016/j.apmr.2023.04.010

Warner, H. J., Shroff, R., Zuanazzi, A., Arenas, R. M., & Jackson, E. S. (2023). Linguistic features of stuttering during spontaneous speech. *Journal of Fluency Disorders*, *78*, 106016. https://doi.org/10.1016/j.jfludis.2023.106016

Yang, J. S., MacWhinney, B., & Bernstein Ratner, N. (2022). The Index of Productive Syntax: Psychometric properties and suggested modifications. *American Journal of Speech-Language Pathology*, *31*(1), 239-256. https://doi.org/10.1044/2021_AJSLP-21-00084

Ye, Z., Hu, S., Li, J., Xie, X., Geng, M., Yu, J., ... & Meng, H. (2021, June). Development of the cuhk elderly speech recognition system for neurocognitive disorder detection using the dementiabank corpus. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6433-6437). IEEE. https://doi.org/10.1109/ICASSP39728.2021.9413634