# Overview

The National Science Foundation (NSF) and the National Institute on Deafness and other Communication Disorders (NIDCD) have recently provided funding to establish FluencyBank (<http://fluency.talkbank.org> ) as a new component of the larger TalkBank system (<http://talkbank.org> ). The purpose of this article is to explain how FluencyBank will work to extend our understanding of the development of fluency and disfluency in both children and adults. We will begin by explaining how the concept of FluencyBank has emerged from the overall TalkBank project for creating databases for the study of spoken language. Second, we will present and discuss the specific 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.

# 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 configuration and its compliance with digital archiving standards, see Bernstein Ratner & MacWhinney (in press). The TalkBank initiative began in 2000 as 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.

## TalkBank Principles

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 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, are freely available and downloadable from the website at http://talkbank.org.

To facilitate use of CLAN analysis programs by researchers or clinicians who have employed other methods of transcription, CLAN includes a series of 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.

## TalkBank Components

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. 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 http://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 data in the various components of TalkBank are freely open for browsing, downloading and analysis. However, users of data in the clinical banks such as AphasiaBank and FluencyBank must request a password, and access to the data in HomeBank requires further attention to methods for safe-guarding the use of untranscribed audio.

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. 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. 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 pages 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.

TalkBank is an international and cross-linguistic project. The free CLAN program provides automatic morphological/syntactic tagging for Cantonese, Chinese, Dutch, English, French, German, Hebrew, Japanese, Italian, and Spanish. Because these 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.

CLAN includes a wide variety of user-tailored search and analysis routines that have been extremely fruitful in evaluating theoretical claims and models. Such 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 marking of the finite verb (Wexler XXX, Pine XXX). 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 HSLLD corpus). In these debates, and many others, the availability of a shared open database has been crucial in the development of analysis and theory.

## Clinical Extensions of TalkBank

TalkBank entered the clinical arena with the creation of the AphasiaBank initiative, funded in 2007, and directed by Audrey Holland and Brian MacWhinney (Forbes, Fromm & MacWhinney, 2014). AphasiaBank 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 normative 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 analysis of clinical data was then extended through the construction of KIDEVAL for child language data. The goal of KidEval is to facilitate faster, more accurate and more informative language sample analysis, by both researchers and practicing clinicians.

Language sample analysis (LSA) for clinical purposes can be quite time-consuming (Overton & Wren 2014). After spending hours of work to create a basic, clinicians and researchers must then devote further time to compute measures such as Developmental Sentence Score (DSS; 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; Evans &Miller, 1999; Miller, 2001; Hassanali, 2014), it is not very frequently used in practice.

Fortunately, the use of free utilities such as CLAN that 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. By extension, this process can benefit clinical assessment, therapy planning, and measurement of therapeutic progress in clinical work in fluency disorders. Media-linked transcripts preserve data in a single integrated, annotatable format that can easily facilitate post hoc hypothesis testing and data exploration.

# Why we need FluencyBank

Because spoken language production is less amenable to controlled study than is comprehension, fluency is under-represented in psycholinguistic research (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, Kam, & Tanenhaus, 2007; Corley & Stewart, 2008; Ferreira, Lau, & Bailey, 2004; MacGregor, Corley, & Donaldson, 2010; Watanabe, Hirose, Den, & Minematsu, 2008). However, 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, et al., 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 (N. de Jong, 2008; Derwing, Munro, Thomson, & Rossiter, 2009; Segalowitz, 2010; Yoshimura & MacWhinney, 2007). Finally, the study of disfluency is important for the development of algorithms for automatic speech recognition (Goldwater, Jurafsky, & Manning, 2010).

The classic psycholinguistic models of speech production (Goldman-Eisler, 1958; Maclay & Osgood, 1959; Fromkin, 1973; Garrett, 1976; Dell & Reich, 1981; Dell & Seaghda, 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 of the tongue from Jaeger (2004) and Stemberger (1989). However, these slips of the tongue 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, et al., 2010; McDaniel, McKee, & Garrett, 2010; Rispoli, Hadley, & Holt, 2008; 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.

