

ARTICLE

Uh and *eh* signal novelty for monolinguals and bilinguals: evidence from children and adults

Elizabeth MORIN-LESSARD* and Krista BYERS-HEINLEIN

Concordia University

*Corresponding Author: Concordia University, Department of Psychology, 7141 Sherbrooke Street West, Montréal, QC, H4B 1R6. Email: morinlessard@gmail.com

(Received 23 April 2018; revised 10 September 2018; accepted 10 December 2018; first published online 4 March 2019)

Abstract

Previous research suggests that English monolingual children and adults can use speech disfluencies (e.g., *uh*) to predict that a speaker will name a novel object. To understand the origins of this ability, we tested 48 32-month-old children (monolingual English, monolingual French, bilingual English–French; Study 1) and 16 adults (bilingual English–French; Study 2). Our design leveraged the distinct realizations of English (*uh*) versus French (*eh*) disfluencies. In a preferential-looking paradigm, participants saw familiar–novel object pairs (e.g., *doll–rel*), labeled in either Fluent (“Look at the *doll/rel!*”), Disfluent Language-consistent (“Look at *thee uh doll/rel!*”), or Disfluent Language-inconsistent (“Look at *thee eh doll/rel!*”) sentences. All participants looked more at the novel object when hearing disfluencies, irrespective of their phonetic realization. These results suggest that listeners from different language backgrounds harness disfluencies to comprehend day-to-day speech, possibly by attending to their lengthening as a signal of speaker uncertainty. Stimuli and data are available at <<https://osf.io/qn6px/>>.

Keywords: speech disfluencies; bilinguals; children; speaker knowledge

Speech is a well-structured signal, and regularities in the speech signal allow listeners to form predictions as they listen to speech in real time. For example, familiarity with English syntax allows listeners to predict that the sentence “Cats like to chase ...” will likely end with the noun *mice*. Listeners can also make predictions from seemingly erratic parts of speech that provide predictive information. For example, unintentional hesitation markers such as *uh* tend to appear at the start of syntactic clauses (Clark & Wasow, 1998), and to be produced immediately after the first word of a new clause, which tends to be *the* in English, for example: “Did you see *thee, uh, chameleon* in the tree?”. *Uhs* and *ums* also tend to precede words that are new or previously unmentioned in the discourse (Arnold, Kam, & Tanenhaus, 2007; Arnold & Tanenhaus, 2007), since speakers may be less acquainted with these new or infrequent

Corresponding author’s current address: Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary, AB, T2N 1N4.

© Cambridge University Press 2019

terms (Smith & Clark, 1993). Monolingual children and adults use such speech disfluencies to predict that a speaker will refer to a new or discourse-new object (Arnold *et al.*, 2007; Kidd, White, & Aslin, 2011; Orena & White, 2015). However, the basis of this prediction is not clear. The current study asks what makes speech disfluencies useful for real-time speech processing for children and adults, by investigating the processing and use of disfluencies by monolingual and bilingual listeners.

The nature of speech disfluencies

Speech disfluencies provide a valuable cue to help language listeners efficiently process natural speech, as they occur regularly in predictable situations. Speakers commonly produce non-pathological stumbles that affect the flow of speech, such as filled pauses (e.g., *uh*, *um*), repetitions (e.g., *the the ...*), and silent pauses (Fox Tree, 1995). The frequency and type of disfluencies produced vary based on factors such as the complexity of a topic (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001), and uncertainty about a topic (Smith & Clark, 1993). Disfluencies are especially likely to occur before a difficult or unfamiliar word, or one that is new in the speech context (Arnold, Losongco, Wasow, & Ginstrom, 2000; Arnold & Tanenhaus, 2007). For example, when English speakers anticipate a short delay (e.g., before producing a new or infrequent word), they often produce filled pauses such as *uh* (henceforth 'speech disfluencies' or 'disfluencies'; Clark & Fox Tree, 2002).

The exact realization of speech disfluencies depends on the language being spoken. For example, typical English disfluencies are *uh* and *um* (Shriberg, 2001). However, typical French disfluencies are *euh* or *eum* (Duez, 1982). The primary difference between these disfluencies is in the vowel, which may be due to each language having a neutral vowel that is close to the vowel of a common determiner in a language (e.g., *uh* in English is close to *the*, and *euh* in French is close to *le*). In the case of English and French, each disfluency only includes vowels present in each respective language. In other words, the vowel sound *uh* is not in the Canadian French phonemic inventory, and conversely the vowel sound *euh* is not in the Canadian English phonemic inventory. Although the realization of disfluencies differs between languages, the change in phonemes between different languages (e.g., *uh* vs. *euh*) does not change the meaning of the disfluency per se. It may merely suggest the language to which the disfluency belongs. Importantly, unlike words, speech disfluencies make no semantic contribution to an utterance.

Listeners' use of speech disfluencies across development

Can listeners use speech disfluencies to make predictions about language? As noted above, disfluencies occur in predictable locations in the speech stream – often when a speaker is uncertain. Research with monolingual adults has demonstrated that the English filled pause *uh* helps them recognize words on a screen faster compared to when *uh* has been edited out (Fox Tree, 2001). Other research with English-speaking adults has revealed that disfluent discourse leads listeners to expect that a speaker will refer to novel or unfamiliar objects (Arnold *et al.*, 2007; Barr & Seyfeddinipur, 2010), whereas fluent discourse makes listeners expect that a speaker will continue talking about previously mentioned objects (Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Tanenhaus, Altmann, & Fagnano, 2004). Building on these findings, recent research suggests that the predictive use of disfluencies may also depend on

situation- and speaker-specific characteristics, such as when a speaker has difficulty naming novel words, or is producing disfluencies in a non-native language (Arnold *et al.*, 2007; Bosker, Quené, Sanders, & de Jong, 2014; Heller, Arnold, Klein, & Tanenhaus, 2015). For these reasons, when listeners hear a disfluency such as *uh*, they can predict that a speaker will refer to a label that is difficult to name and therefore potentially also novel.

Research with monolingual children suggests that speech disfluencies also support language comprehension in younger learners, who use disfluencies to predict which object a speaker will name. In a study particularly pertinent to the present research, Kidd *et al.* (2011) investigated 16- to 32-month-old children's real-time processing of sentences containing a speech disfluency. Monolingual English children viewed object pairs. Half of the objects were novel objects with nonsense labels (e.g., *mog*), while the other half were familiar objects with familiar labels (e.g., *shoe*). On each trial, a familiar–novel object pair was presented three times. On the first two presentations, the familiar object was named to establish its familiarity in the discourse (e.g., “I see the *shoe*! Ooh, what a nice *shoe*!”). On the third presentation, either the familiar or novel target word was named in a fluent sentence (e.g., “Look! Look at the shoe/mog!”) or a disfluent sentence (e.g., “Look! Look at *thee uh shoe/mog*!”). Children's looking at the two objects was measured from the onset of the disfluency until just before the speaker uttered a label (i.e., the disfluency period). Children aged 28–32 months reliably looked predictively towards the novel object when they heard the disfluency. However, children aged 16–20 months did not appear to use the disfluency to predict the speaker's intended target, suggesting that the ability to use disfluencies predictively appears over the course of development.

Subsequent research has extended this work to show that children's predictive use of speech disfluencies is robust with previously reliable speakers, but not with previously unreliable speakers (Orena & White, 2015). Disfluencies may also provide cues to children about a speaker's preference for particular objects (Thacker, Chambers, & Graham, 2018). Additionally, three-year-olds can use disfluencies to predict that a speaker will label an object that is perceptually familiar but novel to the discourse (e.g., a familiar object that has not previously been mentioned; Owens & Graham, 2016). However, while adults can use filled pauses to anticipate reference solely based on object novelty, three- and five-year-olds require that an object be new in the discourse to anticipate reference to a novel object (Owens, Thacker, & Graham, 2018). Together, findings from real-time comprehension studies with monolingual children suggest that their ability to use speech disfluencies predictively emerges around their second birthday, and that they flexibly adapt their predictions based on both speaker and context.

Speech disfluencies as informative cues

Important developmental questions remain about what makes speech disfluencies informative to listeners. One possibility is that children learn to use disfluencies predictively through experience with specific linguistic forms: they learn that *uh* and *um* often precede the labeling of novel or discourse-new referents. Both children and adults are highly sensitive to the statistical regularities present in human language, and they use these regularities for language acquisition and processing. For example, young infants are sensitive to when particular syllables co-occur or pattern in specific ways (Gómez & Gerken, 2000; Saffran, Newport, & Aslin, 1996). Experience with a

language's patterns of co-occurrence allows listeners to gain knowledge of the structure and constraints of their language.

