Journal of Fluency Disorders 56 (2018) 69–80

Contents lists available at ScienceDirect

Journal of Fluency Disorders

journal homepage: www.elsevier.com/locate/jfludis

Fluency Bank: A new resource for fluency research and practice

Nan Bernstein Ratner[a,*], Brian MacWhinney[b]

[a] Department of Hearing and Speech Sciences, University of Maryland, 0100 Lefrak Hall, College Park, MD 20742, United States
[b] Department of Psychology, Carnegie-Mellon University, Pittsburgh, PA 15213, United States

1. Introduction

The National Science Foundation (NSF) and the National Institute on Deafness and Other Communication Disorders (NIDCD) have recently provided funding to establish FluencyBank (https://fluency.talkbank.org) as a new component of the larger TalkBank system (https://talkbank.org). The purpose of this article is to explain how FluencyBank will work to extend our understanding of the nature and development of typical and disordered fluency in both children and adults. To ground our discussion, we review the overarching organization of TalkBank and its component databases, and describe common features of TalkBank datasets. We then address the relation of FluencyBank to the overall TalkBank project. In doing this, we will discuss the specific funded research goals of FluencyBank. Finally, we will describe the resources that TalkBank and FluencyBank provide to fluency researchers, instructors, and clinicians with interests in typical and disordered fluency.

1.1. TalkBank

TalkBank is the world's largest open-access repository of data on spoken language. For an extensive summary of the technical aspects of TalkBank, see MacWhinney (in press). The TalkBank initiative began in 2000 as an extension of the Child Language Data Exchange System (CHILDES), established in 1984 by Brian MacWhinney and Catherine Snow (see MacWhinney & Snow, 1990). In the first years of development of the CHILDES system, most corpora were represented only in the form of computerized transcripts, although a few had accompanying media. Currently, new TalkBank corpora include transcripts linked to media (audio and video) on the utterance level, as well as extensive annotations for morphology, syntax, phonology, gesture, and other features of spoken language. All TalkBank corpora can be browsed online, or downloaded for additional annotation and analysis. As we note, many continue to be used to generate new research findings on an ongoing basis.

1.2. TalkBank features

An important principle underlying the TalkBank approach is that all data are transcribed in a single consistent format, called CHAT (MacWhinney, 2000). This format has been developed over the years to accommodate the needs of a wide range of research communities and disciplinary perspectives. TalkBank also makes available an extensive and free set of analysis programs, called CLAN, that rely on the fact that all TalkBank data use the CHAT transcription format. The CLAN programs and manuals, along with related morphosyntactic taggers that automatically insert part-of-speech and grammatical analysis into transcripts, are freely available and downloadable from the website at https://talkbank.org, for PC, Mac and Unix platforms.

The use of standard formats and codes is particularly important for the field of fluency studies. These conventions include a detailed set of fluency codes that permit automatic computational analysis across data from any language community. For example, blocking is marked with the Unicode symbol ≠ and sound iterations are marked by the Unicode character ↫. Entry of these characters is facilitated through keyboard shortcuts and a dropdown menu. These codes replace the various idiosyncratic codes

* Corresponding author.
E-mail addresses: nratner@umd.edu (N. Bernstein Ratner), macw@cmu.edu (B. MacWhinney).

https://doi.org/10.1016/j.jfludis.2018.03.002
Received 18 August 2017; Received in revised form 31 January 2018; Accepted 27 March 2018
Available online 29 March 2018
0094-730X/

developed in numerous separate labs to mark stuttering and other forms of disfluency. Use of these standards allows users to convert historical datasets to a standard transcription format with good fidelity of fluency marking. We discuss this issue in greater detail later in this article.

To facilitate use of CLAN analysis programs by researchers or clinicians who have employed other methods of transcription, CLAN includes a series of free programs to convert to CHAT from SALT (saltsoftware.com), Praat (praat.org), Phon (childes.talkbank.org/phon), ELAN (tla.mpi.nl/tools/elan), and LENA (lenafoundation.org) formats, among others.

1.2.1. TalkBank databases

TalkBank includes over a dozen specialized open-access language banks, all using the same transcription format and standards. These banks include CHILDES for child language acquisition, AphasiaBank for aphasia and other neurodegenerative language conditions, PhonBank for the study of phonological development and disorder, TBIBank for language in traumatic brain injury, DementiaBank for language in dementia, HomeBank for daylong audio- and video-recordings in the home, CABank for Conversation Analysis, SLABank for second language acquisition, ClassBank for studies of language in the classroom, BilingBank for the study of bilingualism and code-switching, and additional smaller banks that are under development. As noted in the Introduction, the most recently funded initiative is FluencyBank, for the study of the development of fluency and disfluency across the lifespan. Each of these components of TalkBank can be accessed from the overall TalkBank index page at https://talkbank.org.

The current size of the TalkBank text database is 800MB, with an additional 5TB of media data. New data are being added continuously. The majority of data in the various components of TalkBank are freely open for browsing, downloading and analysis. However, access to the research data in the clinical banks such as AphasiaBank and FluencyBank requires a password, and access to the data in HomeBank requires further attention to methods for safe-guarding the use of untranscribed day-long audio gathered in fully naturalistic settings.

These language banks have had a substantial impact on wide areas of research, as measured by the large number of publications that have used the data and programs. To date, CHILDES, which is the oldest and most widely recognized database, has been used to provide data for over 7000 published articles. PhonBank has been used in almost 500 articles, and AphasiaBank has been referenced in over 200 publications in only a decade since creation. To donate data to TalkBank can be rewarding: The contributors of TalkBank corpora benefit from bibliographic attribution and citation in these publications. To systematize the citation process, each corpus is assigned a DOI (Digital Object Identifier) number which users are required to cite. In addition, each corpus is described on a web page that includes links for downloading data and media, DOI information, corpus documentation, photos and contact information for the contributors, and articles to be cited when using the data.

1.2.2. Transcription conventions

TalkBank is an international and cross-linguistic project. Transcription is supported for all orthographies. The free CLAN program provides morphological/syntactic tagging (annotation for part-of-speech and grammatical analysis) for Cantonese, Chinese, Dutch, English, French, German, Hebrew, Japanese, Italian, and Spanish. These quickly and *automatically* insert information about each word's part of speech, grammatical function, and additional linguistic information (such as case, number, tense and gender marking for all words in the transcript). Both the availability of numerous free language parsers and automatic tagging (rather than hand-coding, as in alternatives such as SALT), make CLAN uniquely useful to both researchers and clinicians. To demonstrate, the simple command MOR (short for "morphological analysis") applied to a sentence such as the italicized ones below, produces the added information immediately below it, for one or multiple transcripts, almost instantaneously. The clinician or researcher only needs to type the speaker's intended words and to annotate perceived disfluencies; no linguistic knowledge is necessary to produce this level of analysis:

**SLP: do you think that intensive programs like the Hollins program might be more useful now that you're older?*

%mor: mod|do pro|you v|think pro:dem|that adj|intensive n|program-PL prep|like art|the n:prop|Hollins n|program mod|might cop|be qn|more adj|use&dn-FULL adv|now rel|that pro|you~aux|be&PRES adj|old-CP?

**CLI: and &-you_know I've read up on what I think < that > [/] that &-um &-you_know ↫ph↫how s:tuttering can s:ometimes be cured through &-um ↫ps↫ps:ychological counseling.*

%mor: coord|and pro:sub|I~aux|have v|read&ZERO adv|up prep|on pro:int|what pro:sub|I v|think pro:rel|that adv:int|how n:gerund|stutter-PRESP mod|can adv|sometimes aux|be part|cure-PASTP prep|through adj|psychological n:gerund|counsel-PRESP.

A full, linked, browsable transcript with accompanying media can be found for readers to view at https://fluency.talkbank.org/browser/index.php?url=Examples/Tom.cha We reiterate that the transcriber noted only what the speaker said, and noted disfluencies, since this level of detection is not capable of automation at present. Then the transcript was treated using the simple one-word MOR command, and resulted in the browsable version at the link above. We realize that the grammatical annotations in these transcripts can resemble gobbledygook for those not entrenched in linguistic analysis, and encourage readers to consult the listing of common abbreviations used in syntactic analysis starting on page 19 of the MOR manual at the Talkbank site (https://talkbank.org/manuals/MOR.pdf). We have included definitions for many of them in Appendix 1. The program automatically reads proper names and modifiers via capitalization, and disregards fluency notations that interrupt the citation forms of lexical entries in assigning part of speech and morphological inflections.

