Instructions to Contributors

Dear Contributor:

Enclosed in this document please find the page proofs, copyright transfer agreement (CTA), and offprint order form for your article in the *Seminars in Speech and Language*, Volume 37, Number 2, 2016. Please print this document and complete and return the CTA and offprint order form, along with corrected proofs, within 72 hours.

1) Please read proofs carefully for **typographical** and **factual** errors only; mark corrections in the margins of the proofs in blue or black pen; please be sure to write as clearly as possible so no errors are introduced into your article. **Answer (on the proofs) all author queries marked in the margins of the proofs.** Check references for accuracy. **Please check on the bottom of the 1st page of your article that your titles and affiliations are correct.** Avoid elective changes, because these are costly and time consuming and will be made at the publisher's discretion.

2) Please pay particular attention to the proper placement of figures, tables, and legends. Please provide copies of any formal letters of permission that you have obtained.

3) **Please return the corrected proofs, signed copyright transfer agreement, and your offprint order form.**

4) As a contributor to this journal you will receive a complimentary PDF file of the article after publication.

- If you wish to order offprints, **please circle the quantity required** (left column) **and the number of pages in your article.** If you wish to order copies of the journal please enter the number of copies on the indicated line.

- If you do not want to order offprints or journals simply put a slash through the form, **but please return the form.**

Please return all materials within 72 hours. E-mail is the easiest way to ensure your corrections are received in a timely manner. You may also return materials via fax or overnight mail to:

Joycelyn Reid, Senior Production Editor
Thieme Medical Publishers
333 Seventh Avenue, 18th Floor
New York, NY 10001 USA
Phone: 212-584-4668
Fax: 212-947-1112
Email: joycelyn.reid@thieme.com

Please do not return your materials to the editor or the typesetter.

Please note: Due to a tight schedule, if the publisher does not receive the return of your article proofs within 7 days of the date the e-mail was sent to you, the publisher reserves the right to proceed with publication without author changes. Such proofs will be proofread by the editor and the publisher.

Thank you for your contribution to this journal.

Permission to Publish and Copyright Transfer Agreement

Manuscript Information:

Journal: Seminars in Speech and Language

Manuscript Title: Developing and Using Big Data Archives to Quantify Disfluency and Stuttering in Bilingual Children

Manuscript Number:

Authors: Shelley B. Brundage, Ph.D., CCC, BCS-F Tayler Corcoran, Catherine Wu, Charlotte Sturgil

Corresponding author's contact data: Shelley B. Brundage, Ph.D., CCC, BCS-F, Department of Speech and Hearing Science, George Washington University, 2115 G St. NW, Suite 201, Washington, DC 20052

Corresponding author's e-mail address: brundage@gwu.edu

Contact at the publishers: Joycelyn Reid

E-mail address at the publishers: joycelyn.reid@thieme.com

Dear Author,

Please

- *read this form carefully,*
- *check all Manuscript Information,*
- *sign this form with your digital signature and*
- *return to us.*

Thank you very much in advance.

Assignment of Rights

We – the Thieme Publishing Group – do not accept any manuscript for publication in a journal that has previously been published elsewhere.

Your consent to the following assignments of rights, also on behalf of the other authors (if several authors contribute to the manuscript), and the signing of this Copyright Transfer Agreement is a necessary requirement for the publication of your manuscript.

Upon acceptance of your manuscript by us you assign to us (on behalf of all authors), without geographical or language restriction and for the duration of the legal copyright term, the rights to use your article, for all print runs/updates, including the rights to:

– reproduce and distribute copies of the article in printed form (e.g., in a periodical or journal, medical textbook or other target group oriented book, paperback book, special edition for secondary markets or special customers, brochures, advertising supplements, edited volumes, etc.);

– reproduce and distribute the article in electronic media formats (e.g., magnetic tape, CD-Rom, CDI, DVD, electronic paper, hardware RAM, hard-disk, USB memory stick) and make available to the public (e.g., internet, intranet or other wired or wireless data networks), in particular by displaying on stationary or mobile visual display units, monitors, PDA, mobile phones, smart phones or other devices by download (e.g., e-pub, PDF, App) or retrieval in any other form;

– publish ourselves or to authorize the publication of excerpts in other works or articles, in audio-visual accompanying materials or interactive products or services, and including the transfer of rights of use to third parties (e.g., under the terms of licensing agreements);

– translate, transfer and process into other languages or versions (e.g., podcast, audiobook or other image and sound carriers), broadcast by means of television, cable or satellites, radio or other audio-visual media, to rent out and lend, store in an electronic archive and to use in any other type of format that may become known in the future and – where applicable – for all other rights protected by organizations assessing and/or collecting fees for copyright use. Furthermore you assign to us all statutory royalty claims under relevant law insofar we mandate an organization to administer such rights for publishers and authors; we accept the assignments.

Any adaptions, if appropriate for the exercise of the rights of use granted to us, shall be processed by us.
Please forward any inquiries that are addressed to you regarding the above-mentioned rights of use for our attention and response.

Open Access / Repositories

The rights of use are assigned to us exclusively – subject to your rights in accordance with our Open Access Guidelines. Our Open Access Guidelines state that immediately after the publication of the article by us, you and the other authors are entitled to make the published version of the article available to the public on your homepage and on the homepage of your institution for your own scientific and other non-commercial purposes. Twelve months after publication by us, you and the other authors are entitled to make the accepted manuscript version available to the public on other non-commercial websites, provided that you make full reference to the published version ("Green Open Access"). For further details please click the button "Information on Green Open Access."

OPEN ACCESS
Information on Green Open Access

For more Information on our Open Access Program please visit http://open.thieme.com.

Duties of care

Product liability laws set high standards for your duty of care as the author of a scientific manuscript. This is especially the case when you give therapeutic information and/or specify doses. Therefore please check this information carefully in the typeset page-proofs of your article. Your task will be much easier if you have the information counterchecked – depending on the sensitivity of the information within the article – by specialist colleagues. Only you, as the author, have the specialist knowledge to be able to assess the accuracy of the information. For further information on how to indicate corrections, please click the button "Correction markup symbols".