Knowledge about co-occurrence patterns in language can be used predictively during comprehension. For example, both children and adults predict that a speaker is more likely to refer to a food than a non-food item if they have heard the verb *eat* (e.g., Friedrich & Friederici, 2005; Snedeker & Trueswell, 2004). In the context of speech disfluencies, listeners could leverage their repeated experience hearing specific disfluencies prior to novel or discourse-new words to form predictions that a disfluency is likely to precede a novel or discourse-new label. It may be that children become able to use disfluencies predictively once they accrue sufficient experience that disfluencies such as *uh* tend to precede a speaker's labeling of a novel or discourse-new object. Moreover, this account would predict that children would be best able to use typical, frequently encountered disfluencies as a basis to make predictions about what will follow in the sentence, but would be less able to use atypical, non-native disfluencies.

Some evidence from the adult literature supports this account, showing that the form of the speech disfluency could affect its use in speech prediction. In a study that compared adults' predictions in the context of disfluencies produced by either a native versus an accented speaker, adults increased attention to low-frequency referents (e.g., a sewing machine) only when hearing the native speaker (Bosker *et al.*, 2014). The main interpretation of this result was that listeners attributed the disfluency to retrieval difficulty in the native speaker, but not in the accented speaker. However, an alternate possibility is that the non-native realization of the disfluency itself disrupted adults' ability to use the disfluency predictively. In support of this idea, over the course of the study, listeners in the native condition increased looking to the low-frequency referent upon hearing the disfluency, suggesting that adults detected and learned the association between disfluencies and subsequently labeled referents. However, as both the disfluency and the rest of the sentence were accented, this study does not provide conclusive evidence of whether atypical pronunciation of the disfluency itself disrupted its predictive use. Nonetheless, if listeners' predictions are less robust when hearing a non-native disfluency, this would suggest that such predictions are, at least in part, driven by experience hearing the specific native realization of a disfluency prior to a speaker naming novel or discourse-new referents.

In contrast, it is also possible that listeners can flexibly use a range of speech disfluencies, whether native or non-native, to predict that a speaker will label a novel object. Although languages differ in the typical realizations of disfluencies, disfluencies often involve lengthening (Fox Tree & Clark, 1997). Further, even within a single language, there can be a variety of typical disfluencies, such as *uh*, *um*, and *hmm* in English. Moreover, whatever their form, disfluencies typically occur because a speaker is uncertain about the choice of a word, or is having retrieval difficulties (Beattie & Butterworth, 1979). Thus, it is also possible that children learn that lengthening in general, either by itself or as a marker of speaker uncertainty, can provide a cue about what a speaker is likely to label. In this case, listeners might be able to make equal use of native and non-native disfluencies.

Speech disfluencies and bilingualism

Comparisons between monolinguals and bilinguals can help illuminate the role of language experience in children's predictive use of disfluencies. This is because bilingual

individuals are unique in their experience with disfluencies. The obvious distinction between monolingual and bilingual environments is that, while monolinguals hear only one language, bilinguals hear two. As such, English monolinguals are only exposed to disfluencies specific to the English language, such as *uh*, but do not encounter the *eu*h that is typical in French. Likewise, French monolinguals are only exposed to the French *eu*h, but not the English *uh*. In contrast, English–French bilinguals hear both *uh* and *eu*h. These two disfluencies could be heard in different proportions based on bilinguals' relative exposure to English and French, and might sometimes be encountered in a 'code-switched' manner (Poplack, 1980), such that the language of the disfluency does not match the language of the sentence (e.g., the English *uh* appears in a French sentence, as in "Regarde le *uh* ... chien!").

These distinctions between monolingual and bilingual experiences highlight the fact that the two language groups have different experiences with speech disfluencies. If language-specific experience with particular phonetic forms is important for the development of the predictive use of disfluencies, then bilinguals might develop this ability later than monolinguals as they have less exposure to disfluencies in each language (but likely hear as many total disfluencies across their two languages). On the other hand, we might also find that monolinguals and bilinguals are similar in when they begin to use disfluencies for prediction. Such findings would be evidence against the idea that children begin to use disfluencies predictively when they notice that specific syllables – such as *uh* or *um* – tend to precede novel or discourse-new labels.

The present research

The main goal of this research was to investigate the mechanism underlying the ability to use disfluencies to predict the content of upcoming speech. To do this, our study compared listeners from two different language backgrounds, who diverge in their experience with the specific realization of speech disfluencies: monolinguals and bilinguals. We also leveraged the fact that languages differ in how they realize disfluencies, and that listeners' experience is limited to the type of disfluencies present in their native language(s). Finally, we took a developmental perspective, by investigating both children (aged 32 months: Study 1) and adults (Study 2).

Expanding on Kidd *et al.*'s (2011) study with monolingual English children, we tested monolingual and bilingual 32-month-olds and bilingual adults using a preferential-looking paradigm. Thirty-two months was chosen because at this age monolinguals are consistently able to use disfluencies predictively (Kidd *et al.*, 2011). English and French monolingual children were tested on their predictive use of two types of disfluencies: those consistent with their native language and those inconsistent with their native language. Having English and French monolinguals allowed us to replicate Kidd *et al.*'s study, to test the generalizability of their results, and to test whether monolinguals' predictive use of disfluencies is limited to native-like disfluencies. We also tested English–French bilingual children and adults on their comprehension of language-consistent and language-inconsistent disfluencies, in both of their languages. Findings from these child and adult bilingual groups are key in helping us understand whether the predictive use of speech disfluencies – *uh* and *eu*h in particular – is tied to a specific language, and whether their use emerges later for bilinguals than for monolinguals.

We hypothesized that experience with the specific syllables that are characteristic of a language's disfluencies allows listeners to form predictions about the speech that follows. Thus, we predicted that listeners' predictive use of disfluencies would differ

depending on the specific realization of the disfluency (language-consistent vs. language-inconsistent) and their own language background (monolingual vs. bilingual). Alternatively, it might be that listeners are highly flexible in their use of speech disfluencies, in which case children and adults across different language backgrounds might be equally able to use native, non-native, and code-switched disfluencies in predicting that a speaker will label a novel referent.

Study 1: children

Method

The present research was approved by the Human Research Ethics Committee at Concordia University, and all children were treated in accordance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2; CIHR, NSERC, & SSHRC, 2014). Parents provided informed consent prior to the study.

Study 1 stimuli and data can be found on the Open Science Framework at <<https://osf.io/qn6px/>>.

Participants

Forty-eight 32-month-old children were included in the final sample. Children were recruited from government birth lists, and lived in Montréal, Canada. All children were healthy and typically developing, with no reported hearing or vision problems. Following Kidd *et al.* (2011), we aimed for a sample of 16 participants per group. There were 16 monolingual English children ($M_{\text{age}} = 32.18$, $SD = 17$ days, range = 31.29 to 33.20, females = 10), 16 monolingual French children ($M_{\text{age}} = 32.18$, $SD = 15$ days, range = 31.26 to 33.9, females = 6), and 16 English–French bilingual children ($M_{\text{age}} = 32.22$, $SD = 13$ days, range = 31.27 to 33.7, females = 7). Of the sixteen bilinguals, 7 were English-dominant ($M = 65\%$ English exposure, $SD = 5.2\%$) and 9 were French-dominant ($M = 62\%$ French exposure, $SD = 8.9\%$). On average, bilinguals were exposed to their dominant language 64% of the time (range: 52–74%), and to their non-dominant language 36% of the time (range: 26–48%). Monolingual children were exposed to either English or French at least 90% of the time. English–French bilingual children were exposed to each of their languages a minimum of 25%, and did not have more than 10% exposure to a third language. All bilinguals had acquired their two languages from birth except for three children, who acquired their second language upon entering daycare at 10, 12, and 14 months, respectively. For bilingual children, language dominance was established based on exposure, with the language most often heard considered the dominant language (note that no children had perfectly balanced exposure). Another 29 children were tested but not included in the final sample due to failure to meet pre-established language criteria for monolingualism or bilingualism (15), low birth weight or premature birth (4), fussiness/inattention (5), reported health issues (2), parental interference (2), or equipment error (1). Data collection for this study began in June 2015, and was completed in June 2017.