1.2.3. CLAN analytical programs

MOR is an excellent example of a powerful free analytical program supported by TalkBank (and thus FluencyBank). One might ask: What is the value of the MOR parsers and their resulting grammatical annotation? The answer is that this first level of analysis is critical to most analyses of spoken or written language. Even basic measures of language use require analysis of complex words (those consisting of multiple inflections), and running tallies of unique word forms over all words used in a speech sample. Because TalkBank morphosyntactic analyzers all use a parallel technology and output format, CLAN commands can be applied to each of these 10 languages for uniform computation of indices such as utterance length and complexity, vocabulary diversity, formulation errors, pause duration, and various measures of disfluency. For English, this development has enabled the development of two new and powerful clinical language analysis "bundles" (EVAL and KidEval), as well as a new fluency calculator (FluCalc), that can greatly improve and speed both clinical and research analysis of spoken and written language samples. We describe these in greater detail in Sections 1.4. and 4, below.

Critically, however, CLAN is an open, programmable system that users can adapt to their specific needs. CLAN includes a wide variety of user customizable search and analysis routines that have been extremely fruitful in evaluating theoretical claims and models. Such user-tailored evaluations have been important in understanding children's acquisition of morphology and syntax, in areas such as the English past tense (Marcus et al., 1992, Pinker & Prince 1988, MacWhinney & Leinbach 1991) or finite verb marking (Wexler, 1998, Freudenthal, Pine, Aguado-Orea, & Gobet, 2007). Emergentists (Pine & Lieven 1997) have used CHILDES data to explore theories of how children learn to use determiners, and generativists (Valian, Solt, & Stewart, 2009) have used the same data to argue for the presence of innate categories that guide children's acquisition of syntax. CHILDES data and CLAN programs have also been used to explore the contribution of adult language models and interaction profiles in children's language development (e.g., the many publications stemming from Snow, Tabors and Dickinson's Home-School Study of Language and Literacy Development [HSLLD] corpus, and the large number of investigations of the "learnability" of child-directed speech based on the Bernstein corpus). In these debates, and many others, the availability of a shared open database has been crucial in the development of analysis and theory, as have CLAN's powerful and flexible computing resources. We hope that the same benefits accrue to fluency researchers and clinicians.

1.3. Clinical extensions of TalkBank

After many years as primarily a research resource, TalkBank entered the clinical arena with the creation of the AphasiaBank initiative, funded in 2007 by the US National Institutes of Health, and directed by Audrey Holland and Brian MacWhinney (see summary in MacWhinney, Fromm, Forbes, & Holland (2011) and Forbes, Fromm & MacWhinney, 2014). AphasiaBank currently has 436 video recordings of people with aphasia and 226 non-aphasic controls performing the AphasiaBank protocol, which includes a uniform set of discourse, narrative, and processing tasks. Using the interactive EVAL program, researchers and clinicians can automatically compute in-depth language sample analysis across 32 measures, with reference values for typical adult and aphasic performance (in English) on each task, stratified by age, gender and diagnosis. AphasiaBank also includes smaller amounts of protocol data from Spanish, German, Italian, Mandarin and Cantonese. The framework of the EVAL program for "bundled" analysis of clinical data (rather than having to specify each type of analysis separately) was then extended to other age groups through the construction of KidEval for child language data. The goal of KidEval is to facilitate faster, more accurate and more informative child language sample analysis, by both researchers and practicing clinicians.

Child language sample analysis (LSA) for either clinical or research purposes can be quite time-consuming (Overton & Wren 2014). After spending hours of work to create a basic transcript, clinicians and researchers must then devote further time to compute measures such as Developmental Sentence Score (DSS; Lee & Canter, 1971; Long & Channell 2001; Cochran & Masterson 1995) or the Index of Productive Syntax (IPSYN) (Scarborough 1990). As a result, LSA is not widely used to inform child language assessment, let alone assessment of fluency clients (Bernstein Ratner & MacWhinney, 2016). Although we know that computer-assisted LSA can save time, and improve accuracy and depth of analysis (Heilmann, 2010; Price, Hendricks & Cook, 2010; Miller, 2001; Hassanali, Liu, Iglesias, Solorio, & Dollaghan, 2014), it is only infrequently used in practice. It is also under-exploited in research on children who stutter, when compared to standardized testing (see Ntourou, Conture & Lipsey, 2011); most studies have stopped with simple measures such as mean length of utterance (MLU). Unfortunately, MLU has limited ability to discriminate among child language profiles after the ages of 3–4 years, or an average MLU of 4.0 (Brown, 1973; Bernstein Ratner & MacWhinney, 2016), while other measures are more informative. Moreover, these measures require even more expertise and time expenditure if done by hand.

Fortunately, the use of free utilities such as CLAN, that can link transcription to the audio- or video-recorded record of the client's actual speech sample, can greatly improve the accuracy and informativeness of language sample analysis. When combined with the high accuracy of the automatic morphological parser for English, a simple typed transcript (with no need for overt, clinician coding of morphology, as in systems such as SALT) can be immediately annotated for morphological and grammatical features. These, in turn, can feed programs such as EVAL and KidEval. Each produces dozens of counts and proportions of a wide array of features relevant to language sample analysis. For example, both EVAL and KidEval compute: Mean Length of Utterance (in words and morphemes) of utterances pre-screened for eligibility using Brown's 1973 conventions; multiple alternative computations of vocabulary diversity (such as TTR, MATTR [moving average type-token ratio], number of different words [NDW] and VOCD), and indices of syntactic complexity (such as verbs/utterance). For many purposes, either EVAL or KidEval can be used for language analysis, depending upon the clinician or researcher's desired measures. Given typical concerns in acquired language disorder, EVAL, which was originally written for analysis of language use by people with aphasia, computes distribution of major parts of speech [POS]), while KidEval adds information listing 14 major morphemes in tracked in assessment of English child language development (more commonly

known as "Brown's morphemes), etc. For more details, readers are invited to consult the *Clinicians' Guide to CLAN* at the TalkBank site (https://talkbank.org/manuals/Clin-CLAN.pdf). The KidEval program also prescreens utterances using the sometimes complex and difficult-to-understand rules for inclusion in computations of MLU, DSS, and IPSYN (Sagae, Davis, Lavie, MacWhinney,& Wintner (2007)). It computes these measures automatically, avoiding the additional labor and computational error that will arise, if done manually. All EVAL and KidEval measures can also be computed for samples of written language, if they are composed in MS-Word or plain text. A somewhat abridged example of KidEval output is shown in Appendix 2. Moreover, these facilities are now available for an increasingly large number of languages other than English, including French, Spanish, Chinese, and Japanese with the extension to other languages in preparation. CLAN's computational power can greatly benefit clinical assessment, therapy planning, and measurement of therapeutic progress in clinical work in fluency disorders. Media-linked transcripts also preserve data in a single integrated, annotatable format that can easily facilitate *post hoc* hypothesis testing and data exploration.

2. Why we need FluencyBank

We think that it's important to note that FluencyBank development was supported both by the NIDCD, with its clinical focus on research, as well as the National Science Foundation, which has a focus on understanding typical speech/language production and comprehension. Because spoken language production is less amenable to controlled study than is comprehension, fluency is underrepresented in psycholinguistic research (Altmann, 2001; Fromkin & Bernstein Ratner, 1998). Disfluency in speech is not inherently bad and can be informative for listeners as well as for models of the speech production process. Devices such as filled pauses and simple repetitions can aid listeners' ability to process conversation easily and without error (Arnold ,Fagnano &Tanenhaus, 2003; Arnold, Kam, & Tanenhaus, 2007; Corley & Stewart, 2008; Ferreira, Lau, & Bailey, 2004; MacGregor, Corley, & Donaldson, 2010; Watanabe, Hirose, Den, & Minematsu, 2008).