## 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, et al., 2011; McDaniel, Mckee, & Garrett, 2010) in all speakers, 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.

Stuttering is one of the most prominent and puzzling disorders of language production. This disorder has been attributed variously to problems with sound features, grammatical structures, lexical access, or some combination of these factors (Blodstein & Ratner XXX). Understanding the nature of these barriers can help in tailoring educational or clinical programming to respond to individual profiles. By understanding which aspects of processing are blocking efficient production, we can specify therapies to improve these processes. To do this, we should first determine which fluency features are shared across groups and which are uniquely associated with specific language learning/production conditions and diagnoses.

Work with typical adults shows 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).

For people who stutter, current models cannot predict why normal disfluencies and stutters 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.

To close these various gaps in our understanding of the development of fluency and disfluency, we need detailed longitudinal data collected across many children and adults using a consistent set of tasks, consistently transcribed 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 addition, FluencyBank seeks to bring together multiple communities interested in the development of fluency, including psychoinguists, 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 et al., 2015, re-evaluating widely-publicized claims made by Howell et al., 2009; Byrd, Bedore & Ramos, 2015).

## The coding dilemma and potential solutions

To build FluencyBank, we need to develop highly systematic standards and practices for fluency coding. Unfortunately, fluency coding has been subject to significant reliability problems (Brundage, et al., 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 et al. (2009) found that, in an analysis of the disfluencies in a two-minute, 120-word speech sample by XXX coders, the count of dislfuencies ranged between a low of 12 and a high of 40.

Coders agree better when samples are cut into short (e.g., 5 sec) random intervals (Cordes & Ingham, 1996; Cordes, et al., 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.
2. FluencyBank relies on a consistent system for fluency coding, based on CHAT.
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 create a core set of “gold standard” transcriptions of disfluency patterns.
5. Using these gold standard transcriptions, we can train automatic speech recognition (ASR) systems such as SpeechKitchen (Metze, 2012 XXX) that can 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 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 the underlying acoustic facts are quantified in Praat, the basic phonological facts as characterized in Phon, and the basic lexical and morphosyntactic facts as characterized by CLAN, we can achieve much great levels of coding consistency. On this basis, we can 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.

## Distinguishing stuttering from language encoding difficulty

There is growing evidence that other 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, Bernstein Ratner, & Rescorla, 2002; Guo, Tomblin, & Samelson, 2008; Hall, Yamashita, & Aram, 1993; Smith et al., 2011; Vasic & Wijnen, 2004), reading impairment (RI) (Smith et al., 2006; Smith et al., 2008; Hester & Pellowski, 2014) and Autism Spectrum Disorder (ASD) (Lake, Humphreys, & Cardy, 2011; Scaler Scott et al., 2014; Sisskin & Wasilus, 2014). Notably, Sisskin & Wasilus note that atypical disfluency has been “lost in the literature, but not on the caseload.” A survey of over 200 SLPs (Sisskin & Bernstein Ratner, 2016) 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. The rich analytic capacities of TalkBank and the CLAN programs can aid us in pulling apart the characteristic verbal patterns of each of these disfluency profiles.

## 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 et al., 2014). Yet, efforts to identify *which* children will recover have relied primarily on prediction from familial history, and few other prognostic cues have emerged. 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.

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 children who experience only passing difficulty with speech fluency, rather than the profiles that have the potential to inform the initial phases of a life-long communication disorder. 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 are now able to test some features of early expressive language that may predict persistence or recovery. Our preliminary work with CWS in two large cohorts strongly supports original findings from the Illinois Stuttering Project (ISP; Yairi, et al, 1995). 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, Brown, Bernstein Ratner & Weber, in press; 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.