Measures

An adaptation of the Language Exposure Questionnaire (Bosch & Sebastián-Gallés, 2001), implemented with the Multilingual Approach to Parent Language Estimates

(MAPLE; Byers-Heinlein *et al.*, 2018), was used to assess input from different caregivers across the child's life, and to measure monolinguals' and bilinguals' exposure to their language(s). In a semi-structured interview, the primary caregiver was asked about the family's language background, the child's home environment, daily activities, and typical daily routines. This interview yielded an estimate of the total number of hours spent in each language over the course of the child's life, which was converted to a percentage, and was then averaged with the parent's estimate to yield final exposure percentages.

Finally, a demographics questionnaire gathered information about the family's background and the child's health history.

Stimuli

Visual stimuli consisted of 32 familiar–novel object pairs appearing against a white background. The familiar objects were imageable words from the MacArthur-Bates Communicative Development Inventories (MCDI) vocabulary norms in American English (Fenson *et al.*, 2007) and Québec French (Boudreault, Cabirol, Trudeau, Poulin-Dubois, & Sutton, 2007). Based on MCDI norms, these objects were known by at least 50% of 18-month-olds in English, and by at least 50% of 16-month-olds in French (French norms for 18-month-olds were not available). Images of familiar objects were selected from Google, and images of novel objects were selected from the NOUN-2 Database (Horst & Hout, 2016).

Auditory stimuli were recorded by a native bilingual female speaker of Canadian English and Québec French with no noticeable accent in either language, who produced stimuli in child-directed speech. Auditory stimuli were normalized to a comfortable hearing level of 70dB. Novel words were chosen to sound like possible words in English or French. Names for novel words in English were selected from the NOUN-2 Database (Horst & Hout, 2016), from Kidd *et al.* (2011), or were created for this study. Novel English and French words had the same number of syllables (see Table 1).

Familiar and novel objects were paired for presentation on each trial, and matched for image size and color salience. Auditory stimuli in each pair were phonologically distinct in place of articulation at word onset, did not rhyme, were matched on number of syllables, and in French were matched for grammatical gender (Lew-Williams & Fernald, 2007). Half of French trials presented feminine words (e.g., *la maison* and *la t  lue*), and the other half masculine words (e.g., *le soulier* and *le pafli*).

The sequence of each trial followed Kidd *et al.* (2011), such that novel labels were also discourse-new (see Owens & Graham, 2016, and Owens *et al.*, 2018, for evidence that discourse novelty is a driver of disfluency effects in children). There were three presentations of each object label. On the first presentation, the familiar object was named in an English or a French sentence (e.g., English: "Look at the doll!") as the two images appeared. A black screen then appeared for 1 second. On the second presentation, the same familiar–novel object pair appeared and was named in English or French, in a different sentence (e.g., English: "Ooh! What a nice doll!"). This was again followed by a black screen for 2.5 seconds. On the third presentation, the familiar–novel object pair was shown again. On half of the trials the familiar object was named, and on the other half the novel object was named. Crucially, in this third sentence, half of the trials (8) were fluent, and half (8) were disfluent. On fluent trials, the final sentence contained no disfluency. Of the

Table 1. Familiar–novel object pairs in English and in French

English order		French order		
Familiar	Novel	Det.	Familiar	Novel
book	semp	le	bas [sock]	sèpe
bread	mog	la	bouche [mouth]	taque
cheese	gorp	la	chaise [chair]	moube
cookie	posha	le	chapeau [hat]	brussin
cow	sarl	la	cuillère [spoon]	dimotte
cup	blat	la	dent [tooth]	shème
dog	koob	le	lait [milk]	fide
doll	rel	le	lit [bed]	juve
frog	dak	la	main [hand]	froise
key	mip	la	maison [house]	télue
pig	juff	le	nez [nose]	migue
plate	zel	le	pied [foot]	vuche
stroller	voopi	la	pomme [apple]	kète
toothbrush	glindle	la	porte [door]	tine
truck	shob	le	soulier [shoe]	pafli
window	teeba	le	verre [glass]	triffe

Notes. Familiar and novel words in English and French. Masculine (*le*) and feminine (*la*) determiners (Det. column) are indicated for familiar–novel pairs in French. English translations of French words are indicated in square brackets. Familiar and novel words in English and French were not translations, to ensure that words were equally familiar or novel to both monolinguals and bilinguals (who saw both English and French orders).

disfluent trials, half (4) were Disfluent Language-consistent, where the final sentence had an elongated determiner (*thee/lee/laaa*) and a filled pause consistent with the language of the sentence (e.g., English: “Look at thee uh doll/rel”). The other half of disfluent trials (4) were Disfluent Language-inconsistent, where the final sentence contained an elongated determiner and a filled pause inconsistent with the language of the sentence (e.g., English: “Look at thee euh doll/rel”). As in Kidd *et al.* (2011), *the* was elongated for disfluent trials (pronounced *thee* instead of *thuh*), given that this is typical of determiners preceding disfluencies in natural speech (Fox Tree & Clark, 1997). Since the realization of determiners in both English and French can change when preceding a vowel, all words started with consonants. Each trial lasted 18.5 seconds.

A total of eight study orders were created (4 in English, 4 in French) that counterbalanced the side of presentation of each object, and which object was labeled on each trial (see Figure 1). All sentences within an order were in the same language. Trials were presented quasi-randomly, such that no more than two trials of the same type (Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent) appeared in a row. One of five animated attention-grabbing objects accompanied by music were presented in the center of the screen between each trial to recapture the child’s attention, and to direct their gaze to the center of the screen.

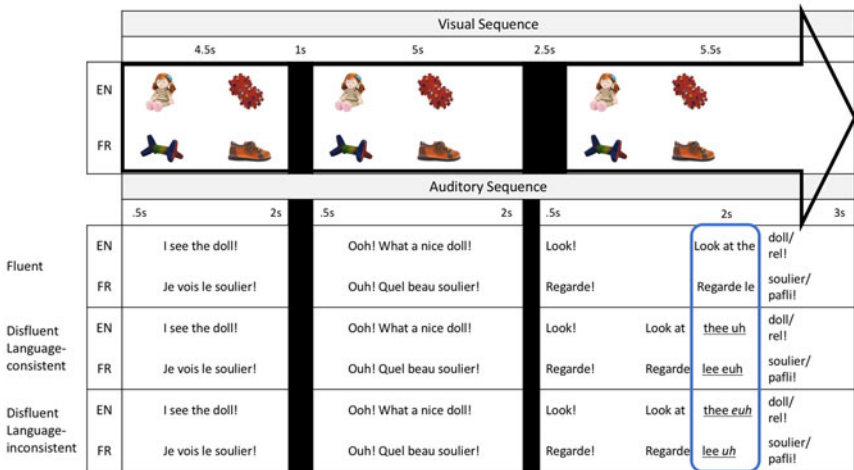


Figure 1. The audio-visual structure of trials in English and French, for each Trial type (Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent). EN denotes an English trial, whereas FR denotes a French trial. The 2 s Disfluency Window of analysis before the onset of the target word in the third sentence is boxed.

Monolingual children were tested in only one order in their native language (randomly-assigned), and saw a total of 16 trials. Bilinguals were tested in two orders (one randomly-assigned English and one French) and saw a total of 32 trials. For bilinguals, the order of presentation of English and French was counterbalanced.

Procedure

A Tobii T60XL eye-tracker was used to present the stimuli on a 24-inch screen, while eye-gaze data were gathered at a sampling rate of 60 Hz. Children sat on their parent’s lap or by themselves on a chair in a dimly lit soundproof room, approximately 60 cm away from the screen. The caregiver was instructed not to interact with their child during the study, and wore opaque sunglasses and headphones through which music was played to ensure that they were blind to the experimental condition. The study began following a 5-point infant calibration routine to calibrate the eye-tracker to children’s eyes. A play break was given to bilingual children between the two experimental orders if needed. The total duration of the experiment was approximately 6 min for monolinguals and 12 min for bilinguals. Questionnaires were completed following the study.

Results

Analytic strategy

Consistent with Kidd and colleagues’ study (2011), the crucial part of the present study was to assess looking during the disfluency period on disfluent trials, compared to an equivalent window on fluent trials. This was a 2-second window before the onset of the target word in the third sentence of each trial, during which participants heard the speech disfluency. The target word always occurred 15 s into each trial (4000 ms into sentence 3), meaning that the Disfluency Window was from 13 to 15 s into

each trial (2000–4000 ms into sentence 3). The window was shifted by 250 ms as in Kidd *et al.* to allow for time needed for children to launch an eye-movement. This window (2250–4250 ms) establishes the predictive nature of disfluencies; if children look to the novel object as they hear a disfluency, and importantly, BEFORE they hear a target word, this means that they effectively used disfluencies to predict that the speaker was going to label a novel object. In addition to the main question of children's looking during the disfluency period, secondary analyses were conducted to assess looking to the target object (vs. the distractor) after it was named in the Label Window, to examine children's processing of the familiar and novel words.