Correction markup symbols

Author's Declaration

I have taken note of the information on the duties of care under product liability law; I agree to the assignments of rights in accordance with the foregoing sections "Assignment of Rights" and "Open Access / Repositories" also on behalf of the other authors (if several authors have contributed to the article).

I declare that no third party rights will be infringed through the publication. Any material contained in the manuscript (including illustrations, tables, or other material) from third-party sources will be identified as such through citation, indicating the source. If required, I have obtained the copyright permission from the publishers concerned.

The above-mentioned assignments of rights also relate to the illustrations in your manuscript. We do not accept any illustrations for which it has not been granted all rights of use in accordance with this contract.

Should one of the foregoing regulations be or become invalid in whole or in part this shall not affect the validity of the other provisions. Any invalid provision shall be replaced by a regulation that comes as close as possible to the purpose of the invalid provision in economic terms, insofar as legally permissible.

This article is ready to print after the execution of the corrections indicated by me.

Date Digital Signature

Information on the digital signature

Order Form for Offprints and additional copies of the Seminars in Speech and Language
(Effective January 2013)

Please circle the cost of the quantity/page count you require (orders must be in increments of 100)

	Pages in Article / Cost				
Quantity	**1 to 4**	**5 to 8**	**9 to 12**	**13 to 16**	**17 to 20**
100	$298	$497	$746	$968	$1,158
200	$397	$646	$970	$1,258	$1,495
300	$496	$798	$1,198	$1,568	$1,869
400	$549	$886	$1,330	$1,735	$2,075
500	$598	$966	$1,450	$1,886	$2,262
1000	$1,076	$1,739	$2,610	$3,385	$3,995

Volume/Issue #: ______________________ Page Range (of your article): ______________

Article Title: ____

MC/Visa/AmEx No: ______________________ Exp. Date: ______________

Signature: ____

Name: ____

Address: ____

City/State/Zip/Country: ____

Corresponding author will receive a complimentary PDF of the article after publication.

Number of **additional** copies of the journal, at the discounted rate of $25.00 each: ______________

Notes

1. The above costs are valid only for orders received before publication of the issue. **Reprints ordered after printing will be substantially more expensive.**

2. **A shipping charge will be added to the above costs.**

3. Reprints are printed on the same coated paper as the journal and saddle-stitched.

4. For larger quantities or late orders, please contact reprints department: Phone: +1(212) 584-4662
Fax: +1(212) 947-1112
E-mail: reprints@thieme.com

Developing and Using Big Data Archives to Quantify Disfluency and Stuttering in Bilingual Children

Shelley B. Brundage, Ph.D., CCC, BCS-F,[1] Tayler Corcoran,[1] Catherine Wu,[1] and Charlotte Sturgil[1,2]

ABSTRACT

Worldwide, bilingualism is the rule rather than the exception, and yet we have surprisingly little research data on the fluency development of bilingual children, and even less information on their potential risk for stuttering. Many variables influence a bilingual child's language, speech, and fluency development (e.g., amount of exposure to each language); controlling these variables in research studies necessitates large numbers of bilingual participants. The frequency and types of typical disfluencies in the speech of young children are also varied. In addition, stuttering is also variable in its presentation, and when we assess bilingual children for the presence of stuttering we are adding yet another layer of complexity. This article reviews research on typical disfluencies in monolingual and bilingual speakers, and how this information might be useful clinically. We provide examples from our laboratory to illustrate how CLAN[Q3] can be used over time to track the behaviors of research participants. We also present data on the identification of stuttering in bilingual children. We discuss challenges to studying bilingual speakers and how big data initiatives such as TalkBank address these challenges to increase our understanding of bilingual fluency development.

Q3

KEYWORDS: Stuttering, bilingual, children, typical disfluency, big data

[1]Department of Speech and Hearing Science, George Washington University, Washington, DC; [2]Ingenuity Prep Public Charter School, Washington, DC.

Address for correspondence: Shelley B. Brundage, Ph.D., CCC, BCS-F, Department of Speech and Hearing Science, George Washington University, 2115 G St. NW, Suite 201, Washington, DC 20052 (e-mail: brundage@gwu.edu).

; Guest Editor, Brian MacWhinney, MD

Semin Speech Lang 2016;37:117–127.
DOI: http://dx.doi.org/10.1055/s-0036-1580739.
ISSN 0734-0478.

Learning Outcomes: As a result of this activity, the reader will be able to (1) describe speech behaviors that are considered typical disfluencies; (2) describe the functions that typical disfluencies are thought to serve in speech production; (3) list challenges to studying fluency development and stuttering in bilingual children; (4) list reasons why big data initiatives are useful for answering questions about speech, language, and fluency development in bilingual children.

Bilingualism can be defined as "the regular use of two or more languages (or dialects)[Q4]."[1] It is estimated that from one-third to over one-half of the world's population is bilingual.[2,3] Given this large proportion of bilingual speakers, it is surprising that there is not more information in the research literature on the frequency of typical disfluencies in bilinguals. In addition, we do not as yet know what constitutes "normal" amounts of typical and/or stuttering-like disfluencies in bilingual children, making assessment of fluency disorders challenging in this population. Compounding the task for clinicians, the bilingual children on their caseloads often speak a myriad of different languages; that is, not everyone is bilingual in the same two languages, making it challenging to complete language sample analyses for many of the children a clinician is likely to encounter in our increasingly bilingual world.