Addressing other questions will require us to examine transcripts linked to high quality media. Is there a 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 did not directly address this question.

How does bilingualism impact 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. 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, Bedore & Ramos, 2015). Of course, stuttering is a disorder with serious lifetime handicaps, whereas second language acquisition is a normal process. To misclassify a child can be educationally and socially handicapping.

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 (N. de Jong, 2008; Derwing, Munro, Thomson, & Rossiter, 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.

# The components of FluencyBank

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

## Database

FluencyBank is the first multi-investigator effort to provide data sharing for describing how children progress in fluent language production from their first words to mature adult-like utterances, and the tracking of profiles of atypical fluency development, such as stuttering. Justice et al. (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 will inform some of our questions.

Critically, FluencyBank will also provide a home and data-sharing access for speech samples collected as part of projects involving 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. 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. This is 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.

## Contributing

FluencyBank facilitates data contributions in two ways. First, using automatic conversion programs, we can transform non-CHAT data into CHAT format for consistent analysis. Second, we facilitate contributions by adapting data access to correspond to IRB requirements. Unless existing video data were gathered with explicit consent for open public use, we will extract the audio from the video and place the data behind password protection. To de-identify data, we will 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 <http://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.

## Teaching with FluencyBank

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 XXX, Johnon XXX) 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.

FluencyBank has also 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 IRB-approved set of interviews with adults who stutter specifically designed to provide multi-media samples (with accompanying diagnostic instrument examples) for open access without password protection. Two projects, called Voices of People who Stutter/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. Over a dozen samples from adults who stutter have been posted since January 2017, and used in numerous fluency disorders classes world-wide; materials from the adults who clutter are anticipated to be available sometime during Autumn 2017. Both projects are open-ended and will continue to recruit and add to the teaching materials site.

## Tools

FluencyBank is working to provide new, more accurate methods for fluency analysis based on media linkage. These include both transcript analysis tools, as well as transcription assistance tools.

**Transcript analysis**: The FluCalc program was released as a part of CLAN in June 2017. This program provides 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 & Ambrose (XXX ). Currently, these values are based on words rather than syllables, 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 segment words into syllables.

**Transcription utilities**: 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.

**Cross-domain analysis of fluency breakdown**. TalkBank currently supports utilities for fully automated morphological, syntactic and phonological 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% (source), 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.

**Corpus development, protocol recommendations and utility testing**. 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; N. H. 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.

## Training and Support

FluencyBank and TalkBank grants 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, and trouble-shooting of problems with data or program use. We currently provide four free manuals: the full CHAT transcription manual, the full CLAN manual, with its large array of data analysis options, the full MOR manual for programs that do morphosyntactic analysis, and the SLP guide to CLAN, which is an abridged user manual for “quick start use” by practicing clinicians 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 individually directed at illustrating specific aspects of transcription and analysis.

We will be introducing new users to FluencyBank measures and methods via presentations and tutorials at the Oxford Dysfluency Conference and the ASHA annual convention, as well as additional meetings in the discipline, such as the International Fluency Association joint congress. In addition, any interested user can request an Internet-facilitated tutorial.

# 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, Phon 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. 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 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 that from 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. We invite readers to explore and use this new resource.

# References:

Altmann, G. T. (2001). The language machine: Psycholinguistics in review. *British Journal of Psychology*, *92* Part 1, 129- 170.

Ambrose, N., & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language and Hearing Research, 42*, 895-909.

Anderson, J. D., Wagovich, S. A., & Hall, N. E. (2006). Nonword repetition skills in young children who do and do not stutter*. Journal of Fluency Disorders, 31*, 177-199.

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*, 314-24.

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, *161*25-35.

Bernstein Ratner, N. (1992). Measurable outcomes of instructions to modify normal parent-child verbal interactions. *Journal of Speech and Hearing Research, 35*, 14-29.