However, as any reader of *JFD* is painfully aware, excessive or atypical disfluency can negatively influence perceptions of speaker typicality, nativeness, language competence, formulation effort, and truthfulness, with associated implications for educational, vocational and social progress, intelligence gathering and trial testimony (Arnold et al., 2007; Boltz, 2005; Bortfeld, Leon, Bloom, Schober, Brennan, 2001; Hartsuiker & Notebaert, 2010; Ozuru & Hirst, 2006). Even in typically developing (TD) children, there is growing evidence that fluency can be a relevant adjunct to standardized assessment findings in isolating expressive language difficulty (Boscolo, Bernstein Ratner & Rescorla, 2002; Guo, Tomblin & Samelson, 2008; Finneran, Leonard, & Miller, 2009; Steinberg, Bernstein Ratner, Berl & Gaillard, 2013). The study of disfluency is also a major emerging issue in second language acquisition (SLA) theory and practice (de Jong, 2008; Derwing, Munro, Thomson, & Rossiter, 2009; Yoshimura & MacWhinney, 2007). Finally, the study of disfluency is important for the development of algorithms for automatic speech recognition (Goldwater, Jurafsky, & Manning, 2010), since human listeners tend to be able to filter disfluencies, while machines find this task quite difficult.

The classic psycholinguistic models of speech production (Goldman-Eisler, 1958; Maclay & Osgood, 1959; Fromkin, 1973; Garrett, 1976; Dell & Reich, 1981; Dell & O'Seaghdha, 1992; Bock & Levelt, 1994; Indefrey & Levelt, 2000) focused on the analysis of "slips of the tongue" and hesitation phenomena in normal adult speakers. For children, there is landmark work on children's "slips" from Jaeger (2004) and Stemberger (1989). However, these types of error are relatively rare, whereas disfluencies in speech are ubiquitous and potentially informative, as adult studies illustrate. More recently, there have been studies of the ways in which fluency develops in typical speakers over the lifespan (e.g., Horton, Spieler, & Shriberg, 2010; Martin, Crowther, Knight, Tamborello 2nd, & Yang, 2010; McDaniel, McKee, & Garrett, 2010; Rispoli, Hadley, & Holt, 2008; Tumanova, Conture, Lambert, & Walden, 2014; Wagovich, Hall, & Clifford, 2009).

Although these various avenues of research have illuminated important aspects of language fluency and its development, we do not yet have a consistent set of analysis methods or a shared open-access database that can allow us to fully understand the development of fluency and disfluency in both normal and atypical speech, across a wide age range, and across language communities. FluencyBank is an effort to remedy this knowledge gap.

2.1. Specification of disfluency mechanisms and functions

Roughly 6% of spoken words in adult 'fluent speech' are disfluent (Fox Tree, 1995), and an even higher proportion are disfluent in child speech (Kowal, O'Connell & Sabin, 1975). Moreover, disfluency increases in tasks demanding more conceptual or linguistic effort (Kemper, Hoffma, Schmalzried, Herman, & Kieweg, 2011; McDaniel et al., 2010) in all speakers (including people who stutter), regardless of age or population. Although current models can capture the general conditions leading to non-fluency or their probable loci, they are less adequate in predicting the *type* of breakdown (e.g., hesitation, fillers, mazes), or their precise loci and distribution.

Readers of *JFD* need little reminding that stuttering remains one of the most prominent and puzzling disorders of language production. Accounts of its nature and cause are numerous and varied (Bloodstein & Ratner, 2008). The use of all possible converging forms of research data (experimental as well as naturalistic) can help us evaluate which accounts are strongest and which are weaker. We need to determine which fluency features are shared across groups and which are uniquely associated with specific language learning/production conditions and diagnoses. To do this, we needed to construct FluencyBank.

Large corpora derived from the speech of typical adults suggest that different disfluency phenomena reflect different speech stressors (Fraundorf & Watson, 2014). For example, the two most frequent English "fillers", *um* and *uh*, may differentiate syntactic as opposed to lexical retrieval difficulty (Clark & Fox Tree, 2002). Similarly, Rispoli, Hadley & Holt (2008)'s work with typically developing children has uniquely identified potential differences between stalls (which seem to reflect encoding difficulty) and revisions (which grow with grammatical development and self-monitoring skills).

However, for people who stutter, current models cannot predict why disfluency and stuttering phenomena types differ qualitatively as well as quantitatively, even within the same speaker. The Covert Repair Hypothesis (Postma & Kolk, 1993), based on Levelt and colleagues' WEAVER + + + model (Levelt, Roelofs, & Meyer, 1999), is a notable exception, although some of its predictions fail to account for both experimental and observational data (Brocklehurst & Corley, 2011) and WEAVER + + + is completely predicated on adult language competence, with little attention paid to fluency development (or disorder) in children.

To better test existing models of stuttering and other forms of disfluency, we need detailed longitudinal data collected across many children and adults using a consistent set of tasks, consistently transcribed so that they may be combined for automatic analysis, using a wide range of well-constructed computer programs. The central goal of FluencyBank is to construct this database and these programs. In this enterprise, FluencyBank seeks to bring together multiple communities interested in the development of fluency, including psycholinguists, speech technologists, speech pathologists, clinicians, second language researchers, and developmental psychologists. By creating a shared database and analysis programs, FluencyBank can stimulate networking opportunities across these multiple communities for examination of overlapping and specific concerns.

There is also a pressing need for research on the differential diagnosis of atypical fluency profiles. The terms "stuttering" and "disfluency" tend to be used interchangeably, resulting in frequent misidentification of bilingual and limited English Proficiency (LEP) children as children with stuttering (CWS), a disorder with serious lifetime handicaps (Sin, Beltran, & Howell, 2015, re-evaluating widely-publicized claims made by Howell, Davis, & Williams, 2009; Byrd, Bedore & Ramos, 2015). We need better specification of how fluency differs in monolingual speakers of different languages, and in bilingual speakers, and how these profiles clearly distinguish themselves from fluency disorders, such as stuttering.

2.2. The coding dilemma and potential solutions

To combine the strengths of multiple research groups, whether historical, current or future, we need to develop highly systematic standards and practices for fluency coding. Unfortunately, fluency coding has been subject to significant reliability problems (Brundage, Bothe, Lengeling, & Evans, 2006; Cordes, 2000; Cordes & Ingham, 1994, 1996; Hubbard, 1998; Lickley & Bard, 1998). This notorious variability in how disfluencies are perceived, coded and localized has led to significant concerns that two listeners may code the same speaker's sample as differently as a single coder might appraise samples over time. This problem extends to studies of normal disfluencies, producing confusions regarding processes and etiologies (compare Clark & Fox Tree, 2002 vs. O'Connell & Kowal, 2005). To illustrate this problem, Gottwald, Bernstein Ratner, Watson, Brundage, and Zebrowski, (2009) found that, in an analysis of the disfluencies in a two-minute, 120-word speech sample by dozens of coders, the count of disfluencies ranged between a low of 12 and a high of 40. Thus, it is unclear how much faith we can place in transcripts not linked to the actual recorded signal.

Coders agree better when samples are cut into short (e.g., 5 s) random intervals (Cordes & Ingham, 1996; Cordes, Ingham, Frank, & Ingham, 1992; Ingham, Cordes, & Finn, 1993), out of context, and when the task is only to render a binary judgment as either fluent or stuttered, rather than to identify each single disfluency. This method improves reliability of fluency counts, but is quite unsuited to most clinical needs, as well as any research that transcends simple tallies, thereby excluding most linguistic studies of normal or atypical fluency behaviors in context. It is not surprising, therefore, that this system has not been widely adopted.

