Journal of Memory and Language 75 (2014) 104–116

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Native ‘um’s elicit prediction of low-frequency referents, but non-native ‘um’s do not

Hans Rutger Bosker [a],*, Hugo Quené [b], Ted Sanders [b], Nivja H. de Jong [b]

[a] *Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands*
[b] *Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands*

ARTICLE INFO

Article history:
Received 25 July 2013
revision received 16 May 2014

Keywords:
Prediction
Disfluency bias
Disfluency
Hesitation
Non-native speech
Speech comprehension

ABSTRACT

Speech comprehension involves extensive use of prediction. Linguistic prediction may be guided by the semantics or syntax, but also by the performance characteristics of the speech signal, such as disfluency. Previous studies have shown that listeners, when presented with the filler *uh*, exhibit a *disfluency bias* for discourse-new or unknown referents, drawing inferences about the source of the disfluency. The goal of the present study is to study the contrast between native and non-native disfluencies in speech comprehension. Experiment 1 presented listeners with pictures of high-frequency (e.g., a hand) and low-frequency objects (e.g., a sewing machine) and with fluent and disfluent instructions. Listeners were found to anticipate reference to low-frequency objects when encountering disfluency, thus attributing disfluency to speaker trouble in lexical retrieval. Experiment 2 showed that, when participants listened to disfluent non-native speech, no anticipation of low-frequency referents was observed. We conclude that listeners can adapt their predictive strategies to the (non-native) speaker at hand, extending our understanding of the role of speaker identity in speech comprehension.

Prediction in human communication lies at the core of language production and comprehension (see Kutas, DeLong, & Smith, 2011; Pickering & Garrod, 2007, for reviews). Most research into language-mediated prediction has focused on prediction elicited by semantic (e.g., Altmann & Kamide, 1999), syntactic (e.g., Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004) or phonological properties (e.g., DeLong, Urbach, & Kutas, 2005) of the linguistic input. But listeners form linguistic predictions not only based on *what* is said, but also on *how* it is said. That is, performance aspects of the speech signal also affect prediction, such as prosodic characteristics (Dahan, Tanenhaus, & Chambers, 2002; Weber, Grice, & Crocker, 2006) and disfluencies (Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Arnold, Hudson Kam, & Tanenhaus, 2007). This study corroborates that disfluency may indeed guide prediction by showing that listeners—upon encountering an *uh*—anticipate reference to a low-frequency referent, but only when listening to a native speaker.

Disfluencies are “phenomena that interrupt the flow of speech and do not add propositional content to an utterance” (Fox Tree, 1995, p. 709), such as silent pauses, filled pauses (e.g., *uh*’s and *uhm*’s), corrections, and repetitions. Disfluency is a common feature of spontaneous speech: it is estimated that six in every hundred words are affected by disfluency (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Fox Tree, 1995). Disfluencies follow a non-arbitrary distribution: they tend to occur before dispreferred or more complex content, such as open-class words (Maclay & Osgood, 1959), unpredictable lexical items (Beattie &

---

* Corresponding author.
E-mail address: HansRutger.Bosker@mpi.nl (H.R. Bosker).

http://dx.doi.org/10.1016/j.jml.2014.05.004
0749-596X/

Butterworth, 1979), color names of low-frequency (Levelt, 1983), or names of low-codability images (Hartsuiker & Notebaert, 2010).

Traditionally, the disfluent character of spontaneous speech was thought to disrupt the mechanisms involved in speech perception (Martin & Strange, 1968). It was assumed to pose a continuation problem for listeners (Levelt, 1989), who were thought to be required to edit out disfluencies in order to process the remaining linguistic input. Thus, disfluencies would uniformly present obstacles to comprehension and need to be excluded in order to study speech comprehension in its 'purest' form (cf. Brennan & Schober, 2001).

Experimental evidence has shown, however, that disfluencies may help rather than hinder the listener. Disfluencies may aid comprehenders to avoid erroneous syntactic parsing (Brennan & Schober, 2001; Fox Tree, 2001), to attenuate context-driven expectations about upcoming words (Corley, MacGregor, & Donaldson, 2007; MacGregor, Corley, & Donaldson, 2010), to speed up word recognition (Corley & Hartsuiker, 2011), and to improve recognition memory (Collard, Corley, MacGregor, & Donaldson, 2008; Corley et al., 2007; Fraundorf & Watson, 2011; MacGregor et al., 2010). Moreover, they may guide prediction of the following linguistic content.