Eye-gaze data collected were summed across time for each trial and within each area of interest using the R package *eyetrackingR* (Dink & Ferguson, 2015; R Core Team, 2017). On each trial, rectangular areas of interest were established about 2 cm around each object. The proportion of looking to the novel object, obtained by dividing the total duration of fixation at the novel object divided by the sum of the total duration of fixation at both the novel and familiar objects, was the main dependent variable. Analyses were performed separately for monolingual and bilingual children, as monolinguals completed the study only in their native language, while bilinguals completed the study in their two languages (dominant and non-dominant).

Monolinguals

Disfluency Window

Average proportion of looking time data are displayed in Figure 2. Preliminary analyses revealed no differences as a function of gender or between the monolingual groups (English monolinguals, French monolinguals), therefore data were collapsed over these factors.

We conducted a one-way repeated-measures ANOVA with Trial type (3 levels: Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent) as the independent variable, and proportion of looking to the novel object as the dependent variable. Results showed a significant main effect of Trial type, [$F(2,62) = 4.99$, $p = .01$, $\eta^2 = 0.073$]. Follow-up paired-samples *t*-tests revealed that monolinguals looked less to the novel object on Fluent trials ($M = .51$, $SD = .17$) than on Disfluent Language-consistent trials [$M = .61$, $SD = .21$, $t(31) = -2.523$, $p = .02$, $d = 0.52$], and on Disfluent Language-inconsistent trials [$M = .62$, $SD = .20$, $t(31) = -2.534$, $p = .017$, $d = 0.59$]. More central to the current research question, there was no difference between Disfluent Language-consistent ($M = .61$, $SD = .21$) and Disfluent Language-inconsistent ($M = .62$, $SD = .20$) trials [$t(31) = -0.310$, $p = .76$, $d = 0.05$]. Results of follow-up *t*-tests compared to chance are presented in Table 2. Monolinguals' proportion of looking to the novel object was above chance on both Disfluent Language-consistent trials and on Disfluent Language-inconsistent trials, but not on Fluent trials. Together, these results replicate previous findings that monolinguals of this age use disfluencies predictively, while replicating this finding in a new group (French monolinguals), and demonstrate that monolinguals are flexible in this capacity across the specific realization of disfluencies.

Label Window

See Figure 3 for a timecourse of average looking to the target. To ensure that monolinguals looked at the target object once it was named, we analyzed looking in

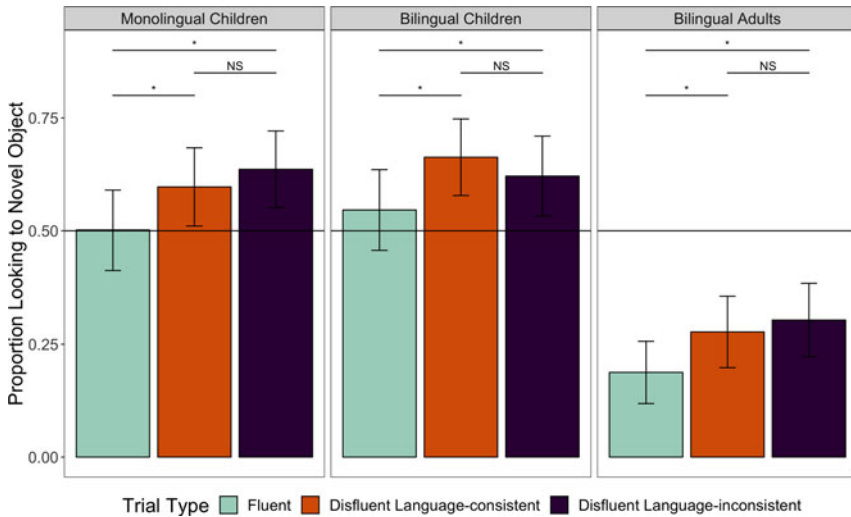


Figure 2. Monolingual children's, bilingual children's, and bilingual adults' looking to the novel object in the Disfluency Window on Fluent (8 trials), Disfluent Language-consistent (4 trials), and Disfluent Language-inconsistent (4 trials) trials. Error bars represent the standard error of the mean. NS denotes $p > .05$, and asterisks denote $p < .05$.

a 2 s window anchored on the target word onset (4000 ms into sentence 3) on each trial type (Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent). As with the Disfluency Window, the Label Window was shifted by 250 ms to allow for the launching of eye-movements (4250–6250 ms). We also assessed whether it was more difficult for monolinguals to identify novel targets compared to familiar targets. A 3 (Trial type: Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent) \times 2 (Target type: Familiar, Novel) repeated-measures ANOVA on the proportion of looking to the target revealed no significant main effect of Trial type, [$F(2,59) = 0.10, p = .91, \eta^2 = 0.082$], and no main effect of Target type, [$F(1,59) = 0.72, p = .41, \eta^2 = 0.613$]. There was also no significant interaction between Trial type and Target type, [$F(2,350) = 1.81, p = .17, \eta^2 = 1.805$]. These results suggest that looking to the target once it was labeled for monolingual children did not differ based on different trial types, nor based on whether the target was novel or familiar. Moreover, children looked at the labeled target significantly above chance on all trial types ($ps < .05$), suggesting that they could equally identify the target on familiar and novel label trials.

Bilinguals

Disfluency Window

Results for bilinguals are displayed in Figure 2. Preliminary analyses revealed no effect of gender, of language of testing (dominant language, non-dominant language), nor of order of testing (dominant language first, non-dominant language first; all $ps > .05$), thus subsequent analyses were collapsed across these factors.

A repeated-measures ANOVA was conducted with Trial type (Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent) on the proportion of looking

Table 2. Disfluency Window analyses of looking to the novel object on Fluent, Disfluent Language-consistent, and Disfluent Language-inconsistent trials compared to chance (0.5) in Study 1 (monolingual and bilingual children) and Study 2 (bilingual adults)

	Monolinguals					Bilinguals				
	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	<i>d</i>
Study 1										
Fluent	.51	.17	0.210	.84	0.04	.54	.21	0.961	.34	0.24
Dis. lang.-consist.	.61	.21	2.966	.006	0.52	.66	.23	3.801	.001	0.69
Dis. lang.-inconsistent.	.62	.20	3.474	.002	0.61	.65	.26	3.054	.005	0.56
Study 2										
Fluent						.17	.17	-11.140	.001	1.97
Dis. lang.-consist.						.30	.26	-4.238	.001	0.69
Dis. lang.-inconsistent.						.30	.26	-4.242	.001	0.56

Notes. *M* = mean. *SD* = standard deviation. *T* values are comparisons to chance looking (0.5).

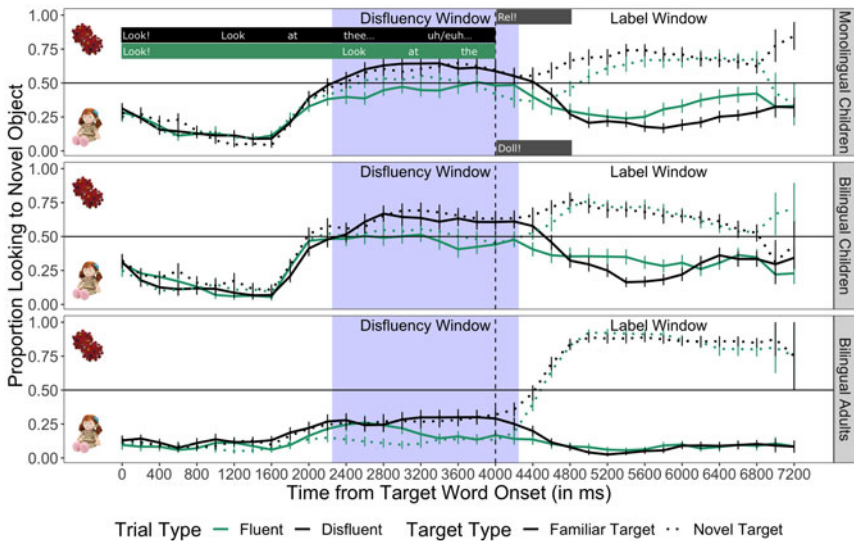


Figure 3. Monolingual children's, bilingual children's, and bilingual adults' looking to familiar and novel target objects following the target word onset. The shaded area represents the 2 s Disfluency Window before the target word onset (2250–4250 ms), which was shifted by 250 ms in the analysis to account for stimulus response latency. The Label Window corresponds to the area following the Disfluency Window (4250–6250 ms), which was also shifted by 250 ms. The vertical dashed line indicates the target word onset. Chance looking (.50) is represented by the horizontal line.