So how do clinicians address these challenges? Some will decide—unfortunately—to assess the child in only one language (often the clinician's own language). This decision can lead to misrepresentations of the child's linguistic abilities and fluency, particularly if the assessed language is not the child's heritage language. For example, if clinicians test only English vocabulary in a bilingual Spanish–English child, they will underestimate the total number of words the child actually knows.[4] There is a better way, and it involves using programs and utilities that have been designed to assess speech samples in a variety of languages. These free utilities, part of the TalkBank initiative, give clinicians the tools to perform cross-linguistic assessments on bilingual children. Importantly for clinical efficiency, clinicians can obtain a variety of information from one speech sample in each language, including measures of utterance length, vocabulary diversity, and frequency of morphemes unique to each language. All of this information can literally be obtained by the touch of a button using the KIDEVAL[Q5] utility in TalkBank. This is possible because TalkBank has programs that contain grammars for several languages, all of which interface with utilities that allow clinicians to analyze speech samples in those languages. TalkBank utilities are also flexible and can address common speech behaviors, such as code switching, seen in bilingual speakers. All of these TalkBank functions—grammatical analysis, vocabulary diversity, MLU[Q6]—are important for the study of disfluency in bilinguals, because disfluency rates may be a proxy for proficiency in the second language (L2[Q7]),[5] and research suggests that stuttering is influenced by linguistic ability.[6]

In this article we will (1) define typical disfluencies and summarize studies of them in monolingual English and Spanish speakers, (2) briefly review a common model of speech production and how this model provides a useful framework for explaining why certain disfluency types occur and where they occur in the speech production process, (3) summarize the nature and frequency of disfluencies in bilingual speakers and how these disfluencies might serve as a proxy for language proficiency, (4) discuss the challenges to assessing disfluency and stuttering in bilingual children in the face of a lack of normative data for this population, (5) illustrate how big data initiatives such as the TalkBank project can help address some of these challenges, (6) present data on stuttering identification in bilingual children, and (7) provide clinical suggestions for assessment of stuttering in bilingual children.

TYPICAL DISFLUENCIES IN MONOLINGUAL ENGLISH AND MONOLINGUAL SPANISH SPEAKERS

Typical disfluencies can be broadly defined as "phenomena that interrupt the flow of speech and do not add propositional content to an utterance."[7(p.709)] Importantly, typical disfluencies can be differentiated from stuttering-like disfluencies, which include part-word repetitions, sound prolongations, and blocks.[8] Many different professionals study speech disfluencies, including speech-language pathologists (SLPs), linguists, and psychologists, and because of this there are many different definitions of what constitutes typical disfluencies. Word and phrase repetitions, revisions, filled pauses, and hesitations all occur in definitions of typical disfluencies. In English, typical disfluencies comprise 5 to 10% of natural conversation,[9] occurring at a rate of about 6 per 100 words.[10] In short, typical disfluencies are common in English spontaneous speech, and this finding has been replicated in other languages, including Chinese, Croatian, Dutch, German, Norwegian, Japanese, Russian, Spanish, Swedish, and Tok Pisin, suggesting the universality of these types of disfluencies in language production.[11–19]

Several corpora-based studies have informed our knowledge of disfluencies in spontaneous speech dialogues in Spanish. In general, these studies report similar types of disfluencies as those reported for monolingual English speakers, including retracings, filled pauses, and hesitations/silent pauses.[20,21] Findings of these studies include an overall disfluency rate of 4 to 6 disfluencies per 100 words in Spanish, which is similar to reports of disfluency rates reported in adult-directed speech in English.[9,17] The common Spanish filled pauses *ah*, *eh*, and *mm* were found to occur at a rate of 1.67 per 100 words. In another corpus set of mostly formal language monologues, speakers produced an average of 5 disfluencies per 100 words (not including silent pauses); around 1.5 per 100 words were fillers.[22] These rates, and the overall rates in particular, are all similar to the rates found in other studies using English spontaneous speech corpora.[10,23]

Disfluencies in the speech of monolingual Spanish-speaking children have also been the studied. Watson and colleagues evaluated the type and frequency of typical disfluencies in the spontaneous speech of 2- and 3-year-old monolingual Spanish children ($n = 32$).[24,25] The children exhibited disfluencies similar to those seen in English-speaking children: revisions, interjections, and word repetitions. The 3-year-old children were significantly more disfluent than the 2-year-old group. Disfluency rates ranged from 0 to 25%, with most of the rates falling in the 3 to 6% range. These authors note that there was significant variability in disfluency rates and the types of disfluencies exhibited by individual children, an issue that we will return to in the assessment challenges section later. Taken together, these studies suggest that the types of disfluencies seen in Spanish are similar to those seen in English and that monolingual Spanish-speaking preschoolers, on average, do not exhibit significantly higher rates of disfluency than their monolingual English peers, although there is significant variability in individual children.

A behavior that occurs so frequently must have a purpose. What function do typical disfluencies serve? Research suggests that these types of disfluencies (1) allow speakers to buy time for speech production planning and/or lexical retrieval, (2) to hold the floor and not cede their conversational turn, (3) to focus the listener's attention on new/novel information, and (4) give listeners time to process the speaker's utterance.[18,26–34] Interestingly, speakers may use *um* and *uh* differentially to communicate how much time their utterance will take to complete, with *um* signaling longer utterances to follow.[29] Any behaviors that are this frequent in conversation must be accounted for in speech production models.

LEVELT'S MODEL OF SPEECH PRODUCTION

Levelt presents a model of speech production in monolingual speakers that involves at least three parts: a *conceptualizer*, a *formulator*, and an *articulator*.[33] The conceptualizer encompasses the speaker's communicative intents and contains "preverbal" aspects of the message. The formulator takes these preverbal messages and develops a "speech plan" that conforms to both the lexical and the phonologic rules of the language spoken.[35] Importantly for studies of bilinguals, Levelt considers the lexicon to be made up of two parts: the lemma and the lexeme. Lemmas include semantic and syntactic properties of words whereas lexemes contain phonological properties of words, both of which might differ across languages. The formulator is responsible for selecting the correct lexical items for the message from the lexicon. This information is passed on to the articulator, where these speech plans are transformed into the spoken message. Levelt's model also includes a speech comprehension system that allows speakers to monitor their spoken messages and correct errors. Although Levelt's model was developed to explain monolingual speech production, it has been adapted with minor changes to account for bilingual speech production.[35]

Different types of disfluencies occur at different levels of Levelt's model. *Stalls* are disfluencies that indicate problems at the conceptualizer or formulator levels of Levelt's model. Stalls are "interruptions that add or change nothing in the linguistic structure being produced" including "long silent pauses, pauses filled with um or uh, and repetitions of linguistic material already produced."[36(p.954)] Thus, whole word repetitions, phrase repetitions, and filled pauses would be considered stalls. Stalls, or "glitches," are thought to result when higher-order linguistic planning is slowed in some way. This slowing could be due to linguistic planning problems, lexical selection problems, or inaccurate self-monitoring. In contrast, revisions involve changing something that has been previously spoken. Revisions are thought to result when the speaker detects a mismatch between the intended message and the actual spoken message and acts to correct it.[37] Thus, revisions by definition occur after at least part of the message has been articulated.