Arnold, Fagnano, and Tanenhaus (2003) and Arnold et al. (2004) investigated whether listeners use the increased likelihood of speakers to be disfluent (e.g., saying 'thee *uh* candle' instead of 'the candle') when speakers refer to new as compared to given information (Arnold, Wasow, Losongco, & Ginstrom, 2000). In eye-tracking experiments using the Visual World Paradigm, participants' eye fixations revealed that, prior to target onset, listeners were biased to look at a discourse-new referent when presented with a disfluent utterance: a *disfluency bias* toward discourse-new referents (Arnold et al., 2003, 2004). Subsequently, Arnold et al. (2007) extended the disfluency bias to the reference resolution of known vs. unknown objects (cf. Watanabe, Hirose, Den, & Minematsu, 2008). Upon presentation of a disfluent sentence such as 'Click on thee *uh* red [target]', listeners were found to look more at an unknown object (an unidentifiable abstract symbol) prior to target onset as compared to a known object (e.g., an ice-cream cone).

Additional experiments in Arnold et al. (2007) and Barr and Seyfeddinipur (2010) demonstrated that the mechanism responsible for the disfluency bias is a perspective-taking process. In the second experiment reported in Arnold et al. (2007), the authors tested whether (1) listeners 'simply' associated unknown or discourse-new referents with disfluency, or that (2) listeners actively made rapid inferences about the source of the disfluency (e.g., when the speaker is perceived to have trouble in speech production, the most probable source of difficulty is the unfamiliarity of the unknown referent). This second experiment was identical to their first experiment, except that participants were now told that the speaker suffered from object agnosia (a condition involving difficulty recognizing simple objects). Results revealed that the preference for unknown referents following a disfluency, observed in the first experiment, disappeared in the second experiment. This suggests that listeners draw inferences about the speaker's cognitive state (e.g., having equal difficulty naming known and unknown objects) which modulates the extent to which disfluency guides prediction.

This raises the question how disfluency affects prediction in a much more common situation, namely when listeners are confronted with disfluencies in non-native speech. Non-native speech is all the more vulnerable to disfluency due to, for instance, incomplete mastery of the second language (L2) or a lack of automaticity in L2 speech production (De Bot, 1992; Segalowitz, 2010). These factors lead to a higher incidence of disfluencies in non-native speech, and it causes a different distribution of disfluencies (relative to the regularities in native disfluency production; Davies, 2003; Kahng, 2013; Riazantseva, 2001; Skehan & Foster, 2007; Skehan, 2009; Tavakoli, 2011). As a consequence, the distribution of disfluencies in non-native speech may be argued to be, from the native listeners point of view, more variable than the disfluency distribution in native speech.

This different distribution of non-native disfluencies might affect listeners' predictive strategies in two possible ways: first, the disfluency bias may be attenuated when listening to non-native speech—similar to what was found for speech from an object-agnosic patient (Arnold et al., 2007). Because of their higher incidence and wider distribution, non-native disfluencies are, to the listener, worse predictors of the word to follow (as compared to native disfluencies). Thus, native listeners may refrain from using non-native disfluencies for prediction, leading to a reduction or elimination of the preference for more complex referents upon hearing non-native disfluent speech.

This first hypothesis is supported by the observed attenuation of comprehension processes when listening to non-native speech. For instance, Hanulíková, Van Alphen, Van Goch, and Weber (2012) report a classical P600 effect for grammatical gender violations in native speech. In contrast, when the same violations were produced by a non-native speaker with a foreign accent, no P600 effect was observed. The foreign accent in the non-native speech presumably served as a cue for listeners to adjust their comprehension strategies of grammatically ill-formed sentences. Hanulíková et al. (2012) argue that prior experience with non-native speakers producing syntactic errors lies at the core of this cognitive modulation. Similarly, prior experience with the different distribution of non-native disfluencies may attenuate listeners' predictive strategies.