to the novel object, and revealed a significant main effect of Trial type, [$F(2,59) = 8.06$, $p < .001$, $\eta^2 = 0.215$]. *T*-tests were used to follow up on the main effect of trial type. Bilinguals showed a similar pattern to monolinguals. They looked significantly less to the novel object on Fluent trials ($M = .54$, $SD = .21$) than on Disfluent Language-consistent trials [$M = .66$, $SD = .23$, $t(30) = -3.274$, $p = .003$, $d = 0.77$], or on Disfluent Language-inconsistent trials [$M = .65$, $SD = .26$, $t(29) = -2.859$, $p = .008$, $d = 0.67$]. As with monolinguals, there was no difference in looking for bilinguals between Disfluent Language-consistent and Disfluent Language-inconsistent trials [$t(29) = 0.254$, $p = .80$, $d = 0.04$]. Results of follow-up *t*-tests compared to chance are presented in Table 2. Looking to the novel object was statistically above chance on both Disfluent Language-consistent trials and on Disfluent Language-inconsistent trials, but not on Fluent trials. These comparisons suggest that bilingual children were successful at using disfluencies to predict novelty, and that they were equally successful when disfluencies were consistent and inconsistent with the language of the sentence.

Label Window

Last, we assessed whether bilinguals looked to the target object after it was labeled (see Figure 3 for a timecourse of average looking results). We analyzed looking in the 2 s after the target object was labeled. A 3 (Trial type: Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent) \times 2 (Target type: Familiar, Novel) repeated-measures ANOVA on the proportion of looking to the target revealed no main effect of Trial type [$F(2,25) = 0.89$, $p = .42$, $\eta^2 = 0.072$]. However,

there was a marginally significant main effect of Target type [$F(1,57) = 2.98, p = .09, \eta^2 = 0.052$], suggesting that bilinguals may not look to novel targets as much as familiar targets once they have been labeled. Finally, there was no significant interaction between Trial type and Target type [$F(2, 56) = 1.24, p = .30, \eta^2 = 0.043$]. These results suggest that bilingual children's looking to the target differed when the target was familiar vs. novel, but not when trials were fluent or disfluent.

We further investigated looking to familiar versus novel targets using *t*-tests. Overall, bilingual children spent more time looking at the target when its label was familiar ($M = .71, SD = .27$) than when its label was novel [$M = .61, SD = .23, t(30) = 2.26, p = .03, d = 0.41$]. However, looking to both familiar [$t(30) = 17.582, p < .001, d = 6.42$] and novel [$t(30) = 39.397, p < .001, d = 14.38$] targets was significantly above chance (.50), and this result held when looking at this factor by Trial type ($ps < .05$). Note that success at novel label trials requires the use of a disambiguation heuristic such as mutual exclusivity, a point we will return to in the discussion. These findings suggest that bilingual children could identify the referent of both familiar and novel labels, but that their performance was less robust for novel labels.

Discussion

Study 1 tested monolingual and bilingual 32-month-olds on their responses to disfluencies that were language-consistent and language-inconsistent. Performance upon hearing the disfluency was similar for both monolinguals and bilinguals: All children successfully used disfluencies to predict novelty, whether or not the language of the disfluencies was consistent with the language of the sentence. Contrary to our hypothesis, there was no effect of any of our manipulations. We did not find evidence that children benefit from experience with specific realizations of disfluencies. Instead, children appear to be highly flexible, using both same-language and code-switched disfluencies to predict that a speaker will label a novel object.

While our main interest was in children's performance during the disfluency period, we also examined their looking towards the labeled target object. Importantly, all children looked at the target object above chance. For monolinguals, looking was similar for familiar and novel labels. However, bilinguals' looking was less robust during novel label trials. Indeed, a direct comparison revealed a significant difference between monolinguals and bilinguals on novel label trials ($t = 3.590, p < .001$) but not on familiar label trials ($t = 0.179, p = .51$). In our task, children had to infer the meaning of the novel label using a heuristic such as mutual exclusivity (i.e., knowing that objects typically only have one label; Markman & Wachtel, 1988). Our findings are in line with previous reports that bilinguals are less robust in their use of mutual exclusivity than monolinguals, presumably because bilingual children learn two labels for each object (e.g., Byers-Heinlein & Werker, 2009, 2013; Houston-Price, Caloghris, & Raviglione, 2010). Yet, this monolingual-bilingual difference has not been found in all studies (Byers-Heinlein, Chen, & Xu, 2014; Frank & Poulin-Dubois, 2002). We note that, while many infant studies have used looking paradigms and found monolingual-bilingual differences, looking-based paradigms have only rarely been used to compare monolinguals and bilinguals older than age two years. Our current results raise the possibility that looking-based measures might be a more sensitive way for future studies to examine this question.

One possible alternate explanation of our disfluencies results is that children were leveraging their specific experience with disfluencies, but that the difference between

uh and *eu*h was too subtle for children to discriminate between the two. In this case, language-consistent and language-inconsistent disfluencies would have been perceived identically, and that is why children showed no difference in their performance. This explanation is somewhat unlikely, since previous research suggests that children of our participants' age and much younger can differentiate small speech-sound contrasts (see review by Byers-Heinlein & Fennell, 2014). Nonetheless, a discrimination task is needed to rule out the possibility that children could not perceive *uh* and *eu*h as two different sounds.

While it would be ideal to test 2.5-year-old children on their discrimination of the English and French disfluencies, this is methodologically challenging as there is a lack of valid and reliable tasks for testing speech sound discrimination in two-year-olds (Holt & Lalonde, 2012). Speech-sound discrimination is typically tested in infants using habituation or head-turn paradigms (Jusczyk & Aslin, 1995), both of which are inappropriate for preschool-aged children (Johnson & Zamuner, 2010; Werker, Polka, & Pegg, 1997). While older children and adults can be tested with more explicit tasks that use verbal responses or button presses, looking-based methods are better-suited for younger children and infants (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Werker, Cohen, Lloyd, Casasola, & Stager, 1998). Tasks appropriate for two- and three-year-olds are currently under development, but data suggest that they are less valid with two-year-olds, especially when children are tested on subtle sound contrasts (Holt & Lalonde, 2012). Given these challenges, we addressed this issue with data from adults. In Study 2, we tested adults on the same comprehension task on which children were tested in Study 1, and in addition, adults completed a discrimination task to assess whether they could perceive the difference between English and French disfluencies.

Study 2: adults

Study 2 set out to compare the processing of disfluencies in children to their processing by adults, who possess an expert language system. Given the lack of difference between monolinguals and bilinguals in Study 1, and the difficulty of recruiting monolingual participants in Montréal (where most adults interact regularly with both native French and native English speakers), we focused on bilingual adults. Adults were tested on the same comprehension task given to children in Study 1, as well as a discrimination task to rule out the possibility that the difference between the sounds *uh* and *eu*h was too subtle to be readily perceived.

Methods

As in Study 1, research in this study followed institutional and national guidelines for research with human participants. Adult participants read and signed a consent form.

Study 2 stimuli and data are available at <<https://osf.io/qn6px/>>.

Participants

Sixteen bilingual English–French adult participants (females = 15; 14 English-dominant, 2 French-dominant) were included in the study, and ranged in age from 19 to 43 years old, with one not reporting ($M_{\text{age}} = 23\text{y}$). All were students at

Concordia University, an English-speaking institution in Montréal, a city in which both English and French are used in the day-to-day lives of residents.

Participants filled out the Language Background Questionnaire (Segalowitz, 2009), which provided self-report data on participants' language background and proficiency. This was also used to determine participants' eligibility in the study. To be included in the study, adults' native language had to be either English, French, or both. Language criteria for inclusion also required that participants have scores of at least 4/5 on self-rated comprehension and production in both French and English, with 1/5 or less on any other languages spoken. Adult participants' English comprehension scores averaged 4.9/5, while English production scores averaged 4.9/5. French comprehension scores averaged 4.7/5, and French production scores averaged 4.3/5. For 12 adults, their dominant language was also the language that they acquired first. All adults reported having acquired their second language before the age of 6.5 years ($M_{\text{age}} = 3.8\text{y}$). An additional 6 adults were tested but not included in the final sample for having a native language other than French or English (4), for falling asleep during the study (1), or for difficulty calibrating the participant's eyes to the eye-tracker (1). Adults were recruited via a participant pool for Psychology students.