Bernstein Ratner, N. (1997). Stuttering: a psycholinguistic perspective. In R. F. Curlee & G. Siegel (Eds.), *Nature and treatment of stuttering: new directions (2nd ed.)* (pp. 99-127). Boston, MA: Allyn & Bacon.

Bernstein Ratner, N. (2000). Performance or capacity, the model still requires definitions and boundaries it doesn't have. *Journal of Fluency Disorders, 25*, 337-346.

Bernstein Ratner, N., & Guitar, B. (2006). Treatment of very early stuttering and parent-administered therapy: the state of the art. In N. Bernstein Ratner & R. Tetnowski (Eds.), *Current issues in stuttering research and practice* (pp. 99-124). Mahwah, NJ: Lawrence Erlbaum Associates.

Bernstein Ratner, N. & Sih, C.C. (1987).The effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders, 52*, 278-287.

Bloodstein, O., & Bernstein Ratner, N. (2008). *A Handbook on Stuttering (6th ed.)*. Delmar, NY: Cengage.

Blumgart, E., Tran, Y., & Craig, A. (2010a). An investigation into the personal financial costs associated with stuttering. *Journal of Fluency Disorders, 35*, 203-215

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., Bernstein Ratner, N., & 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.

Brownell, R. (2010). *Expressive and Receptive One-Word Vocabulary Tests– 4th Edition*. San Antonio, TX: Pearson Assessments.

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.

Chang, S-E., Zhu, D., Choo, A., & Angstadt, M. (2015) White matter neuroanatomical differences in young children who stutter. *Brain*. doi:10.1093/brain/awu400.

Chow, H-M., Spray, G., Bernstein Ratner, N. & Chang, S-E. (2015) Speech-language development trajectories may predict persistent developmental stuttering during childhood. American Speech Language Hearing Association convention, Denver.

Clark, H. H., & Fox Tree, J. E. (2002) Using uh and um in spontaneous speaking. *Cognition, 84*, 73-111

Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 696-­‐702.

Cooper, E., & Cooper, C. (2003). *Personalized fluency control therapy, 3rd edition*. Austin, TX: Pro-Ed.

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.

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.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It’s the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, *105*, 658‐668.

Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass, 2*, 589-602.

Curlee, R. F. (1981). Observer agreement on disfluency and stuttering. *Journal of Speech and Hearing Research, 24*, 595-­‐600.

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. (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.

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.

De Nil, L., Brutten, G., Hancock, K., Craig, A., McCready, C., McCaul, A., & ... Gilmore, G. (1998). *Communication Attitude Test--Revised*. *Journal of Speech, Language, and Hearing Research*, *41,* 1242-1252.

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.

Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental Trends in Lexical Diversity. *Applied Linguistics*, *25*(2), 220-242.

Ehlen, P., Schober, M. & Conrad, F. (2007). Modeling speech disfluency to predict conceptual misalignment in speech survey interfaces. *Discourse Processes, 44*, 245-265.

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.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. New York: McGraw-Hill.

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.

Fromkin, V., & Bernstein Ratner, N. (1998). Speech production. In J. Berko Gleason & N. Bernstein Ratner (Eds.), *Psycholinguistics, 2nd Edition*. New York: Cengage.

Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with Specific Language Impairment and their typically developing peers on nonverbal cognitive tests: a meta-analysis. *Journal of Speech, Language & Hearing Research*, *57*, 1363-1382.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's test of Nonword Repetition: A test of phonological working memory, *Memory, 2*, 103-127.

Gayraud, F., Lee, H.-R., & Barkat-Defradas, M. (2011). Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. *Clinical Linguistics and Phonetics, 25*, 198-209.

Goldman, R., & Fristoe, M. (2000). *The Goldman‐Fristoe Test of Articulation -2*. San Antonio: Pearson Assessments.

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. (2011). Stuttering prevention and early intervention. In Guitar, B. & McCauley, R. (eds.) *Treatment of stuttering: established and emerging interventions*. Baltimore: Lippincott, Williams and Wilkins.