Alternatively, the disfluency bias for dispreferred or more complex referents may be enhanced when listening to non-native speech. Naming more complex objects is (even) more cognitively demanding for a non-native speaker than it is for a native speaker. As a consequence, the likelihood of a disfluency preceding a more complex word may be argued to be higher in non-native speech than in native speech. If native listeners take this into account, their anticipation of more complex information following non-native disfluency may be enhanced (relative to native speech). As such, the disfluency bias may have an even stronger presence in the comprehension of non-native speech.

This second hypothesis is supported by studies of lexical retrieval of monolinguals and bilinguals. For instance, the ‘weaker links hypothesis’, as proposed by Gollan, Montoya, Cera, and Sandoval (2008), argues that bilingualism has an indirect effect on lexical retrieval. Because non-native speakers have limited exposure to the words in their non-native language, the frequency contrast between high-frequency and low-frequency words is larger for a non-native speaker than for a native (cf. Duyck, Vanderelst, Desmet, & Hartsuiker, 2008). If listeners are aware of this difference, they may anticipate low-frequency referents to occur more often following non-native disfluencies relative to native disfluencies.

This study reports two eye-tracking experiments which target listeners’ attributions of disfluency to the speaker having trouble in lexical retrieval, by studying the reference resolution of high-frequency (e.g., a hand) vs. low-frequency (e.g., a sewing machine) lexical items. Frequency of occurrence is known to affect lexical retrieval (Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Caramazza, 1997; Jescheniak & Levelt, 1994; Levelt, Roelofs, & Meyer, 1999), and it has been identified as a factor affecting the distribution of disfluencies (Hartsuiker & Notebaert, 2010; Kircher, Brammer, Levelt, Bartels, & McGuire, 2004; Levelt, 1983; Schnadt & Corley, 2006). For Experiment 1, using native speech materials, we hypothesize that if we present listeners with two known objects—one having a high-frequency and the other having a low-frequency name—we may find a disfluency bias towards low-frequency objects. Experiment 2, using non-native speech materials, allows for a comparison between the processing of native and non-native disfluencies.

Method

Experiment 1

Participants. Forty-four native Dutch participants, recruited from the UiL OTS participant pool, were paid for participation. All participated with implicit informed consent in accordance with local and national guidelines, and reported to have normal hearing and normal or corrected-to-normal eye-sight. Data from 3 participants were lost due to technical problems. Data from 6 other participants were excluded from further analyses because their responses on a post-experimental questionnaire indicated suspicion about the experiment (see below). The mean age of the remaining 35 participants was 23.8 years ($SD = 8.4$; 11m/24f).

Design and materials. The design of the current experiments resembles that of Arnold et al. (2007). However, in the Visual World Paradigm (cf. Huettig, Rommers, & Meyer, 2011; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) used by Arnold et al. (2007), participants viewed visual arrays on a screen consisting of four pictures: a known object in color A, the same known object in color B, an unknown object in color A and the same unknown object in color B. The spoken instruction contained a color adjective preceding the target word (e.g., ‘Click on thee *uh* red [target]’), disambiguating between target and competitor in color A and distractors in color B. In Arnold et al. (2007) only looks to the known and unknown objects that matched the color word in the input were analyzed (i.e., the target and competitor), based on the argument that “if disfluency creates a bias toward unfamiliar objects, we should see more looks to the unfamiliar color-matched object than the familiar color-matched object in the disfluent condition but not the fluent condition” (p. 919). Because Arnold et al. (2007) only performed statistical analyses on looks to the two relevant objects, and because there is for the present study no principal reason to include two additional distractors in the visual arrays, our experiments use visual arrays consisting of only two objects (one low-frequency, one high-frequency).

A set of low-frequency (LF; $N = 30$) and a set of high-frequency (HF; $N = 30$) pictures (see Appendix) was selected from the picture set from Severens, Lommel, Ratinckx, and Hartsuiker (2005) on the basis of log frequencies: mean (SD) log frequency LF = 0.38 (0.28); HF = 2.07 (0.29). All pictures selected the Dutch common article *de* (as opposed to the neuter article *het*) and had high name agreement: mean (SD) name agreement LF = 96.7 (3.64); HF = 97.3 (3.49). The stimuli also differed on dimensions that are strongly correlated with frequency of occurrence (see Appendix), such as word length in syllables [LF = 2.43 (0.86); HF = 1.30 (0.60)], and age of acquisition [AoA; LF = 6.28 (1.14); HF = 4.70 (0.86)]. We decided to use these picture sets because (i) within this picture database, it was not viable to select different picture sets with low intercollinearity of frequency, AoA, and word length and (ii) word length and AoA are also known to affect lexical retrieval (Belke, Brysbaert, Meyer, & Ghyselinck, 2005; Meyer, Roelofs, & Levelt, 2003; Morrison & Ellis, 2000; Severens et al., 2005). Therefore, if we would find a disfluency bias for the LF picture set, we may infer that listeners attribute the disfluency to speaker trouble in lexical retrieval.