Comprehension and discrimination tasks

For the comprehension task, the equipment and stimuli used in the present study were identical to Study 1. Adults sat on a chair, and were instructed to attend to the screen. After the comprehension task, participants completed a discrimination task by listening to free-field auditory stimuli on a computer. Participants listened to individual *uh* and *eu* tokens that were excised from sentences in the comprehension task, and were asked to circle on a piece of paper whether they heard an English or a French disfluency. Participants had 8 s to listen to each individual disfluency and to circle their answer before the next sound was played, with no possibility of playing a sound twice. There were a total of 32 disfluencies played (16 English, 16 French) presented in a quasi-random order. The discrimination task was about 4 min in duration.

Procedure

Adults first filled out language background questionnaires to assess their language proficiency, then completed the comprehension task. The study began after the same calibration routine used with children. As with the bilingual children, adults then watched two experimental orders, one in French and one in English (counterbalanced). Next, adults completed the discrimination task, before being debriefed at the end of the study.

Results

Analytic strategy

For the eye-tracking task, the analytic strategy for the comprehension task followed that used in Study 1. For the discrimination task, we summed the number of trials on which participants correctly identified the language of the disfluency, which yielded a maximum score of 32 if adults identified all tokens correctly.

Disfluency Window

See [Figure 2](#) for an illustrated depiction of the proportion of looking averages. As only 1 male participated in the study, gender was not included in the analyses. Preliminary analyses including language of testing (dominant language, non-dominant language) and order of testing (dominant language first, non-dominant language first) revealed that these factors did not affect adults' looking (all p s > .05), and thus subsequent analyses were collapsed over these factors. A repeated-measures ANOVA was conducted with Trial type (Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent) on the proportion of looking to the novel object. This analysis revealed a main effect of Trial type, [$F(2,62) = 9.00, p < .001, \eta^2 = 0.225$]. As was observed in the children's results, there was no significant difference in looking between the Disfluent Language-consistent ($M = .30, SD = .26$) and Disfluent Language-inconsistent trials [$M = .30, SD = .26, t(31) = -0.064, p = .95, d = 0.04$]. However, participants looked at the novel object significantly less on Fluent trials ($M = .17, SD = .17$) than on Disfluent Language-consistent trials, [$t(31) = -3.2265, p = .003, d = 0.59$], or on Disfluent Language-inconsistent trials [$t(31) = -3.73, p = .001, d = 0.59$]. Results compared to chance are displayed in [Table 2](#). Surprisingly, comparisons against chance revealed a somewhat different pattern from the children, in that participants looked towards the novel object significantly *BELOW* chance across all three trial types: Disfluent Language-consistent trials, Disfluent Language-inconsistent trials, and Fluent trials. Recall that across all trials, the familiar object was labeled twice prior to the disfluency period, which might have biased adults to look towards the familiar object more than the novel object prior to hearing the final label. Despite this difference in baseline attention in adults as compared to children, adults' looking to the novel object increased on disfluent trials compared to on fluent trials, demonstrating their predictive use of disfluencies.

Label Window

See [Figure 3](#) for an illustrated depiction of the looking-to-target results of adults over the course of a trial. To ensure that adults were completing the main task, we examined adults' looking to the target once it was labeled. Across the three trial types, adults' looking at the target during the Label Window was significantly above chance: Fluent [$M = .85, SD = .07, t = 27.267, p < .001$], Disfluent Language-consistent [$M = .84, SD = .11, t = 17.266, p < .001$], and Disfluent Language-inconsistent [$M = .85, SD = .09, t = 21.647, p < .001$]. Thus, irrespective of hearing fluent versus disfluent speech, adults looked at the target after it was labeled.

Last, we also investigated whether looking to the target was more difficult for adults when words were novel compared to familiar. A 3 (Trial type: Fluent, Disfluent Language-consistent, Disfluent Language-inconsistent) \times 2 (Target type: Familiar, Novel) repeated-measures ANOVA revealed no main effect of Trial type [$F(2,59) = 0.19, p = .83, \eta^2 = 0.006$], and no interaction between Trial type and Target type, [$F(2,59) = 1.54, p = .22, \eta^2 = 0.052$]. However, there was a significant main effect of Target type, [$F(1,374) = 81.97, p < .001, \eta^2 = 0.219$], indicating that adults more accurately looked at familiar targets compared to novel targets. These results suggest that, although accuracy for both target types was very high, adults were more accurate at identifying the target word when it was familiar compared to novel, a finding reminiscent of results with bilingual children.

Discrimination task

On average, participants identified 97.7% of speech disfluencies correctly as belonging to either English or French, or 31.3 out of 32 (range = 24 to 32, median = 32). Thus, all participants could clearly tell apart English from French disfluencies, and obtained very high discrimination scores, well above what would be expected from chance (50% or 16 correct answers) [$t(15) = 30.25, p < .001$].

Discussion

In Study 2, we tested bilingual English–French adults on their predictive use of disfluencies. Like children, adults looked towards the novel object more when hearing disfluent compared to fluent speech, even when the disfluency was code-switched. This pattern was not due to a failure to discriminate different types of disfluencies, as adults could readily tell apart the English *uh* and the French *eu* when tested in a discrimination task. Combined with evidence that bilinguals are sensitive to a range of vowel contrasts at a far younger age than the 32-month-olds tested in Study 1 (e.g., Byers-Heinlein & Fennell, 2014), this suggests that bilinguals' similar performance in the two disfluent conditions was likely not due to a failure to detect the language mismatch. Given that adults detected but were not influenced by the specific realization of the disfluency (*uh* vs. *eu*), these results also fail to support our hypothesis that listeners associate particular disfluency realizations with novelty, instead highlighting listeners' flexibility across different realizations of speech disfluencies.

An intriguing finding from this study was that, after the object was labeled, adults performed better on familiar label trials than on novel label trials. In Study 1, this pattern was also seen with bilingual children, although not with monolingual children. This raises the possibility that differences in the use of mutual exclusivity-related behaviors such as disambiguation are affected by bilingualism into adulthood. However, the current study did not have a monolingual control group, and this possibility remains a hypothesis to be tested directly.

General discussion

We tested monolingual and bilingual children (aged 32 months) and bilingual adults on their predictive use of disfluencies that were either consistent (e.g., "Look at thee *uh* doll!") or inconsistent (e.g., "Look at thee *eu* doll!") with the language of the sentence. Across all groups, listeners increased attention to a novel object rather than a familiar object upon hearing a disfluency. This effect was not modulated by whether the language of the disfluency was consistent vs. inconsistent with the rest of the sentence, or by bilingual participants' language dominance. Results from a control study revealed that bilingual adults could readily discriminate between *uh* and *eu*, showing that the similar response across conditions was not due to an inability to differentiate the English and French disfluencies. Overall, these results suggest that speech disfluencies are a highly robust cue that allow listeners to predict that a speaker is more likely to refer to a novel, discourse-new object.

These studies were designed to test the hypothesis that listeners become sensitive to the predictive nature of disfluencies as they gain associative experience that disfluencies regularly predict the labeling of a novel or discourse-new object. We predicted that hearing typical speech disfluencies in one's ambient language would support listeners

to predict that the following word is likely to be novel. Results from our two studies failed to support this hypothesis. Specifically, we predicted that the predictive use of a disfluency would be affected by whether it was language-consistent or language-inconsistent, and that 2.5-year-old bilinguals might be less able to use disfluencies predictively than monolinguals. However, we did not find evidence that any of these factors affected listeners' ability to use disfluencies predictively.

Our results thus leave open the question of whether and how language experience contributes to listeners' ability to use speech disfluencies predictively. One possibility is that listeners do learn through experience that disfluencies predict novel or discourse-new referents, but that the representation of these disfluencies is inclusive of both native and non-native disfluencies. Under this account, rather than encoding a specific disfluency such as *uh* as being predictive, speakers might group together a range of speech hesitations, for example attending to the lengthening that is characteristic of disfluencies. Even though bilinguals' experience with specific disfluencies is divided across their two languages, they likely encounter a similar total number of disfluencies as monolinguals (or perhaps even more if they encounter more non-native speakers, or produce more disfluencies themselves). Moreover, to the degree that language-consistent and language-inconsistent disfluencies are both perceived as disfluencies, they may be sufficient to activate speakers' knowledge of the association between disfluencies and novel referents.