The coding of disfluencies suffers not only from reliability problems, but also from workload problems. Marking the exact duration of unfilled pauses in a transcript can take even more time than creating the initial transcription. Identifying the timing of segment repetitions, drawls, and retraces requires still further effort. However, without this coding, we cannot properly characterize patterns of disfluency. The creation of FluencyBank offers a fundamental solution to this dilemma. There are five components of this solution:

1. FluencyBank methods link transcription directly to the audio record, thereby tightening the linkage of codes to data. CLAN "chunks" the original signal in small portions, and enables looping of the segment while typing the transcript, making transcription faster and more accurate.
2. FluencyBank has a consistent system for fluency coding that is computationally compatible with all CLAN computing utilities. A single transcript can be analyzed for fluency, as well as linguistic, phonological and acoustic properties.
3. By encouraging data-sharing, FluencyBank will be able to create a large inventory of well-transcribed and well-coded data linked to audio.
4. Because CLAN links directly to Praat, it is possible to facilitate acoustical analyses of spoken language, and to create a core set of "gold standard" transcriptions of disfluency patterns across different speaker populations.
5. Using these gold standard transcriptions, we can eventually train automatic speech recognition (ASR) systems such as SpeechKitchen (Metze, Riebling, Fosler-Lussier, Plummer, & Bates, 2015) to perform automatic diarization and segmentation on input recordings. We have shown that this method is particularly powerful when participants are asked to repeat target sentences or passages. Their productions can then be segmented on the level of the individual phoneme using acoustic models, rather than word-based models. This diarization then provides us with exact time values for the beginning and end of each sound segment and each unfilled pause. Although ASR methods are still imperfect, their accuracy has improved markedly in recent years through the introduction of algorithms such as "deep learning" and "end-to-end" processing.

Automatic diarization and segmentation will address many parts of the workload problem. However, we will still need human input and further analysis to dig more deeply into the coding reliability problem and basic issues in the study of disfluencies. By grounding disfluency coding on acoustic features quantified in Praat, on the basic phonological facts as characterized in Phon

(TalkBank's phonological analysis program), and the basic lexical and morphosyntactic facts as characterized by CLAN, we can achieve much greater levels of coding consistency for the behavioral features of stuttering, which can then be linked, in turn, to discoveries in speech-motor control and brain imaging. On the basis of such converging evidence, we hope to create a data-based understanding of patterns in disfluency. We can then link these methods to the examination of data across speakers with alternative clinical profiles and ages, performing various linguistic tasks. This should help us not only clinically, but by assuring better description of participants in research studies.

2.3. Distinguishing stuttering from language encoding difficulty

There is growing evidence that developmental disorders are accompanied by visibly atypical fluency profiles and slowed rate of language production; in children, these include Late Talkers (LT), Specific Language Impairment (SLI) (Boscolo et al., 2002; Guo et al., 2008; Hall, Yamashita, & Aram, 1993; Smith, Hall, Tan, & Farrell, 2011; Vasic & Wijnen, 2004), reading impairment (RI) (Smith, Roberts, Smith, Locke, & Bennett, 2006; Smith, Smith, Locke, & Bennett, 2008; Hester & Pellowski, 2014) and Autism Spectrum Disorder (ASD) (Lake, Humphreys, & Cardy, 2011; Scaler Scott, Tetnowski, Flaitz, & Yaruss, 2014; Sisskin & Wasilus, 2014). Notably, Sisskin & Wasilus observe that atypical disfluency has been "lost in the literature, but not on the caseload." A survey of over 200 SLPs (Sisskin & Bernstein Ratner, 2015) reveals widespread clinical concern over referrals that parents/teachers make for stuttering that do not fit the full diagnostic criteria for stuttering. These fluency profiles bear only superficial resemblance to stuttering, and there is absolutely no evidence that *any* respond to stuttering therapy (although they may respond to treatment that would almost certainly aggravate fluency in a PWS, as shown by Sisskin & Wasilus, 2014). Differential diagnostic questions will require us to examine transcripts linked to high quality media. First, is there a typical developmental progression in the frequency of utterance disruptions, and their types and loci? To what degree is this progression influenced by syntactic demand? How do length and complexity of target utterances factor into the rate of disfluency in both spontaneous and elicited utterances? Do disfluency profiles change over the course of early development, as a function of age and/or gender? Adults are known to mark syntactic and lexical demand by use of distinct markers (Clark & Fox Tree, 2002; Corley & Stewart, 2008; Fraundorf & Watson, 2014). When does this profile emerge? To date, only one study has attempted to isolate differential functions of fluency disruptors in 3- to 4-year-old children (Hudson Kam & Edwards, 2008), and that study did not directly address the basic developmental question.

A related question asks how bilingualism impacts fluency. Worldwide, most children speak two or more languages, and we need to understand how the need to control the use of multiple languages can impact the growth of fluency (see Skehan, 2009). In addition, there is emerging evidence that bilingual and limited English Proficiency (LEP) children are frequently misidentified as children who stutter (CWS; Sin et al., 2015, Byrd et al., 2015; Howell et al., 2016; Schmid & Fägersten, 2010). Of course, stuttering is a disorder with serious lifetime handicaps, whereas second language acquisition is a normal process. To misclassify a bilingual child as a stutterer can be educationally and socially handicapping. The rich analytic and media-linked capacities of TalkBank and the CLAN programs can aid us in pulling apart the characteristic verbal patterns of each of these potentially distinct disfluency profiles.

2.4. Identifying pathways of fluency development and disorder

Both language delay and stuttering have significant recovery patterns. More than half of late talkers (LTs) and children who stutter (CWS) achieve normal diagnoses by age 5 (Bloodstein & Bernstein Ratner, 2008; Dale, McMillan, Hayiou-Thomas, & Plomin, 2014). Yet, efforts to identify *which* children will recover have relied primarily on prediction from familial history. To date some prognostic cues have emerged, starting with the groundbreaking work of Yairi, Ambrose, Paden & Throneburg (1996), using longitudinal data from the Illinois Stuttering Project (numerous publications summarized in Yairi & Ambrose, 2004). These include profiles of disfluency over time (also identified by Ryan, 2001), and initial phonological/articulatory skills (e.g., the Purdue cohort (Smith & Weber(Fox) and colleagues [e.g., Spencer & Weber-Fox, 2014]), and others [e.g., Kloth, Kraiimaat, Janssen & Brutten, 2000]). Conture and colleagues (e.g., Louko, Edwards & Conture, 1990) have also reported phonological skill differences in numerous publications, primarily between CWS and typical peers. Other studies have identified potential language profiles (e.g., Yairi & Ambrose, 1999; the Illinois, Purdue, Iowa and Syracuse/Vanderbilt cohorts have each issued reports).

It is impossible to do due diligence to this wide body of work in this article. However, the fact remains that even distinguishing between profiles of typically fluent and stuttering children may require meta-analysis of numerous smaller reports (Ntourou et al., 2011). The combined power of larger cohorts and converging evidence from longitudinal and cross-sectional analyses offer one means for identifying additional factors affecting risk and recovery. These include going beyond standardized assessments to examination of spontaneous communication profiles, and even potential differences in the communicative profiles of CWS and their parents (e.g., Kloth et al., 2000; Miles and Ratner, 2001).

To place the need for combined data sets into context, for any individual research endeavor in childhood stuttering, even a large, federally-funded initiative, the demographics of stuttering persistence and recovery pose a significant statistical challenge, with only a small number of children seen in any study likely to progress to persistent stuttering (~20%). Moreover, in non-longitudinal samples, even statistically meaningful findings that distinguish stuttering from typically fluent children may tell us more about profiles of the larger majority of children who experience only passing difficulty with speech fluency, rather than inform the initial stages of the life-long communication disorder that will challenge the smaller proportion of this cohort. Statistical power and generalizability of data analysis were among the primary motivations for CHILDES, AphasiaBank and PhonBank, and are even more strongly desirable in stuttering research, where frequent remission combines with relatively small sample sizes for most published studies to severely limit what can be learned about pathways and predictors in early childhood stuttering.

By combining datasets in CHAT format, we will be able to determine which features of early expressive language predict persistence or recovery. Our preliminary work with CWS, taking advantage of uniform CHAT transcription conventions, and CLAN software programs across two large cohorts strongly supports original findings from the Illinois Stuttering Project (ISP; Yairi et al., 1996). The ISP found that the children who are most likely to persist in stuttering have lower language skill close to onset, as measured by language screening tests. Using growth-curve analyzed data from age-/gender-matched peers, persistence appears to be signaled by delays in aspects of language growth over time, relative to peers who have never stuttered, or who have recover (Leech, Bernstein Ratner & Weber, 2017; Chow, Spray, Bernstein Ratner & Chang, 2015). Using additional data, we can grow the power of this and similar analyses by incorporation of additional cohorts; the availability of over 2500 typically-developing English-speaking children in the CHILDES database that can be selectively age-, gender- and SES-matched can additionally increase the reference sample of typically-fluent children against smaller stuttering research projects' findings.