DISFLUENCY IN BILINGUAL SPEAKERS

Adults

Hilton developed a corpus of L2 speech using TalkBank software and coding conventions.[38] She compared native speakers of English, French, and Italian to a heterogeneous group of adult L2 speakers of those languages. She found that scores on tests of vocabulary and grammar had a strong negative correlation with mean length of hesitation, rate of hesitation, rate of retracing (i.e., repeating words or phrases), and rate of error. Intriguingly, she then used these measures to divide her L2 speakers into two groups: fluent learners, who used hesitations less often, and disfluent learners, who used them more often. When compared with fluent learners, the disfluent learners produced significantly fewer words, significantly more frequent and longer hesitations, and more retracings. Interestingly, the disfluent learners hesitated (paused) more frequently within clauses than the native speakers did, suggesting that the location of the pauses might be just as important as their overall frequency as a marker of L2 proficiency. Tavakoli found similar midclause pauses in L2 speakers, and suggested these may be due to "reformulation and online planning" rather than lexical retrieval.[39(p.71)] Hilton linked these disfluencies to difficulties with lexical access and retrieval in L2. Other studies have documented greater disfluency rates in L2 compared with the primary language in adult speakers and ascribed them to difficulties with lexical access in L2.[5,14,19,40,41] These findings support the view that "deviations from the native norm in the use of disfluency markers can initially be assumed to be rather large in the L2 but to decrease as the speaker's proficiency advances."[41(p.756)] In other words, disfluency rates might be a good proxy for L2 proficiency in bilingual language learners.[5]

Children

Do children who are learning two languages exhibit the similar types and frequency of disfluencies as monolinguals do? The literature on speech fluency in bilingual children is sparse at best, consisting primarily of case studies of small numbers of children without information on the children's exposure to each language, or comparisons to age-matched monolingual controls.[42] There are, however, a few well-controlled studies of groups of bilingual children. We will review a couple of these studies here.

Bedore and colleagues studied maze behaviors in narratives of 66 children between the ages of 4 and 7 years.[43] These 66 children were divided into three groups based on language exposure; this grouping resulted in a functionally monolingual English-speaking group, a functionally monolingual Spanish-speaking group, and a bilingual English–Spanish group. There were no significant differences in either the percentage of utterances with mazes or the distribution of maze types (repetitions and filled pauses, among others) between groups. However, both the monolingual Spanish and bilingual groups speaking in Spanish exhibited significantly more "grammatical revisions" than bilingual and monolingual children speaking in English. These results indicate that language differences may influence disfluency type and frequency, and that linguistic structure may dictate disfluency patterns to some extent. This study also illustrates the need for large groups of children to carefully control variables—such as amount of exposure to each language—that might reasonably influence language and fluency skills of bilingual children.

Byrd and colleagues analyzed the frequency of typical disfluencies and stuttering-like disfluencies in the narratives of 18 bilingual (Spanish–English) children who did not have stuttering diagnoses.[44] These children were divided into groups of balanced language dominance, English dominance, and Spanish dominance (6 children in each group). Disfluency rates did not differ across language dominance groups. Across all groups, typical disfluencies (defined by these authors as revisions, interjections, phrase repetitions, and unfinished words) were present at a rate of 5% in English and 8.5% in Spanish.

CHALLENGES TO STUDYING BILINGUALS AND HOW BIG DATA ARCHIVES CAN HELP

Big data archives will be necessary if we wish to answer clinically relevant questions about bilingual speakers. As the study by Bedore et al nicely illustrates, we need large sample sizes in bilingual studies, because many variables influence language and fluency in bilingual children.[43] Notice how the relatively large sample of 66 children was divided into three groups of 22 each to control for how often the children used each language and for the amount of exposure to each language, two variables known to influence language production in bilingual children.[45] Most studies of bilingual children are either case studies or relatively small numbers of children reported, making the control of influential variables impossible.[42] There is also significant variability in bilingual proficiency both between and within speakers, and larger samples are necessary to see small differences in the presence of this variation.[3,44] Relatedly, we know that socioeconomic status (SES) influences language development in both bilingual and monolingual children,[46,47] suggesting the need to control for this important variable either by studying children from only one SES level or studying a large enough group to sample across SES levels.

The nature of the tasks used in speech samples influences proficiency—as well as the frequency of typical disfluencies and stuttering—in both bilingual and monolingual speakers.[42,48–50] Similar to stuttering-like disfluencies,[51] typical disfluencies increase when the number of different words possible to complete the task increases,[50] the task involves more complex speech planning,[52,53] and when speakers are under time pressure.[53]

ILLUSTRATION OF HOW BIG DATA CAN HELP: ANALYSES FROM OUR LABORATORY

Longitudinal studies are the only types of designs that will "answer questions about relative language strength over time" in bilingual children.[54(p.38)] As part of a large, longitudinal study of bilingual language, phonology, and fluency development, our laboratory is measuring the fluency development of 151 bilingual children from 2.5 to 5 years of age. Our main aim is to obtain normative data on disfluency rates in bilingual Spanish–English preschool children, and to compare it to a group of age- and gender-matched monolingual children.

All of the bilingual children in our study live in southern Florida; participant families were selected for the study if either one or both parents were immigrants from a Spanish-speaking country. All children were typically developing at 30 months, had begun producing words in both languages, and were exposed to both Spanish and English at home, with at least 10% of input in each language by the age of 24 months. All children were born in the United States at full term with no history of hearing problems. Assessment visits occurred at 6-month intervals between the ages of 2.5 to

5 years. Parents completed extensive language histories at each visit and reported on their child's vocabulary in both languages, using the MacArthur Communicative Development Inventories (MCDI) for English and its Spanish equivalent the Inventario del Desarrollo de Habilidades Comunicativas (IDHC).[55] The children completed standardized tests in both English and Spanish during these visits and were videotaped playing with their mothers in English and in Spanish.