Another important consideration is that cues correlated with the filled pause might be sufficient to activate such knowledge. In our design, filled pauses were always preceded by a lengthened determiner consistent with the language of the sentence (e.g., *thee uh*, *thee euh*, *lee/laa euh*, *lee/laa uh*). This lengthening was a deliberate choice, given evidence that nearly all filled pauses are preceded by a lengthened determiner (with a vowel alteration in English; Fox Tree & Clark, 1997). Thus, our design cannot rule out the possibility that the lengthened determiner, either by itself or together with the filled pause, signaled that a novel referent would follow – a topic which will need to be investigated in future research. Our results nonetheless demonstrate that the presence of an atypical filled pause (e.g., Language-inconsistent) is not sufficient to disrupt listeners' predictive use of the disfluency.

A second very different possibility is that listeners become sensitive to the predictive nature of disfluencies when they understand them as a marker of a speaker's uncertainty. This framework would also predict that hearing any type of speech disfluency – whether native or non-native – helps listeners understand that a speaker who produces these disfluencies is uncertain, and will likely name a novel or discourse-new object as opposed to a familiar one. Here, children might understand that speech disfluencies signal a speaker's processing difficulty (Clark & Fox Tree, 2002), and infer that a novel referent is what caused the difficulty. Indeed, uncertain speakers tend to be disfluent and make more statements such as "I don't know" (Smith & Clark, 1993). Beginning around age two years, children are increasingly aware that a speaker's verbal cues and actions may indicate a lack of confidence, and this affects their learning from these speakers (Brosseau-Liard & Poulin-Dubois, 2014; Stock, Graham, & Chambers, 2009).

While we did not manipulate our speaker's knowledge state in this study, previous research suggests that speaker knowledge state matters for how listeners interpret disfluencies. For instance, adults' attention to novel objects is not modulated by disfluencies when they are told that a speaker suffers from object agnosia and has difficulty retrieving familiar words (Arnold *et al.*, 2007). Relatedly, 3.5-year-old

children use disfluencies to predict the labeling of a novel object when hearing a knowledgeable speaker, but not when hearing a forgetful speaker (Orena & White, 2015). Thus, both children and adults appear to modulate their predictive use of disfluencies based on characteristics and knowledge of the speaker.

Sensitivity to a speaker's knowledge state would likely be rooted in children's emerging theory of mind skills. Infants track others' perceptions, beliefs, and goals from early in life (Baillargeon, Scott, & He, 2010; Onishi & Baillargeon, 2005), and these skills become increasingly sophisticated over the preschool years (Astington & Edward, 2010). Previous work indicates that younger children, aged 16–20 months, are not able to use speech disfluencies predictively (Kidd *et al.*, 2011). This could indicate that between ages 16 and 32 months there is important development of the social and causal reasoning abilities underlying this ability. It will be important for future research to more explicitly investigate the link between such abilities and the predictive use of disfluencies to more clearly understand how it develops.

Intriguingly for this study, previous research has suggested that bilinguals have accelerated theory of mind development relative to monolinguals (Goetz, 2003; Kovács, 2009). If children's predictive use of disfluencies is rooted in an understanding of a speaker's uncertainty, we might have expected bilinguals to be even more sensitive than monolinguals to the predictive nature of disfluencies, given they may have earlier insight into social signals (Yow & Markman, 2011). However, we found no difference between these two groups. One explanation is that both monolinguals and bilinguals aged 32 months have developed the necessary social reasoning capacity to succeed at our task. It would be important for future studies to test younger bilinguals at an age at which monolinguals previously performed more poorly, around 16–24 months, to determine whether they might show an advantage at this age.


Our study provides an important contribution to the literature on children's predictive use of disfluencies, which had previously been limited to studies of monolingual English-learning children (Kidd *et al.*, 2011; Orena & White, 2015; Owens & Graham, 2016; Owens *et al.*, 2018; Thacker, Chambers, & Graham, 2018). We replicated these findings with English monolinguals, and extended them to French monolinguals and French–English bilinguals. Moreover, we showed that these groups can flexibly use both native, non-native, and code-switched disfluencies to make predictions. Our results suggest that children from various language backgrounds are sensitive to language-general characteristics of disfluencies, such as lengthening of the determiner or the filled pause itself. However, an important caveat is that English and French disfluencies (*uh* vs. *euh*) are somewhat acoustically similar. Filled pauses can sound quite different in other languages: for example, Japanese uses the bisyllabic disfluency *eeto* (Swerts, 1998). An important direction for future research will be to examine whether Japanese learners can use this disfluency predictively from the same age as children in our study, and whether such a disfluency could be used by speakers of English or French, for example. It is possible that speakers of English and French may require the lengthening of a single vowel as a key marker and indicator of a hesitation, and thus may not be able to use a bisyllabic disfluency such as *eeto* predictively. Additional cross-language research could better pinpoint what type of knowledge is being acquired across development, and how universal children's sensitivities are.

Importantly, our results show that both toddler and adult listeners, from monolingual and bilingual backgrounds, can flexibly take advantage of speech

disfluencies to make predictions about upcoming speech. Moreover, this occurs in real time; listeners increase their attention to a novel object while hearing the disfluency, prior to hearing the noun. Prediction during language comprehension is an important component of efficient language processing, especially for young learners (Lew-Williams & Fernald, 2007; Rabagliati, Gambi, & Pickering, 2016). Since disfluencies occur frequently in human language (Shriberg, 2001), the ability to use them for prediction could make an important contribution to language acquisition. Our evidence shows that children are flexible across different realizations of disfluencies, even those that are not typical of their everyday language input. This suggests that children and adults can interpret disfluencies from interlocutors across a wide variety of backgrounds, including non-native speakers who may produce atypical disfluencies.

Converging with previous research (Arnold *et al.*, 2004; Kidd *et al.*, 2011), our findings confirm that disfluencies direct listeners' attention to novel objects. Our results also showed that, upon hearing the novel label, attention to the novel object was maintained relative to when a familiar label was heard. However, this finding was somewhat less robust for bilingual children, replicating previous findings (Byers-Heinlein & Werker, 2009, 2013; Kalashnikova, Mattock, & Monaghan, 2015). What remains untested is how the presence of disfluencies contributes to subsequent association of the novel object with a novel label. On the one hand, disfluencies direct attention to novel or discourse-new objects, which could help children associate the novel objects with an upcoming novel word. On the other hand, children may avoid learning words that follow a disfluency if they are less willing to learn from speakers who have been unreliable or verbally inaccurate (Brooker & Poulin-Dubois, 2013; Kim, Kalish, & Harris, 2012). Studies investigating word learning processes have found cases where children can identify the referent of a novel word, but have difficulty forming a word-object association (Horst & Samuelson, 2008). If disfluencies paradoxically increase attention to a novel object while decreasing children's willingness to associate that object with a novel label, this provides additional evidence for a processing of disfluencies that is based on children's understanding of a speaker's knowledge. We are addressing this possibility in an ongoing study with monolingual and bilingual children.

Given the growing literature on bilingual children's processing of speech, the investigation of speech disfluencies is a timely topic that illuminates the mechanisms allowing listeners to comprehend everyday speech, and demonstrate listeners' flexibility in the face of phonetic variation. Based on the findings in this study, caregivers and educators should be made aware that some natural imperfections in speech may be beneficial during comprehension. Although our *uhs* and *euhs* are often deemed peripheral to communication, our findings add to evidence that the presence of disfluencies in natural speech is useful for children of different language backgrounds to make predictions about upcoming words during listening. Without hesitation, we contend that speech disfluencies could be an important element of children's language development.

Author ORCIDs.  Elizabeth MORIN-LESSARD, 0000-0001-7218-1611, Krista BYERS-HEINLEIN, 0000-0002-7040-2510

Acknowledgements. We thank our colleagues Drs Katherine White and Celeste Kidd for insightful design discussions. Many thanks to the members of the Concordia Infant Research Lab for their comments on an earlier version of this manuscript, with particular thanks to Kay Goly for her assistance with stimuli

preparation and data collection. Finally, we thank all the families and adult participants who took part in this study. This research was supported by a FRQSC scholarship to EML and an NSERC Discovery Grant to KBH. KBH holds the Concordia University Research Chair in Bilingualism.