Finally, we note that the study of the development of fluency in childhood has important broader implications for other segments of cognitive science. The study of disfluency is an important issue in second language acquisition (SLA) theory and practice (de Jong, 2008; Derwing et al., 2009; Mora, 2006; Segalowitz, 2010; Yoshimura & MacWhinney, 2007), where there is the same concern about linguistic knowledge and planning windows as for first language learners. As noted earlier, rich new data on the development of fluency would also constitute a major challenge and stimulus for models of speech production, which currently have no real developmental component.

3. The components and resources of FluencyBank

As with past TalkBank initiatives, the FluencyBank initiative emerged out of several years of discussion among researchers in typical speech/language development, stuttering, language disorders and second language acquisition. As with the other TalkBank projects and sites, its primary components are a database, transcription and analytical tools, and teaching resources.

3.1. Database

FluencyBank is the first multi-investigator effort to provide data sharing for describing the development of fluency. It seeks to track children's productions from first words to adult-like utterances, and to track the profiles of atypical fluency development across a range of ages and language communities. Obviously, we could tackle this problem by simply doing more new studies. However, Justice, Breit-Smith, and Rogers (2010) emphasized the critical need to use existing data to more fully exploit our research capacity in developing relevant advances in understanding and treating atypicalities in children's communication development. Simply put, we should not be gathering entirely new data when old data, particularly those that can be combined to provide greater statistical power and generalizability, will inform some of our questions.

FluencyBank will also provide a permanent home and data-sharing access point for speech samples collected as part of projects involving adult stuttering and other communication disorders. Although FluencyBank holdings are not yet extensive, clinical and research groups around the world have made commitments to contribute large amounts of both older and newer data. We need to preserve our valuable, landmark research data for future generations. Our priority in working with *pre-gathered data* from other sites is to convert and link labs' existing data to the CHAT format and linkage and then to *return these data to the original teams for potential additional research analyses,* before posting the data for use by others. We return data first to contributors because the utilities in CLAN and the linkage to utilities such as PRAAT and Phon enable researchers to make additional publishable use of their existing data. These researchers then commit to eventual donation of the data to FluencyBank, under conditions of password protection and user agreements.

Funding to create and support the database is currently about 18 months old. Thus, FluencyBank is a toddler enterprise. To date, we have converted and posted all primary speech sample data from the fluency studies of Bernstein Ratner and colleagues (these include the studies co-authored with Miles, Wagovich/Silverman, Hakim, and the POLER study (Strekas, Bernstein Ratner, Berl, & Gaillard, 2013). We are actively converting numerous historical and current data sets from major laboratories in stuttering and typical fluency to the CHAT format and linking to original media, while preserving any existing confidentiality arrangements. Because of its unique character, it may take a few years for the FluencyBank to grow in size and scope to resemble its other TalkBank siblings. On the other hand, we have begun to achieve the goal of constructing efficient and consistent methods for fluency coding and analysis.

3.2. Contributing to grow the database

FluencyBank facilitates data contributions in two ways. First, using automatic conversion programs, we can transform non-CHAT data into CHAT format for consistent analysis. Many researchers in our field have used SALT to transcribe data. We can turn these data into CHAT rather easily, and then link the data to either audio or video records (something not possible in SALT). Second, we facilitate contributions by adapting data access to align with IRB requirements. Unless existing video data were gathered with explicit consent for open public use, we extract the audio from the video, using FluencyBank funding, and place the data behind password protection. To de-identify data, we eliminate all references to personal identifiers and addresses in both the transcripts and the audio. Proper names are replaced by generic placeholders, such as *Childname* in transcripts, and the segment of the linked audio with that reference is silenced. Increasingly, researchers are now using specialized consent forms with graduated permission to archive data in various formats. Examples of such suggested templates for obtaining participant support to archive research data are provided at

https://talkbank.org/share/irb/ The site also posts a post-hoc consent form for use at the completion of a longitudinal study, after participants have a greater understanding of the nature of their collected samples.

3.3. Teaching with FluencyBank

Over the years, TalkBank has provided valuable support for student education and training. Starting with CHILDES, numerous contributors have developed teaching activities to exploit some of the open access data sets for teaching typical language development in both clinical and non-clinical coursework. These methods have figured both in published research (Sokolov & Snow, 1994) and materials available from the CHILDES website. Because of its standard protocol and accompanying multimedia, AphasiaBank has generated a well-exploited set of multi-media educational activities, including Grand Rounds demonstrations of the characteristics of different types of aphasia. A Google Scholar search will show hundreds of citations and references to its assets.

FluencyBank has likewise made a commitment to provide educators with access to materials that can improve the quality of clinical education in fluency disorders. The bank already includes an open-access, IRB-approved set of interviews with more than two dozen adults who stutter, specifically designed to provide multi-media samples (with accompanying diagnostic instrument examples); these do ***not*** require a password for access, and are freely downloadable by both instructors and students. Two projects, called Voices of People who Stutter and Voices of People who Clutter, were initiated with the assistance of the National Stuttering Association and the International Cluttering Association, who helped to vet interview questions and accompanying questionnaires contributed by volunteers. Both projects are open-ended, and will continue to recruit and add to the teaching materials site.

How might University instructors make use of these teaching materials? For example, the Voices of PWS project uses a common set of interview questions, accompanied by a completed, but unscored copy of the speaker's *OASES* questionnaire, and the "Friuli" reading passage from the *Stuttering Severity Instrument-4 (SSI-4)* (both with permission of the instrument publishers). This permits students to complete an *SSI-4*, if their clinic has copies of the instruments and their scoring data. Transcripts are roughly transcribed in CHAT and linked to video media, but disfluencies are not marked, to permit guided scoring of fluency and concomitant features of stuttering. Interview questions were also selected to maximize discussion of how behavioral, affective and cognitive components of chronic stuttering may differ among PWS. See https://fluency.talkbank.org/teaching.html for a list of suggested teaching activities; we solicit and welcome additional suggestions from instructors, as do the other teaching sites at Talkbank). Given numerous requests after the first year by users, as well as requests to contribute from parents and children who stutter, the University of Maryland also plans to obtain IRB approval for collection of data from children with fluency disorders in the coming academic year.

4. FluencyBank analysisTools

FluencyBank is working to provide new, more accurate methods for fluency analysis based on transcription-media linkage. These include preconfigured analysis tools, and support for program interoperability (movement between transcription programs, such as CLAN, SALT, ELAN, etc., and other analysis platforms, such as Praat and Phon).

4.1. Preconfigured analyses

The free FluCalcprogram was released as a part of the open-access CLAN program in June 2017. This program provides preconfigured analysis of raw and proportioned counts of individual types of typical and atypical (SLD) disfluencies (e.g., prolongations, blocks, unfilled and filled pauses), average repetition unit (iteration) frequency for word and part-word repetitions, overall counts and proportions of typical vs SLD behaviors, together with a weighted SLD score based on the work of Yairi and Ambrose (2004). These values can be based on words or syllables (currently for English only), as is more traditional in the stuttering literature for historical reasons. We are currently exploring linkage of FluencyBank utilities with those in Phon, which can automatically perform phonological analysis of the sample. A sample output of the FluCalc program, run on the same children as in Appendix 2 (KidEval), is shown in Appendix 3. Both printouts use abbreviations to save space; a guide to the annotated transcript shown earlier, as well as the KidEval and FluCalc spreadsheets, was provided in Appendix 1.

4.2. Program interoperability

Although the basic technology to link media to transcripts emerged two decades ago, the full computational exploitation of this benefit has been much more recent. Currently, TalkBank files interact with the Phon program for phonological analysis (Rose, Hedlund, Byrne, Wareham, & MacWhinney, 2007) and the Praat program for acoustic analysis (Boersma & Weenink, 1996). However, a great deal of programming work is still needed to maximize this linkage for the automatic and semi-automatic analysis of disfluencies. In this area, FluencyBank hopes to break new ground in terms of computational methods.