The parent–child interactions were transcribed and coded for disfluency and stuttering using a standard set of disfluency codes.[56,57] Code-switching behavior was also identified and coded in each transcript using CLAN[Q8] coding conventions. Coding each transcript for disfluency, stuttering, and code-switching behaviors is straightforward, and involves adding special characters to the transcript. We then used a CLAN utility called FREQ[Q9] to count the number of each code that was present in each transcript. To date we have analyzed the typical disfluencies in both languages of 20 children at 30 months. On average, these 20 children were exposed to Spanish 65% of the time and English 35% of the time at home.

Because some children spoke more than others, we calculated a disfluency rate for each sample rather than reporting total disfluencies in the sample. In English, the disfluency rate averaged 3% of words spoken and in Spanish it was 2%. A matched pairs t test revealed no significant differences between languages in disfluency rate ($t[19] = 1.59$, $p = 0.13$; power $= 0.80$) for these 20 children. English disfluency rate was significantly correlated with MCDI raw score in English at 30 months ($r = 0.50$, $p = 0.04$); Spanish disfluency rates were not significantly correlated with IDHC raw scores ($r = 0.05$, $p = 0.86$). Filled pauses were the most commonly occurring disfluency type in both languages.

Putting our findings in context with the work of Byrd and colleagues, it appears that the disfluency rates of bilingual children are similar in both languages early in development, and later diverge and increase over time.[44] We do not, however, know the trajectory of this increase or whether the trajectories are similar across languages. Recall that Byrd and colleagues studied children between the ages of 5 years, 6 months and 6 years, 7 months and found rates of typical disfluencies in the 5 to 8% range.

Future research using big data archives will fill in the gap in our knowledge of fluency development from 30 months to 5 years of age and provide us with normative data on typical disfluencies in bilingual children. Additional future research questions include analyzing our transcripts to see if certain parts of speech "attract" disfluencies in each language. This question is answerable in large part due to the CLAN utility MOR[Q10], which labels parts of speech in each utterance in a transcript (i.e., it identifies all the nouns, verbs, adjectives, adverbs, and so on, produced by the speaker). Once clinicians or researchers have completed a MOR analysis of a transcript—again, literally with the touch of a button—they can then use the FREQ command to find the frequency of each part of speech in the transcript. Combining the MOR utility with other utilities in CLAN allows clinicians to count combinations of parts of speech, such as verb phrases or noun phrases, which may be clinically useful in documenting treatment progress. To date, the MOR utility includes the grammars of 11 languages: Cantonese, Chinese, Danish, Dutch, English, French, German, Hebrew, Japanese, Italian, and Spanish.*

A FEW WORDS ABOUT STUTTERING-LIKE DISFLUENCIES IN BILINGUAL CHILDREN

Stuttering is a fluency disorder with a multifactorial origin characterized by "abnormally high frequency and/or duration of stoppages in the forward flow of speech."[58(p.10)] In contrast to TD[Q11]s, stuttering-like disfluencies are produced with tension in the vocal tract and include three primary speech behaviors: part-word repetitions, sound prolongations, and blockages of air. In addition, persons diagnosed

* *Other software packages are available for language transcription and analysis and can be purchased online. The SALT*[Q13] *(Systematic Analysis of Language Transcripts) programs can analyze English and Spanish transcripts.*

with stuttering often exhibit affecting and cognitive reactions to their stuttering[59]; these reactions include anxiety about speaking and beliefs about one's abilities as a speaker.

Approximately 1% of the monolingual population stutters.[51] The reported incidence/prevalence of stuttering in bilingual speakers varies across studies; some report higher rates in bilingual speakers,[60–62] whereas others do not.[63] A recent study of 18 bilingual (Spanish–English) typically fluent children between the ages of 5 years, 6 months and 6 years, 7 months found stuttering-like disfluencies occurring at significantly higher rates (3 to 22%) than expected based typical monolingual (English) norms.[44] Studies with higher prevalence rates have led some researchers to suggest that bilingualism conferred elevated risk for stuttering and that clinicians should advise parents to speak in only one language to their stuttering bilingual child.[60,64] However, such drastic advice needs to be based in strong research evidence—evidence that is currently lacking. Methodological differences in (1) how bilingualism was assessed, (2) who made the diagnosis of stuttering and how close to stuttering onset the diagnosis was made, (3) the languages studied, and (4) when exposure to L2 occurred, make it challenging if not impossible to compare findings across studies.[65]

In terms of who should be making the stuttering diagnosis in bilingual children, recent findings in our laboratory suggest that parents may not be accurate judges of the presence of stuttering in preschool bilingual children. From the 151 children in the longitudinal study described previously, a subset of 26 children (15 boys, 11 girls) were identified by their parents as stuttering at least once between 30 to 60 months of age. Two fluency specialists evaluated the videotaped samples for the presence of stuttering. There were three possible identification options: (1) stuttering present, (2) stuttering not observed, or (3) ambiguous/not enough speech to diagnose. These specialist SLP ratings were then compared with those of the parents. Twenty-six (17%) of the 151 bilingual children were identified by their parents as children who stutter. Based on parental report alone, stuttering is more prevalent in bilingual preschoolers than has been reported for monolinguals (1 to 4%, historically; up to 11% from a recent study[66]). Parents of 21 of the subset of 26 bilingual children (81%) reported that the stuttering occurred in both languages. Of the remaining five, three were reported to stutter in English only and four in Spanish only.

Of the 21 children whose parents reported stuttering in both languages, a total of 35 time points were identified in the children between 2.5 to 5 years of age. Two stuttering specialists then evaluated these 35 samples. The specialists agreed with the parental reports of stuttering presence in only 26% of the samples (9 of 35). Specialists disagreed with parent reports 63% of the time (22 of 35). The remaining 11% (4 of 35) were ambiguous (sparse child speech, unclear speech behaviors).