LF pictures were paired with HF pictures to form a visual array of pictures for one trial. There was no phonological overlap between pair members. Together with these experimental pictures, an equal number of LF and HF filler pictures was selected following the same criteria as used for the experimental pictures. However, half of the filler target objects selected the neuter article *het* in order to maintain a natural ratio of grammatical genders.

The audio materials consisted of instructions to click on one of the two objects. These instructions were either fluent or disfluent. A corpus study, based on the Corpus of Spoken Dutch (CGN; Oostdijk, 2000), was conducted to decide on the position of the disfluency in our disfluent sentences. The study targeted the position of the Dutch filled pauses *uh* and *uhm*. It was found that the most common position of Dutch filled pauses was the position preceding the article *de* ($N = 4111$; as compared to the position following the article: $N = 754$). Therefore, the disfluency in our disfluent condition always preceded the article (cf. Arnold et al., 2007, where the disfluency followed the article).

For the speech materials of Experiment 1, a female native Dutch speaker (age = 30) was recorded. Recordings were made in a sound-attenuated booth using a Sennheiser ME-64 microphone. The speaker was instructed to produce half of the target words (50% HF, 50% LF) in the fluent template (i.e., *Klik op de [target]*, 'Click on the [target]'), and the other half of the target words using a disfluent template, produced 'as naturally as possible' (i.e., *Klik op uh de [target]*, 'Click on uh the [target]'). From all sentences that were recorded, three fluent and three disfluent sentence templates were excised that sounded most natural. These templates extended from the onset of *Klik* to the onset of the article *de* (boundaries set at positive-going zero-crossings, using Praat; Boersma & Weenink, 2012). Additionally, the target words were excised from the same materials, starting at the onset of the article *de* at a positive-going zero-crossing. These target fragments were randomly assigned to one of the three fluent sentence templates (resulting in a fluent instruction), and to one of the three disfluent sentence template (resulting in a disfluent instruction). Thus, target words were identical across fluent and disfluent conditions.

As a consequence of the described cross-splicing procedure, the differences between fluent and disfluent stimuli were located in the sentence templates (i.e., fluent *Klik op*, 'Click on'; and disfluent *Klik op uh*, 'Click on uh'). In order to have the recorded instructions sound natural, the contrast between disfluent and fluent stimuli also involved several prosodic characteristics (cf. Arnold et al., 2007). For instance, the words *Klik op*, 'Click on', in the disfluent condition were longer and had a higher pitch as compared to the fluent condition (see Table 1 for prosodic properties of the native, Experiment 1, and non-native, Experiment 2, sentence templates). Nonetheless, the primary difference between the fluent and disfluent condition concerned the absence or presence of the filled pause *uh*.

Filler trials were recorded in their entirety; no cross-splicing was applied to these sentences. Instead of counterbalancing the two fluency conditions across the LF and HF filler targets, each LF filler target was recorded in the disfluent condition and each HF filler target was recorded in the fluent condition. The reason for this design was that we aimed at a fluent:disfluent ratio across the two frequency conditions which resembled the ratio in spontaneous speech (with disfluencies occurring more often before low-frequency words; Hartsuiker & Notebaert, 2010; Kircher et al., 2004; Levelt, 1983; Schnadt & Corley, 2006). Using our design, the fluent:disfluent ratio was 1:3 for low-frequency targets and 3:1 for high-frequency targets. Disfluent filler trials contained all sorts of disfluencies (filled pauses in different positions, vowel lengthening, corrections, repetitions, etc.).

Table 1
Duration (in ms) and pitch (in Hz) for the three fluent and three disfluent sentence templates in the native and non-native speech.