4.3. Morphosyntactic analysis

As noted earlier, TalkBank currently supports utilities for fully automated morphosyntactic analysis of transcriptions in 11 languages (in alphabetical order: Cantonese (yue), Chinese (zho), Danish (dan), Dutch (nld), English (eng), French (fra), German (deu), Hebrew (heb), Japanese (jpn), Italian (ita), and Spanish (spa). For English, accuracy of fully automated morphological tagging is currently estimated at 95% (Huang, 2016), with excellent coding of syntactic function. Because a simple typed and fluency coded

transcript can now directly interface with this automated morphosyntactic analysis, FluencyBank utilities can provide new potential for linguistic and grammatical analysis of the correlates of dis/fluency, particularly in early stages of language learning and stuttering.

4.4. Fluency tagging

Large cohorts of very young stuttering children, bilinguals and late talkers will provide a new challenge for algorithms built to automatically tag speech samples for fluency, which to date have mainly been tested on typically fluent adult speech (Bakker, 1999; De Jong & Wempe, 2009; Horton et al., 2010; Schiel, Heinrich, & Barfüßer, 2011). Tool development to train systems that assign syllable peak profiles to measure speech rate automatically is part of the ongoing technological component of FluencyBank.

5. Current work and work in progress

5.1. User support

Thanks to longstanding NIH and NSF funding, the FluencyBank and TalkBank projects enable a large array of user support services, such as lab- or project-specific instruction in file linking and transcription, use of programs for research and clinical purposes, email support services, and trouble-shooting of problems with data or program use. We currently provide three free manuals for typical user purposes: the CHAT transcription manual (https://talkbank.org/manuals/CHAT.pdf), and the CLAN manual for the analytic programs (https://talkbank.org/manuals/CLAN.pdf). In addition, there is an SLP guide to CLAN (https://talkbank.org/manuals/Clin-CLAN.pdf), which is an abridged user manual for "quick start use" by practicing clinicians and clinical researchers who want to use the clinical "bundle" programs for adult and child language sampling (EVAL, KidEval), and fluency assessment (FluCalc). The TalkBank site also hosts a large number of short screencasts designed to illustrate specific aspects of transcription and analysis (see https://talkbank.org/screencasts/). In addition, any interested user can request a personal Internet-facilitated tutorial from the authors.

5.2. Data collection

In addition to data from the majority of the first author's published research, the FluencyBank grant from NIH supports our collection of new longitudinal data from children who stutter, typical peers, late-talking children and preschool children acquiring both English and Spanish. Our goal is to identify common and distinct features of the disfluency profiles seen in these child populations. We are currently working with numerous labs around the world to convert existing research transcript data to TalkBank-compatible formatting, and to link audio media to these files. This can be a somewhat time-consuming process, as most fluency labs either developed their own idiosyncratic transcript conventions, or used SALT. Neither of these typical options used standard codes for fluency behaviors, or linked media to transcripts. At least two of these influential data archives, both of which represent major, federally-funded, large scale longitudinal studies of CWS, should be publically available by the end of 2018. More critically, we have been providing assistance to ongoing fluency projects in numerous parts of the world to shift their new data collection to TalkBank-compatible transcript conventions. This should greatly accelerate research data holdings in the next few years. What kinds of data will be valuable? All data. It is impossible to know in advance how existing data can be applied to new research problems. Our experience with CHILDES showed that some data evolved to become major influences on research across diverse disciplines, to a level that could not have been predicted when they were contributed. For example, our own Bernstein Ratner corpus (one of the first in the CHILDES archive) produced fewer than a half dozen articles by its author based on its original set of questions. To date, it has spurred almost 300 research projects by other laboratories, all fully attributed.

We have also acquired additional teaching materials in the process of conversion and posting, most recently a collection of almost 100 small media clips from Glen Tellis at Miseracordia University. Like many contributions we are discussing with researchers, these clips did not have accompanying transcripts, which are being generated with the federal grant support to the FluencyBank project. We continue to solicit additional teaching as well as research materials, and welcome inquiries about potential contributions from *JFD* readers. We especially welcome contributions from languages other than English. We are currently working with datasets in French and Dutch for eventual inclusion in FluencyBank, in addition to our funded Spanish language work, but welcome more linguistic diversity in our holdings.

6. Conclusion

Understanding the bases of fluency, disfluency and stuttering is central to both theory and clinical practice. As demonstrated by the success of CHILDES, AphasiaBank, PhonBank, and HomeBank, data sharing speeds the discovery of knowledge in a discipline and provides power to analyses as well as ensuring greater generalizability of findings. The US NIH currently requires applicants to specify a data sharing plan for all grant applications for a reason. Notably, CHILDES, which has resulted in over 7000 published articles, has shown how data sharing can reshape the academic landscape of a research area. AphasiaBank, which has established a fixed protocol that is now being applied by self-enrolled supporters, shows how a clinical discipline can combine data-sharing with tightly defined data collection and analysis to enable more reliable differential diagnosis of multiple types of language dissolution following disease and trauma, as well as measure response to intervention. PhonBank has immensely enlarged the body of data

available to understand the complex nature of phonological development and disorder over childhood. HomeBank has further extended TalkBank to provide securely curated records of the exploding number of studies utilizing daylong recording technology, such as those collected using the LENA system. FluencyBank joins this TalkBank community with the goals of preserving classic data in our field, enabling new shared study of large numbers of linguistically diverse speakers, and improving the research and clinical base in fluency development and disorder. Initiatives such as those reported by Bauman, Hall, Wagovich, Weber-Fox, & Bernstein Ratner, 2012 and Leech et al., 2017 Leech et al. (2017) offer examples of such collaboration. We invite readers to explore and use this new resource.

Acknowledgments

This work was supported by the National Institutes of Health [NIDCD: 1 R01 DC015494-01] and the National Science Foundation [BCS-1626300/1626294].

Appendices: Supplementary Data.

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jfludis.2018.03.002.