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.

Guitar, B. (2014). *Stuttering: an integrated approach to its nature and treatment*. Baltimore, MD: Lippincott, Williams & Wilkins.

Hakim, H. B., & Bernstein Ratner, N. (2004). Nonword repetition abilities of children who stutter: an exploratory study. *Journal of Fluency Disorders, 29*, 179-199.

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., & Kolk, H. J. (2001). Error monitoring in speech production: A computational test of the Perceptual Loop Theory*. Cognitive Psychology, 42*, 113-148.

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology, 57*, 169-177.

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.

Hof, Y., Wijnen, F., & Dejonckere, P. (2009). Language planning disturbances in children who clutter or have learning disabilities. *International Journal of Speech-Language Pathology, 11*, 496-508.

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.

Hubbard, C. P. (1998). Reliability of judgments of stuttering and disfluency in young children's speech. *Journal of Communication Disorders, 31*, 245-259.

Ingham, R. & Cordes, A. K. (1997). Identifying the authoritative judgments of stuttering: comparisons of self-judgments and observer judgments. *Journal of Speech, Language, and Hearing Research*, 40, 581-594.

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.

Jones, M., Onslow, M., Packman, A., O'Brian, S., Hearne, A., Williams, S., . . Schwarz, I. (2008). Extended follow-up of a randomized controlled trial of the Lidcombe Program of Early Stuttering Intervention. *International Journal of Language and Communication Disorders, 43*, 649-661.

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.

Kolk, H., & Heeschen, C. (1992). Agrammatism, paragrammatism, and the management of language. *Language and Cognitive Processes, 7*, 89-­‐129.

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.

Larsen-Freeman, D. (2009). Adjusting expectations: the study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*, 579-589.

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. (in review). Linguistic predictors of persistence and recovery in children who stutter*. Journal of Fluency Disorders*.

Leech, K., Salo, V., Rowe, M., & Cabrera, N. (2013). Father input and child vocabulary development: The importance of 'Wh', questions and clarification requests. *Seminars in Speech and Language*, *34*(4), 249-259.

Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-38.

Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (2015). A developmental vocabulary assessment for parents (DVAP): validating parental report of vocabulary size in 2-to 7-year-old children. Journal of Cognition and Development, (ahead-of-print), 1-13.

Lickley, R. J., & Bard, E. C. (1998). When can listeners detect disfluency in spontaneous speech? *Language and Speech, 41*, 203‐226.

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.

Maclay, H., & Osgood, C. (1959). Hesitation phenomena in spontaneous English speech. *Word, 25*, 19-44.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk (3rd ed.)*. Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. & Bernstein Ratner, N. (2015 and to appear). Using CHILDES archive and TalkBank tools to inform clinical decision making in developmental language disorders. LSA *Linguistic Linked Open Data (LLOD) Resources* *for Collaborative Data-Intensive Research in the Language Sciences Workshop,* Chicago, July*.* To appear in *Language Resources and Evaluation Journal* Special Issue: “Development of Linguistic Linked Open Data Resources for Collaborative Data-Intensive Research in the Language Sciences.”

MacWhinney, B., & Osser, H. (1977). Verbal planning functions in children’s speech. Child Development, 48, 978-985.

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.

Miles, S., & Bernstein Ratner, N. (2001). Parental language input to children at stuttering onset*. Journal of Speech, Language and Hearing Research, 44*, 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.

Nation, P. (1989). Improving speaking fluency. *System, 17,* 377-384.

Nip, I. B., & Green, J. R. (2013). Increases in cognitive and linguistic processing primarily account for increases in speaking rate with age. *Child Development*, *84*, 1324-1337.

Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics, 30*, 555-578.