These findings suggest that parents may overestimate the presence of stuttering in their bilingual children. Why might this be? The specialists identified moments of disfluency that were apparently due to language formulation and/or lexical retrieval difficulties but lacked behavioral and affective features of stuttering (e.g., lack of tension, awareness, struggle; only single iterations of repeated segments; whole-word repetitions only). In all such cases, moments of disfluency were rare during fairly lengthy language samples (<1% disfluent words).

So, does bilingualism confer risk for stuttering in preschoolers? It may depend who you ask. Discrepancies were noticed between parental and professional judgments of stuttering. Thus, some of the reported incidence discrepancies reported across various studies may be due to the differences in judgments about what constitutes stuttering versus language formulation disfluency more commonly seen in bilingual children. Our impression that stuttering may be overdiagnosed in bilingual children's speech is strengthened by two additional findings. First, there was an obviously atypical gender distribution in parent-reported cases (approximately equal numbers of both genders). Second, most children were reported to "stutter" at only a single point, often late in development rather than early in development, but did not show "stuttering" consistently across the sampling period of 30 to 60 months (e.g., "stuttering" at 48 months but not at earlier or later points).

This sample provides a truly unique opportunity to investigate linguistic features of stuttering within a single speaker of two languages close to the onset of stuttering symptoms. Future endeavors include comparing the language abilities of the nine stuttering children to age- and gender-matched controls. Future studies should report standard appraisals of stuttering, such as those derived from SSI-4 Q12 scores, rather than from parental/self-report or percent disfluent words/syllables alone.

CLINICAL APPLICATIONS

In terms of stuttering assessment in bilingual children, SLPs need to know that it is relatively easy to mistake typical disfluencies related to language formulation/lexical retrieval difficulties for stuttering-like disfluencies, particularly if one is not trained to make this distinction. Fortunately, SLPs can accurately gauge stuttering severity even in a language they do not speak themselves.[67] Stuttering assessments in bilingual children should include speech samples assessments in both languages, as well as assessments of affective and cognitive reactions that accompany stuttering. The presence of tension and struggle during speech production, and negative affective and cognitive reactions to speaking situations are signs of stuttering. The nature of the disfluencies seen may also help distinguish the two types: stuttering includes part-word repetitions, sound prolongations, and blocks of air, whereas revisions and restarts are more likely to signal language formulation difficulties. Assessing language comprehension and production in both languages is also crucial, given the relationship between stuttering and language skills,[6] and between typical disfluency and language proficiency.

We do not yet know what constitutes normative amounts of disfluency in bilingual children. The preliminary data reviewed here suggest that at 30 months, bilingual children exhibit disfluency rates in both languages that are similar to monolingual norms (3% disfluency). But by the time bilingual children reach school age, their disfluency rates—and their rates of stuttering-like disfluencies—increase markedly.[44] So, when should we be concerned? Certainly, the presence of large amounts of typical disfluencies would warrant further assessment of the child's language abilities, particularly in terms of word retrieval and language formulation. It is this need for normative information that makes big data initiatives so critical in bilingualism.

Exploring the nature of a child's language formulation disfluencies may be useful in planning treatment targets. Clinicians can use Levelt's model to guide them. Recall from Levelt's speech production model that hesitations and filled pauses signal "upstream" problems at conceptualizer/formulator level, whereas revisions signal problems postproduction of the utterance. Clinicians can use this information to increase or decrease task difficulty. In general, the more difficult the speaking task, the more disfluent the speaker will be.[10,14,32,33,49] Task difficulty can be manipulated by increasing syntactic complexity and/or lexical diversity (formulator and articulator level), and by varying the amount of novel information that the speaker must convey (conceptualizer level). At the conceptualizer level, tasks that require speakers to produce large amounts of abstract, complex information are more challenging than tasks that involve providing concrete information.[53] At the level of the formulator, lexical selection can be made more challenging by choosing targets that are rare in a language, or by choosing targets which are very similar in form to other nontarget words.

CONCLUSION: BILINGUALISM, DISFLUENCY, STUTTERING, AND BIG DATA ARCHIVES

As a research and clinical community, we should consider developing a standardized protocol for data collection with bilingual children. Standardized protocols exist for other big data projects such as AphasiaBank. Standardization insures that the data collected are interpretable and of interest to the larger research and clinical communities. In addition, tools are currently available that allow for fast and accurate analysis of large data sets. These tools (e.g., KIDEVAL utility in CLAN) can assist clinicians and researchers in making their assessments more efficient, and enable them to track language and fluency changes over time.

When developing this protocol, we will need to reach consensus on what information to collect about bilingual children and their language environments. At the very least, we should collect information on variables known to influence language development, such as: percentage of input in each language, SES, and maternal education level. It would also be useful to decide on a set of standardized tests to augment the speech and language data collected in speech samples. Of course these tests will vary by the age of the children being studied as well as the languages that they speak, but we would suggest at least one standardized measure of expressive and receptive vocabulary in each language.

Big data initiatives help answer questions that would otherwise remain unanswerable due to the variation seen in bilingual language acquisition. Large data sets are necessary to control variables known to influence language and fluency development. Both researchers and clinicians can assist in this effort by contributing data (with their client's permission of course!) to the TalkBank archive. In this way we all contribute to the research base of our profession that is crucial to evidence-based practice.

ACKNOWLEDGMENTS AND DISCLOSURES

We thank Dr. Nan Bernstein Ratner for comments on an earlier version of this manuscript. We acknowledge members of the GWU Bi-Lingual Language Development Laboratory: Sebastian Corrales, Diego Martinez, Francesca Rush, Sophia Santa-Cruz, Hannah Rowe, Katherine Winters, and Eleni Ziogas. Portions of this article were presented at the ASHA convention and the World Congress on Fluency Disorders, 2015.

Dr. Brundage received financial compensation as a coinvestigator on 1R01HD068421 to Florida Atlantic University (PI: Hoff).