References

Altmann, G. T. (2001). The language machine: Psycholinguistics in review. *British Journal of Psychology, 92*(Part 1), 129–170.
Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um: New information. *Journal of Psycholinguistic Research, 32*, 25–36.
Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*, 914–930.
Bakker, K. (1999). Technical solutions for quantitative and qualitative assessments of speech fluency. *Seminars in Speech and Language, 20*, 185–195.
Bauman, J., Hall, N., Wagovich, S., Weber-Fox, C., & Bernstein Ratner, N. (2012). Past tense marking in the spontaneous speech of preschool children who do and do not stutter. *Journal of Fluency Disorders, 37*, 314–324.
Bernstein Ratner, N., & MacWhinney, B. (2016). Your laptop to the rescue: Using the CHILDES Archive and CLAN programs to improve language sample analysis. *Seminars in Speech and Language, 37*(2), 74–84.
Bernstein Ratner, N., & MacWhinney, B. (2018). Use of big data child language assessment. In B. Lust (Ed.). *Development of linguistic linked open data resources for collaborative data-intensive research in the language sciences*Cambridge, MA: MIT Press [in press].
Bloodstein, O., & Ratner, N. Bernstein (2008). *A handbook on stuttering* (6th ed.). Delmar, NY: Cengage.
Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.). *Handbook of psycholinguistics* (pp. 945–984). .
Boersma, P., & Weenink, D. (1996). Praat, a system for doing phonetics by computer. *Tech. Rep. 132* [Amsterdam : Institute of Phonetic Sciences of the University of Amsterdam].
Boltz, M. G. (2005). Temporal dimensions of conversational interaction: The role of response latencies and pauses in social impression formation. *Journal of Language and Social Psychology, 24*, 103–138.
Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(part 2), 123–147.
Boscolo, B., Ratner, N. B., & Rescorla, L. (2002). Fluency of school-aged children with a history of specific expressive language impairment: An exploratory study. *American Journal of Speech Language Pathology, 11*, 41–49.
Brocklehurst, P. H., & Corley, M. (2011). Investigating the inner speech of people who stutter: Evidence for (and against) the Covert Repair Hypothesis. *Journal of Communication Disorders, 44*, 246–260.
Brown, R. (1973). *A first language.* Cambridge, MA: Harvard University Press.
Brundage, S. B., Bothe, A., Lengeling, A., & Evans, J. (2006). Comparing judgments of stuttering made by students, clinicians: And highly experienced judges. *Journal of Fluency Disorders, 31*, 271–283.
Byrd, C., Bedore, L., & Ramos, D. (2015). The disfluent speech of bilingual Spanish–English children: Considerations for differential diagnosis of stuttering. *Language, Speech, and Hearing Services in Schools, 46*, 30–43.
Chow, H.-M., Spray, G., Ratner, N. B., & Chang, S.-E. (2015). *Speech-language development trajectories may predict persistent developmental stuttering during childhood.* Denver: American Speech Language Hearing Association convention.
Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73–111.
Cochran, P. S., & Masterson, J. J. (1995). NOT using a computer in language assessment/intervention: In defense of the reluctant clinician. *Language, Speech, and Hearing Services in Schools, 26*(3), 213–222.
Cordes, A. K., & Ingham, R. (1994). The reliability of observational data: II. Issues in the identification and measurement of stuttering events. *Journal of Speech and Hearing Research, 37*, 279–294.
Cordes, A. K., & Ingham, R. (1996). Time-interval measurement of stuttering: Establishing and modifying judgment accuracy. *Journal of Speech and Hearing Research, 39*, 298–310.
Cordes, A. K., Ingham, R., Frank, P., & Ingham, J. (1992). Time-interval analysis of interjudge and intrajudge agreement for stuttering event judgments. *Journal of Speech and Hearing Research, 35*, 483–494.
Cordes, A. K. (2000). Individual and consensus judgments of disfluency types in the speech of persons who stutter. *Journal of Speech, Language, and Hearing Research, 43*, 951–964.
Dale, P. S., McMillan, A. J., Hayiou-Thomas, M. E., & Plomin, R. (2014). Illusory recovery: Are recovered children with early language delay at continuing elevated risk? *American Journal of Speech-Language Pathology, 23*(3), 437–447.
De Jong, N. H., & Wempe, T. O. N. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385–390.
Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*(1), 287–314.
Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 611–629.
Derwing, T., Munro, M., Thomson, R., & Rossiter, M. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*, 533–557.
de Jong, N. (2008). Second language learning of grammar: Output matters too. In M. Young-Scholten, & T. Piske (Eds.). *Input matters in SLA* (pp. 95–115). Bristol, UK: Multilingual Matters.
Ferreira, F., Lau, E. F., & Bailey, K. G. D. (2004). Disfluencies, language comprehension: And tree adjoining grammars. *Cognitive Science, 28*, 721–749.
Finneran, D., Leonard, L., & Miller, C. (2009). Speech disruptions in the sentence formulation of school-age children with specific language impairment. *Language and Communication Disorders, 44*, 271–286.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709–738.
Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic dimensions of variation in disfluency production. *Language. Cognition and Neuroscience, 29*, 1083–1096.
Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science, 31*(2), 311–341.
Fromkin, V., & Bernstein Ratner, N. (1998). Speech production. In J. Berko Gleason, & N. Bernstein Ratner (Eds.). *Psycholinguistics*(2nd ed.). New York: Cengage.
Fromkin, V. (1973). *Slips of the tongue.* San Francisco: WH Freeman.
Garrett, M. F. (1976). Syntactic processes in sentence production. *New Approaches to Language Mechanisms, 30*, 231–256.
Goldman-Eisler, F. (1958). The predictability of words in context and the length of pauses in speech. *Language and Speech, 1*, 226–231.
Goldwater, S., Jurafsky, D., & Manning, C. (2010). Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication, 52*, 181–200.
Gottwald, S., Bernstein Ratner, N., Watson, J. B., Brundage, S. B., & Zebrowski, P. (2009). Helping graduate students acquire knowledge and skills in fluency disorders. *Paper presented at the sixth international world congress on fluency disorders, rio de janiero*.
Guo, L.-Y., Tomblin, J. B., & Samelson, V. (2008). Speech disruptions in the narratives of English-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 51*, 722–738.
Hall, N. E., Yamashita, T. S., & Aram, D. M. (1993). Relationship between language and fluency in children with developmental language disorders. *Journal of Speech and Hearing Research, 36*, 568–579.
Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology, 57*, 169–177.
Hassanali, K. N., Liu, Y., Iglesias, A., Solorio, T., & Dollaghan, C. (2014). Automatic generation of the index of productive syntax for child language transcripts. *Behavior Research Methods, 46*(1), 254–262.
Heilmann, J. J. (2010). Myths and realities of language sample analysis. *Perspectives on Language Learning and Education, 17*(1), 4–8.
Hester, E., & Pellowski, M. (2014). Speech disruptions in the narratives of African American Children with reading disabilities. *Journal of Developmental & Physical Disabilities, 26*, 83–92.
Horton, W. S., Spieler, D. H., & Shriberg, E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology and Aging, 25*, 708–713.
Howell, P., Davis, S., & Williams, R. (2009). The effects of bilingualism on stuttering during late childhood. *Archives of Disease in Childhood, 94*, 42–46.
Howell, P., Tang, K., Tuomainen, O., Chan, S. K., Beltran, K., Mirawdeli, A., et al. (2016). Identification of fluency and word-finding difficulty in samples of children with diverse language backgrounds. *International Journal of Language & Communication Disorders*.
Huang, R. (2016). *An evaluation of POS taggers for the CHILDES corpus. unpublished dissertation.* City University of New York Retrieved August 7, 2016 from http://academicworks.cuny.edu/cgi/viewcontent.cgi?article=2634&context=gc_etds (Unpublished dissertation).
Hubbard, C. P. (1998). Reliability of judgments of stuttering and disfluency in young children's speech. *Journal of Communication Disorders, 31*, 245–259.
Hudson Kam, C. L., & Edwards, N. A. (2008). The use of uh and um by 3-and 4-year-old native English-speaking children: Not quite right but not completely wrong. *First Language, 28*(3), 313–327.
Indefrey, P., & Levelt, W. J. (2000). *The neural correlates of language productionThe new cognitive neurosciences* (2nd ed.). MIT press845–865.
Ingham, R. J., Cordes, A. K., & Finn, P. (1993). Time-interval measurement of stuttering: Systematic replication of Ingham, Cordes, and Gow (1993). *Journal of Speech and Hearing Research, 36*, 1168–1176.
Jaeger, J. J. (2004). *Kids' slips: What young children's slips of the tongue reveal about language development.* Psychology Press.
Justice, L. M., Breit-Smith, A., & Rogers, M. (2010). Data recycling: Using existing databases to increase research capacity in speech-language development and disorders. *Language, Speech, and Hearing Services in Schools, 41*, 39–43.
Kemper, S., Hoffman, L., Schmalzried, R., Herman, R., & Kieweg, D. (2011). Tracking talking: Dual task costs of planning and producing speech for young versus older adults. *Aging, Neuropsychology and Cognition, 18*, 257–279.
Kloth, S. A. M., Kraaimaat, F. W., Janssen, P., & Brutten, G. J. (2000). Persistence and remission of incipient stuttering among high-risk children. *Journal of Fluency Disorders, 24*(4), 253–265.
Kowal, S., O'Connell, D. C., & Sabin, E. F. (1975). Development of temporal planning and vocal hesitations in spontaneous narratives. *Journal of Psycholinguistic Research, 4*, 195–207.
Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin and Review, 18*, 135–140.
Lee, L. L., & Canter, S. M. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech & Hearing Disorders, 36*, 315–340.
Leech, K., Bernstein Ratner, N., & Weber, C. (2017). Preliminary evidence that growth in productive language differentiates stuttering persistence and recovery. *Journal of Speech Language & Hearing Research, 60*(11), 3097–3109.
Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.
Lickley, R. J., & Bard, E. C. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech, 41*, 203–226.
Louko, L. J., Edwards, M. L., & Conture, E. G. (1990). Phonological characteristics of young stutterers and their normally fluent peers: Preliminary observations. *Journal of Fluency Disorders, 15*(4), 191–210.
MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia, 48*, 3982–3992.
MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40*(1), 121–157.
MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language, 17*(2), 457–472.
MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307.
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
Maclay, H., & Osgood, C. (1959). Hesitation phenomena in spontaneous English speech. *Word, 25*, 19–44.
Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., et al. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development,* i-178.
Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F., 2nd, & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition, 116*, 177–192.
McDaniel, D., McKee, C., & Garrett, M. F. (2010). Children's sentence planning: Syntactic correlates of fluency variations. *Journal of Child Language, 37*, 59–94.
Metze, F., Riebling, E., Fosler-Lussier, E., Plummer, A., & Bates, R. (2015). The speech recognition virtual kitchen turns one. *Sixteenth annual conference of the international speech communication association*.
Miles, S., & Ratner, N. B. (2001). Parental language input to children at stuttering onset. *Journal of Speech, Language, and Hearing Research, 44*(5), 1116–1130.
Mora, J. (2006). Age effects on oral fluency development. In C. Muñoz (Ed.). *Age and the rate of foreign language learning.* Clevedon, GB: Multi-lingual matters.
Ntourou, K., Conture, E. G., & Lipsey, M. W. (2011). Language abilities of children who stutter: A meta-analytical review. *American Journal of Speech-Language Pathology, 20*(3), 163–179.
O'Connell, D. C., & Kowal, S. (2005). Uh and um revisited: Are they interjections for signaling delay? *Journal of Psycholinguistic Research, 34*(6), 555–576.
Overton, S., & Wren, Y. (2014). Outcome measurement using naturalistic language samples: A feasibility pilot study using language transcription software and speech and language therapy assistants. *Child Language Teaching and Therapy, 30*(2), 221–229.
Ozuru, Y., & Hirst, W. (2006). Surface features of utterances, credibility judgments and memory. *Memory and Cognition, 34*, 1512–1526.
Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*(1), 73–193.
Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Pre-articulatory repair processes in normal speech. *Journal of Speech and Hearing Research, 36*, 472.