Nye, C., Vanryckeghem, M., Schwartz, J. B., Herder, C., Turner III, H. M., Howard, C., & ... Bosshardt, H. (2013). Behavioral stuttering interventions for children and adolescents: A systematic review and meta-analysis. *Journal of Speech, Language & Hearing Research*, *56*), 921-932.

Ozuru, Y., & Hirst, W. (2006). Surface features of utterances, credibility judgments, and memory. *Memory and Cognition, 34*, 1512-1526.

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually-developing two-year-olds. *Journal of Experimental Child Psychology*, *108*(1), 113–125.

Pellegrino, F., Farinas, J., & Rouas, J.‐L. (2004). Automatic estimation of speaking rate in multilingual spontaneous speech. Paper presented at the Proceedings of Speech Prosody 2004, Nara, Japan.

Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics, 35*, 184-207.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Pre-articulatory repair processes in normal speech. *Journal of Speech and Hearing Research, 36*, 472.

Riley, J., Riley, G., & Maguire, G. (2004). Subjective Screening of Stuttering Severity, Locus of Control and Avoidance: research edition. *Journal of Fluency Disorders, 29*, 51‐62.

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). Stroudsburg: ACL.

Rowe, M. L. (2012). A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. *Child Development*, *83*(5), 1762-1774.

Sadri Mirdamadi, F., & De Jong, N. H. (2015). The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research*, *31*, 105-116.

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. Prague: ACL.

Sagae, K., Lavie, A., & MacWhinney, B. (2005). Automatic measurement of syntactic development in child language*. Proceedings of the 43rd Meeting of the Association for Computational Linguistics* (pp. 197-204). Ann Arbor: ACL.

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*, article and corpus at <https://www.phonetik.uni-muenchen.de/Bas/BasALCeng.html>

Schmid, M., & Fägersten, K. (2010). Disfluency markers in L1 attrition. *Language Learning*, *60*, 753-791.

Segalowitz, N. (2010). *Cognitive bases of second language fluency.* New York: Routledge.

Semel, E., & Wiig, E. (1994). *The Clinical Evaluation of Language Function. Third Edition*. San Antonio: The Psychological Corporation.

Sevilla, Y., Maldonado, M., & Shalóm, D. E. (2014). Pupillary dynamics reveal computational cost in sentence planning. *Quarterly Journal of Experimental Psychology, 67*, 1041-1052.

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](http://www.theifa.org).

Sisskin, V. & Bernstein Ratner, N. (in progress). My client isn’t fluent, but is it stuttering? Stuttering Foundation of America Newsletter, Summer 2015. <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., Hall, N., Tan, X., & Farrell, K. (2011). Speech timing and pausing in children with Specific Language Impairment. *Clinical Linguistics & Phonetics*, *25*, 145-154.

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.

Solomon, R., Van Egeren, L. A., Mahoney, G., Huber, M. Q., & Zimmerman, P. (2014). PLAY Project Home Consultation intervention program for young children with autism spectrum disorders: A randomized controlled trial. *Journal of Developmental and Behavioral Pediatrics*, *35*, 475-485.

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.

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.

Tsai, Peitzu, Brundage, Shelley, Lim, Valerie and Bernstein Ratner, Nan (2010). Linguistic analysis of bilingual stuttering: concepts and methods. In P. Howell & J. Van Borsel (eds.). Fluency disorders and language diversity. Taylor & Francis.

Tumanova, V., Conture, E., Lambert, E., & Walden, T. (2014). Speech disfluencies of preschool-age children who do and do not stutter. *Journal of Communication Disorders*, 49, 25-41.

Vanryckeghem, M., & Brutten, GJ. (2007) Communication Attitude Test for Preschool and Kindergarten Children Who Stutter (KiddyCAT) San Diego, CA: Plural Publishing.

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., & Bernstein Ratner, N. (2007). Frequency of verb use in young children who stutter. *Journal of Fluency Disorders, 32*, 79-94.

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.

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.

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.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-­‐line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24*, 1-27.