REFERENCES

1. Grosjean F. Studying Bilinguals. Oxford, UK: Oxford University Press; 2008
2. Kennison S. Bilingualism. In: Salkind N, ed. Encyclopedia of Human Development. Thousand Oaks, CA: Sage Publications; 2005:169–172
3. Kohnert K. Language Disorders in Bilingual Children and Adults. San Diego, CA: Plural; 2007
4. Hoff E, Core C, Place S, Rumiche R, Señor M, Parra M. Dual language exposure and early bilingual development. J Child Lang 2012;39(1):1–27
5. Lennon P. Investigating fluency in EFL: a quantitative approach. Lang Learn 1991;40:387–417
6. Ntourou K, Conture EG, Lipsey MW. Language abilities of children who stutter: a meta-analytical review. Am J Speech Lang Pathol 2011;20(3): 163–179
7. Fox Tree J. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. J Mem Lang 1995;34:709–738
8. Manning W. Clinical Decision Making in Fluency Disorders, 3rd ed. Clifton Park, NY: Delmar; 2010
9. Shriberg EE. Phonetic consequences of speech disfluency. Paper presented at: Proceedings of the International Congress on Phonetic Sciences; 1999 Q14;1:619–622 Q14
10. Bortfeld H, Leon SD, Bloom JE, Schober MF, Brennan SE. Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. Lang Speech 2001;44(Pt 2):123–147
11. de Leeuw E. Hesitation markers in English, German, and Dutch. Journal of Germanic Linguistics. 2007;19(2):85–114 Q15 Q15
12. Eklund R. Crosslinguistic disfluency modeling: a comparative analysis of Swedish and Tok Pisin human-human ATIS dialogues. Paper presented at: ICSLP; 2000 Q16; Beijing, China; 991–994 Q16
13. Eklund R, Shriberg E. Crosslinguistic disfluency modeling: a comparative analysis of Swedish and American English human-human and human-machine dialogues. Paper presented at: International Conference on Spoken Language Processing; 1998 Q17; Sydney, Q18 Australia Q17 Q18
14. Hartsuiker RJ, Notebaert L. Lexical access problems lead to disfluencies in speech. Exp Psychol 2010;57(3):169–177
15. Hlavac J. Hesitation and monitoring phenomena in bilingual speech: a consequence of code-switching or a strategy to facilitate its incorporation. J Pragmatics 2011;43:3793–3806
16. Riazantseva A. Second language proficiency and pausing: a study of Russian speakers of English. Stud Second Lang Acquis 2001;23:497–526
17. Rodríguez LJ, Inés Torres M. Spontaneous speech events in two speech databases of human-computer and human-human dialogs in Spanish. Lang Speech 2006;49(Pt 3):333–366
18. Watanabe MM, Hirose K, Den Y, Minematsu N. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. Speech Commun 2008;50(2):81–94

19. Wu C. Filled pauses in L2 Chinese: a comparison of native and non-native speakers. Paper presented at: Proceedings of the 20th North American Conference on Chinese Linguistics; [Q19]2008;1:213–227 Q19
20. Rodríguez LJ, Torres MI, Varona A. Annotation and analysis of disfluencies in a spontaneous speech corpus in Spanish. Paper presented at: ITRW on Disfluency in Spontaneous Speech; August, 2001[Q20]; Edinburgh, Scotland[Q21] Q20
21. Moreno Sandoval A, Campillos Llanos L, Toledano D. A quantitative study of disfluencies in formal, informal and media spontaneous speech in Spanish. Paper presented at: Proceedings of Iber-Speech; November, [Q22]2012:164–172[Q23] Q22 Q23
22. Campillos L, Alcantara M. Speech disfluencies in formal context: analysis based on spontaneous speech corpora. Paper presented at: Proceedings of the Corpus Linguistics Conference; 2009[Q24] Q24
23. Shriberg EE. Preliminaries to a Theory of Speech Disfluencies. [Ph.D. dissertation]. Berkeley, CA: University of California at Berkeley; 1994
24. Carlo EJ, Watson JB. Disfluencies of 3- and 5-year old Spanish-speaking children. J Fluency Disord 2003;28(1):37–53
25. Watson J, Anderson R. Disfluencies of 2- and 3-year-old Spanish-speaking children from Puerto Rico. Contemporary Issues in Communication Sciences and Disorders. 2001;28:140–150
26. Arnold JE, Fagnano M, Tanenhaus MK. Disfluencies signal theee, um, new information. J Psycholinguist Res 2003;32(1):25–36
27. Arnold JE, Kam CL, Tanenhaus MK. If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. J Exp Psychol Learn Mem Cogn 2007; 33(5):914–930
28. Arnold JE, Tanenhaus MK, Altmann RJ, Fagnano M. The old and thee, uh, new: disfluency and reference resolution. Psychol Sci 2004;15(9):578–582
29. Clark HH, Fox Tree JE. Using uh and um in spontaneous speaking. Cognition 2002;84(1): 73–111
30. Corley M, Hartsuiker RJ. Why um helps auditory word recognition: the temporal delay hypothesis. PLoS ONE 2011;6(5):e19792
31. Ferreira F, Bailey KG. Disfluencies and human language comprehension. Trends Cogn Sci 2004; 8(5):231–237
32. Goldman-Eisler F. Continuity of speech utterance, its determinants, and its significance. Lang Speech 1961;4:220–231
33. Levelt WJM. Speaking: From Intent to Articulation. Cambridge, MA: The MIT Press; 1989
34. Maclay H, Osgood CE. Hesitation phenomena in spontaneous speech. Word. 1959;15:19–44
35. de Bot K. A bilingual production model: Levelt's "speaking" model adapted. Appl Linguist 1992; 13(1):1–24
36. Rispoli M, Hadley P, Holt J. Stalls and revisions: a developmental perspective on sentence production. J Speech Lang Hear Res 2008;51(4):953–966
37. Wagovich SA, Hall NE, Clifford BA. Speech disruptions in relation to language growth in children who stutter: an exploratory study. J Fluency Disord 2009;34(4):242–256
38. Hilton H. The link between vocabulary knowledge and spoken L2 fluency. Lang Learn J 2008;36(2): 153–166
39. Tavakoli P. Pausing patterns: differences between L2 learners and native speakers. ETL Journal. 2010;65:71–79
40. Riggenbach H. Toward an understanding of fluency: a microanalysis of nonnative speech conversations. Discourse Process 1991;14:423–441
41. Schmid MS, Fagersten KB. Disfluency markers in L1 attrition. Lang Learn 2010;60(4):753–791
42. Carias S, Ingram D. Language and disfluency: four case studies on Spanish-English bilingual children. J Multilingual Commun Disord 2006; 4:149–157
43. Bedore LM, Fiestas CE, Pena ED, Nagy VJ. Cross-language comparisons of maze use in Spanish and English in functionally monolingual and bilingual children. Biling Lang Cogn 2006;9(3): 233–247
44. Byrd CT, Bedore LM, Ramos D. The disfluent speech of bilingual Spanish-English children: considerations for differential diagnosis of stuttering. Lang Speech Hear Serv Sch 2015;46(1):30–43
45. Gutiérrez-Clellen V, Kreiter J. Understanding child bilingual acquisition using parent and teacher reports. Appl Psycholinguist 2003;24:267–288[Q25] Q25
46. Calvo A, Bialystok E. Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. Cognition 2014; 130(3):278–288
47. Hart B, Risley T. The Social World of Children Learning to Talk. Baltimore, MD: Paul H. Brookes Publishing; 1999
48. Derwing T, Rossiter M, Munro M, Thomson R. Second language fluency: judgments on different tasks. Lang Learn 2004;54:655–679
49. Schachter S, Christenfeld N, Ravina B, Bilous F. Speech disfluency and the structure of knowledge. J Pers Soc Psychol 1991;60:362–367
50. Schachter S, Rauscher F, Christenfeld N, Crone K. The vocabularies of academia. Psychol Sci 1994; 5:37–41
51. Bloodstein O, Bernstein Ratner N. A Handbook on Stuttering, 6th ed. Clifton Park, NY: Delmar; 2008
52. Clark HH, Wasow T. Repeating words in spontaneous speech. Cognit Psychol 1998;37(3):201–242
53. Skehan P. Modeling second language performance: integrating complexity, accuracy, fluency, and lexis. Appl Linguist 2009;30:510–532