Price, L. H., Hendricks, S., & Cook, C. (2010). Incorporating computer-aided language sample analysis into clinical practice. *Language, Speech, and Hearing Services in Schools, 41*(2), 206–222.
Rispoli, M., Hadley, P., & Holt, J. (2008). Stalls and revisions: A developmental perspective on sentence production. *Journal of Speech Language and Hearing Research, 51*, 953–966.
Rose, Y., Hedlund, G., Byrne, R., Wareham, T., & MacWhinney, B. (2007). Phon 1.2: a computational basis for phonological database elaboration and model testing. In P. Buttery, A. Villavicencio, & A. Korhonen (Eds.). *Proceedings of the workshop on cognitive aspects of computational language acquisition, 45th annual meeting of the association for computational linguistics* (pp. 17–24).
Ryan, B. P. (2001). A longitudinal study of articulation, language, rate, and fluency of 22 preschool children who stutter. *Journal of Fluency Disorders, 26*(2), 107–127.
Sagae, K., Davis, E., Lavie, E., MacWhinney, B., & Wintner, S. (2007). High-accuracy annotation and parsing of CHILDES transcripts. *Proceedings of the 45th meeting of the association for computational linguistics*.
Scaler Scott, K., Tetnowski, J., Flaitz, J., & Yaruss, J. S. (2014). Preliminary study of disfluency in school-aged children with autism. *International Journal of Language & Communication Disorders, 49*, 75–89.
Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics, 11*, 1–22.
Schiel, F., Heinrich, C., & Barfüßer, S. (2011). Alcohol language corpus: The first public corpus of alcoholized German speech. *Language Resources and Evaluation, 46*(3), 503–521.
Schmid, M., & Fägersten, K. (2010). Disfluency markers in L1 attrition. *Language Learning, 60*, 753–791.
Sin, K., Beltran, K., & Howell, P. (2015). Language history and its impact on stuttering and word-finding disfluency. *Proceedings of the eighth world congress on fluency disorders*. www.theifa.org.
Sisskin, V., & Bernstein Ratner, N. (2015). *My client isn't fluent, but is it stuttering?* Stuttering Foundation of America Newsletter. (in progress) Summer http://www.stutteringhelp.org/My-client-isnt-fluent-but-is-it-stuttering.
Sisskin, V., & Wasilus, S. (2014). Lost in the literature: But not the caseload: Working with atypical disfluency from theory to practice. *Seminars in Speech and Language, 35*, 144–152.
Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics, 30*, 510–532.
Smith, A., Roberts, J., Smith, S., Locke, J., & Bennett, J. (2006). Reduced speaking rate as an early predictor of reading disability. *American Journal of Speech-Language Pathology, 15*, 289–297.
Smith, A. B., Smith, S. L., Locke, J. L., & Bennett, J. (2008). A longitudinal study of speech timing in young children later found to have reading disability. *Journal of Speech Language & Hearing Research, 51*, 1300–1314.
Smith, A., Hall, N., Tan, X., & Farrell, K. (2011). Speech timing and pausing in children with Specific Language Impairment. *Clinical Linguistics & Phonetics, 25*, 145–154.
Sokolov, J. L., & Snow, C. E. (Eds.). (1994). *Handbook of research in language development using CHILDES*. Hillsdale, NJ: Lawrence Erlbaum.
Spencer, C., & Weber-Fox, C. (2014). Preschool speech articulation and nonword repetition abilities may help predict eventual recovery or persistence of stuttering. *Journal of Fluency Disorders, 41*, 32–46.
Steinberg, M., Bernstein Ratner, N., Berl, M., & Gaillard, W. (2013). Fluency patterns in narratives from children with localization-related epilepsy. *Journal of Fluency Disorders, 38*, 193–205.
Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language, 28*(2), 164–188.
Strekas, A., Bernstein Ratner, N., Berl, M., & Gaillard, W. D. (2013). Narrative abilities of children with epilepsy. *International Journal of Language and Communication Disorders, 48*, 207–219.
Tumanova, V., Conture, E. G., Lambert, E., & Walden, T. (2014). Speech disfluencies of preschool-aged children who do and do not stutter. *Journal of Communication Disorders, 49*, 25–41.
Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language, 36*(4), 743–778.
Vasic, N., & Wijnen, F. (2004). Stuttering as a monitoring deficit. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.). *Phonological encoding and monitoring in normal and pathological speech*. Hove, East Sussex: Psychology Press.
Wagovich, S. A., Hall, N. E., & Clifford, B. A. (2009). Speech disruptions in relation to language growth in children who stutter: An exploratory study. *Journal of Fluency Disorders, 34*, 242–256.
Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication, 50*, 81–94.
Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 106*(1–4), 23–79.
Yairi, E., & Ambrose, N. (1999). Early childhood stuttering I: Persistency and recovery rates. *Journal of Speech, Language and Hearing Research, 42*, 1097.
Yairi, E., & Ambrose, N. (2004). *Early childhood stuttering*. PRO-ED Inc.
Yairi, E., Ambrose, N. G., Paden, E. P., & Throneburg, R. N. (1996). Predictive factors of persistence and recovery: Pathways of childhood stuttering. *Journal of Communication Disorders, 29*(1), 51–77.
Yoshimura, Y., & MacWhinney, B. (2007). The effect of oral repetition in L2 speech fluency: System for an experimental tool and a language tutor. *SLATE Conference,* 25–28.

Nan Bernstein Ratner is Professor, Hearing and Speech Sciences, at the University of Maryland, College Park. She has frequently published in the area of fluency and fluency disorders. She has been recognized for her research by the International Fluency Association (of which she is President-Elect) and the American Speech-Language-Hearing Association (of which she is a Fellow and Honors Recipient). Together with Brian MacWhinney, she co-directs the new FluencyBank project, funded by NIH and NSF.

Brian MacWhinney is Professor of Psychology, Carnegie-Mellon University. He has published extensively in child language acquisition, computational analysis of language and psycholinguistics. He was the first recipient of the International Association for Child Language Roger Brown award for distinguished contributions to language acquisition research. He is founder of the TalkBank project, and its director. With Nan Bernstein Ratner, he co-directs the new FluencyBank project, funded by NIH and NSF.