References

- Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal thee, um, new information. *Journal of Psycholinguistic Research*, 32(1), 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(5), 914–30.
- Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55.
- Arnold, J. E., & Tanenhaus, M. K. (2007). Disfluency effects in comprehension: how new information can become accessible. In E. Gibson & N. Perlmutter (Eds.), *The processing and acquisition of reference* (pp. 197–217) Cambridge, MA: MIT Press.
- Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The old and thee, uh, new disfluency and reference resolution. *Psychological Science*, 15(9), 578–82.
- Astington, J. W., & Edward, M. J. (2010). The development of theory of mind in early childhood. *Social Cognition in Infancy*, 5, 16–20.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14(3), 110–18.
- Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes*, 25(4), 441–55.
- Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, 22(3), 201–11.
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: effects of age, relationship, topic, role, and gender. *Language and Speech*, 44(2), 123–47.
- Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, 2(1), 29–49.
- Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native ‘um’ elicit prediction of low-frequency referents, but non-native ‘um’s do not. *Journal of Memory and Language*, 75, 104–16.
- Boudreault, M. C., Cabirol, É. A., Trudeau, N., Poulin-Dubois, D., & Sutton, A. (2007). Les inventaires MacArthur du développement de la communication: validité et données normatives préliminaires. *Revue canadienne d'orthophonie et d'audiologie*, 31(1), 27–37.
- Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? The effect of speaker labeling accuracy on infants’ word learning, imitation, and helping behaviors. *Infancy*, 18(s1), E46–E68.
- Brousseau-Liard, P. E., & Poulin-Dubois, D. (2014). Sensitivity to confidence cues increases during the second year of life. *Infancy*, 19(5), 461–75.
- Byers-Heinlein, K., Chen, K. H., & Xu, F. (2014). Surmounting the Tower of Babel: monolingual and bilingual 2-year-olds’ understanding of the nature of foreign language words. *Journal of Experimental Child Psychology*, 119, 87–100.
- Byers-Heinlein, K., & Fennell, C. T. (2014). Perceptual narrowing in the context of increased variation: insights from bilingual infants. *Developmental Psychobiology*, 56(2), 274–91.
- Byers-Heinlein, K., Schott, E., Gonzalez-Barrero, A. M., Brouillard, M., Dubé, D., Laoun-Rubenstein, A., Morin-Lessard, E., Mastroberardino, M., Jardak, A., Pour Iliaci, S., Salama-Siroishka, N., & Tamayo, M. P. (2018). MAPLE: a Multilingual Approach to Parent Language Estimates. *Bilingualism: Language and Cognition*. Online 10.31234/osf.io/r2q3u
- Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: infants’ language experience influences the development of a word-learning heuristic. *Developmental Science*, 12(5), 815–23.
- Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: evidence from bilingual infants. *Cognition*, 128(1), 407–16.
- CIHR, NSERC, & SSHRC (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada)

- (2014). *Tri-council policy statement: ethical conduct for research involving humans*. Retrieved from <<http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/>> (last accessed 16 May 2017).
- Clark, H. H., & Fox Tree, E. F.** (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111.
- Clark, H. H., & Wasow, T.** (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3), 201–42.
- Dink, J. W., & Ferguson, B.** (2015). eyetrackingR: an R library for eye-tracking data analysis. Retrieved from <<http://www.eyetrackingr.com>>.
- Duez, D.** (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, 25(1), 11–28.
- Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J.** (2007). *MacArthur-Bates Communicative Development Inventories*. Baltimore, MD: Paul H. Brookes Publishing Company.
- Fox Tree, J. E. F.** (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34(6), 709–38.
- Fox Tree, J. E. F.** (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, 29(2), 320–26.
- Fox Tree, J. E. F., & Clark, H. H.** (1997). Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition*, 62(2), 151–67.
- Frank, I., & Poulin-Dubois, D.** (2002). Young monolingual and bilingual children's responses to violation of the Mutual Exclusivity Principle. *International Journal of Bilingualism*, 6(2), 125–46.
- Friedrich, M., & Friederici, A. D.** (2005). Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children. *Neuroreport*, 16(16), 1801–4.
- Goetz, P. J.** (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, 6(1), 1–15.
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L.** (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(1), 23–45.
- Gómez, R. L., & Gerken, L.** (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4(5), 178–86.
- Heller, D., Arnold, J. E., Klein, N., & Tanenhaus, M. K.** (2015). Inferring difficulty: flexibility in the real-time processing of disfluency. *Language and Speech*, 58(2), 190–203.
- Holt, R. F., & Lalonde, K.** (2012). Assessing toddlers' speech-sound discrimination. *International Journal of Pediatric Otorhinolaryngology*, 76(5), 680–92.
- Horst, J. S., & Hout, M. C.** (2016). The Novel Object and Unusual Name (NOUN) database: a collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–409.
- Horst, J. S., & Samuelson, L. K.** (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128–57.
- Houston-Price, C., Caloghiris, Z., & Raviglione, E.** (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, 15(2), 125–50.
- Johnson, E. K., & Zamuner, T.** (2010). Using infant and toddler testing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 73–94). Amsterdam: John Benjamins.
- Jusczyk, P. W., & Aslin, R. N.** (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29(1), 1–23.
- Kalashnikova, M., Mattock, K., & Monaghan, P.** (2015). The effects of linguistic experience on the flexible use of mutual exclusivity in word learning. *Bilingualism: Language and Cognition*, 18(4), 626–38.
- Kidd, C., White, K. S., & Aslin, R. N.** (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science*, 14(4), 925–34.
- Kim, S., Kalish, C. W., & Harris, P. L.** (2012). Speaker reliability guides children's inductive inferences about novel properties. *Cognitive Development*, 27(2), 114–25.
- Kovács, Á. M.** (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12(1), 48–54.
- Lew-Williams, C., & Fernald, A.** (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–8.
- Markman, E. M., & Wachtel, G. F.** (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20(2), 121–57.

- Onishi, K. H., & Baillargeon, R.** (2005). Do 15-month-old infants understand false beliefs? *Science*, 308 (5719), 255–8.
- Orena, A. J., & White, K. S.** (2015). I forget what that's called! Children's online processing of disfluencies depends on speaker knowledge. *Child Development*, 86(6), 1701–9.
- Owens, S. J., & Graham, S. A.** (2016). Thee, uhh disfluency effect in preschoolers: a cue to discourse status. *British Journal of Developmental Psychology*, 34(3), 388–401.
- Owens, S. J., Thacker, J. M., & Graham, S. A.** (2018). Disfluencies signal reference to novel objects for adults but not children. *Journal of Child Language*, 45(3), 581–609.
- Poplack, S.** (1980). Sometimes I'll start a sentence in Spanish y termino en Español: toward a typology of code-switching. *Linguistics*, 18(7/8), 581–618.
- R Core Team** (2017) R: a language and environment for statistical computing. Retrieved from: <<https://www.R-project.org/>>.
- Rabagliati, H., Gambi, C., & Pickering, M. J.** (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience*, 31(1), 94–105.
- Saffran, J. R., Newport, E. L., & Aslin, R. N.** (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, 35(4), 606–21.
- Segalowitz, N.** (2009). *The Language Background Questionnaire*. Unpublished instrument, Department of Psychology, Concordia University, Montreal, QC, Canada.
- Shriberg, E.** (2001). To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31(1), 153–69.
- Smith, V. L., & Clark, H. H.** (1993). On the course of answering questions. *Journal of Memory and Language*, 32(1), 25–38.
- Snedeker, J., & Trueswell, J. C.** (2004). The developing constraints on parsing decisions: the role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49 (3), 238–99.
- Stock, H. R., Graham, S. A., & Chambers, C. G.** (2009). Generic language and speaker confidence guide preschoolers' inferences about novel animate kinds. *Developmental Psychology*, 45(3), 884–8.
- Swerts, M.** (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30(4), 485–96.
- Thacker, J. M., Chambers, C. G., & Graham, S. A.** (2018). When it is apt to adapt: flexible reasoning guides children's use of talker identity and disfluency cues. *Journal of Experimental Child Psychology*, 167, 314–27.
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L.** (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, 34(6), 1289–309.
- Werker, J. F., Polka, L., & Pegg, J. E.** (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and Parenting*, 6(34), 171–8.
- Yow, W. Q., & Markman, E. M.** (2011). Bilingualism and children's use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition*, 14(4), 562–9.

Cite this article: Morin-Lessard E, Byers-Heinlein K (2019). *Uh* and *euh* signal novelty for monolinguals and bilinguals: evidence from children and adults. *Journal of Child Language* 46, 522–545. <https://doi.org/10.1017/S0305000918000612>