54. Miller JF, Heilmann J, Nockerts A, Iglesias A, Fabiano L, Francis DJ. Oral language and reading in bilingual children. Learn Disabil Res Pract 2006;21(1):30–43
55. Fenson L, Dale P, Reznick J, Thal D, Bates E, Hartung J. The MacArthur Communicative Development Inventories: User's Guide and Technical Manual. San Diego, CA: Singular Publishing Group, Inc.; 1993
56. MacWhinney B. The CHILDES Project: Tools for Analyzing Talk, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2000
57. Bernstein Ratner N, Brundage S. A clinician's complete guide to CLAN and PRAAT. 2015. Available [Q26] at: http://childes.psy.cmu.edu/clan/Clin-CLAN.pdf. Q26
58. Guitar B. Stuttering: An Integrated Approach to its Nature and Treatment. Baltimore, MD: Williams & Wilkins; 1998
59. Yaruss JS, Quesal RW. Stuttering and the international classification of functioning, disability, and health: an update. J Commun Disord 2004;37(1):35–52
60. Howell P, Davis S, Williams R. The effects of bilingualism on stuttering during late childhood. Arch Dis Child 2009;94(1):42–46
61. Travis L, Johnson W, Shrover J. The relation of bilingualism to stuttering. J Speech Disord 1937;2:185–189
62. VanBorsel J, Maes E, Foulon S. Stuttering and bilingualism: a review. J Fluency Disord 2001;26:179–205
63. Au-Yeung J, Howell P, Davis S, Charles N, Sackin S. UCL survey on bilingualism and stuttering. Paper presented at: 3rd World Congress on Fluency Disorders; August, 2000[Q27]; Nyborg, Denmark Q27
64. Karniol R. Stuttering out of bilingualism. First Lang 1992;12:255–283
65. VanBorsel J. Review of the research on the relationship between bilingualism and stuttering. In: Howell P, VanBorsel J, eds. Multilingual Aspects of Fluency Disorders. Bristol, UK: Taylor & Francis; 2011:247–270
66. Reilly S, Onslow M, Packman A, et al. Natural history of stuttering to 4 years of age: a prospective community-based study. Pediatrics 2013;132(3):460–467
67. Lee AS, Robb MP, Ormond T, Blomgren M. The role of language familiarity in bilingual stuttering assessment. Clin Linguist Phon 2014;28(10):723–740

Author Query Form (SSL/00599)

Special Instructions: Author please write responses to queries directly on proofs and then return back.

Q1: AU: Shortened running head OK?
Q2: AU: Please provide degree for all authors.
Q3: AU: if CLAN is an abbreviation, please spell out here. Please provide name and location of software manufacturer.
Q4: AU: Please provide page number for this quote.
Q5: Au: If KIDEVAL is an abbreviation, please spell out.
Q6: AU: Please spell out MLU.
Q7: AU: L2 spelled out correctly?
Q8: AU: if CLAN is an abbreviation, please spell out here, first use in text. Please provide name and location of software manufacturer.
Q9: AU: if FREQ is an abbreviation, please spell out.
Q10: AU: Please spell out MOR.
Q11: AU: Please spell out TD.
Q12: AU: Please spell out SSI-4.
Q13: AU: If SALT is a trademarked name, include manufacturer info.
Q14: AU: Please provide complete date and location of this conference. If presentations were published as a book, please provide name and location of publisher. Please clarify 1: volume 1?
Q15: Medline cannot find the journal "Journal of Germanic Linguistics." (in reference 11 "de Leeuw, 2007"). Please check the journal name.
Q16: AU: Please provide exact date of conference. If published as a book, please provide name and location of publisher.
Q17: AU: Please provide exact date of conference.
Q18: Au: Title of reference 13 is the same as title of reference 12. Please check for accuracy.
Q19: AU: Please provide complete date and location of this conference. If published as a book, please provide name and location of publisher. Please clarify 1: volume 1?
Q20: AU: Please provide exact date.
Q21: AU: Please check this reference.
Q22: AU: Please provide complete date and location of this conference.
Q23: AU: Please check this reference.
Q24: AU: Please provide complete date and location of this conference.
Q25: AU: Please check the reference for accuracy.
Q26: AU: Please provide accessed date.
Q27: AU: Please provide exact date of conference.