	Duration	Maximum pitch
Native speech		
Fluent		
Klik	199, 194, 215	227, 217, 220
op	180, 166, 146	237, 222, 214
Disfluent		
Klik	212, 263, 218	262, 282, 261
op	265, 244, 283	269, 270, 260
uh	889, 933, 871	244, 263, 246
Non-native speech		
Fluent		
Klik	222, 222, 212	237, 228, 225
op	191, 197, 196	255, 227, 230
Disfluent		
Klik	221, 262, 240	273, 278, 287
op	263, 233, 253	282, 287, 304
uh	891, 897, 950	254, 259, 280

Apparatus and Procedure. Prior to the actual eye-tracking experiment, participants were told a cover story that the purpose of the eye-tracking experiment was to test the extent to which listeners could correctly follow up instructions from all sorts of speakers (following Arnold et al., 2007). Purportedly, recordings had been made of 20 speakers, including both native and non-native speakers of Dutch, who had been presented with pictures just like the ones the participant was about to see. The speaker's task was then to name a pre-selected picture using a standard instruction template, namely *Klik op de [object]*, 'Click on the [object]'. The presence of the cover story was motivated by the need to justify the presence of disfluencies in the speech. Moreover, it meant that listeners might plausibly attribute the disfluency to difficulty in word retrieval. Participants in Experiment 1 were told they would be listening to speech from a native speaker.

In the eye-tracking experiment, eye movements were recorded with a desktop-mounted SR Research EyeLink 1000 eye-tracker, controlled by ZEP software (Veenker, 2012), sampling the right eye at 500 Hz. Visual materials were presented on a 19-in. computer screen (within a sound-attenuated eye-tracking booth) at a viewing distance of approximately 60 cm. Participants used a standard computer mouse. Speech was heard through loudspeakers at a comfortable listening volume. After participants were informed about the procedure, each experiment started with a thirteen-point calibration procedure followed by a validation procedure. After this validation, participants performed eight practice trials. A drift correction event occurred before every trial (a red dot appearing in the center of the screen). When the participants had fixated the dot, the two visual stimuli were presented. The onset of the visual stimuli preceded the onset of the audio instructions by 1500 ms. The position of LF and HF pictures on the screen (left or right) was randomized and a Latin Square design was used to counterbalance frequency and disfluency in the experimental trials.

Following the eye-tracking experiment, participants were presented with a post-experimental questionnaire that collected ratings of (1) the naturalness of the speech, (2) the accentedness of the speech, (3) the fluency of the speech, and (4) the experience participants had with listening to non-native speakers of Dutch. Following Barr (2008b), participants rated their level of agreement with four statements on a scale from 1 to 9 (1 = strong disagreement; 9 = strong agreement). If a participant's naturalness rating was lower than 5, it was taken as evidence of suspicion towards the stimuli ($n = 6$, see above) and the data from these participants were excluded from any further analyses. However, inclusion of these data did not result in different interpretations of our results.

Experiment 2

Experiment 2 was identical to Experiment 1 except for the fact that now non-native speech materials were used.

Participants. Forty-two native Dutch participants, recruited from the UiL OTS participant pool, were paid for participation. All participated with implicit informed consent in accordance with local and national guidelines, and reported to have normal hearing and normal or corrected-to-normal eye-sight. Data from 6 participants were excluded because their responses on a post-experimental questionnaire indicated suspicion about the experiment (having provided naturalness ratings below 5 in the post-experimental questionnaire). The mean age of the remaining 36 participants was 22.6 years ($SD = 3.3$), 5m/31f.

Design and materials. The visual stimuli were identical to those used in Experiment 1. For the speech materials of Experiment 2, a non-native speaker of Dutch was recorded (female, L1 Romanian, age = 25, LoR = 3.5 years). She reported having rudimentary knowledge of Dutch (self-reported CEFR level A1/A2) and very limited experience using Dutch in daily life. In order to minimize the contrast between the native and non-native recordings, the recording procedures of Hanulíková et al. (2012) were adopted: the non-native speaker first listened to a native utterance after which she imitated the native speech, sentence by sentence. This resulted in non-native speech recordings that were similar to the native recordings except for a noticeable foreign accent (see Table 1 for prosodic properties of the native and non-native speech stimuli). The remaining procedure was identical to Experiment 1.

Apparatus and procedure. The cover story, the instructions, and the post-experimental questionnaire were identical to Experiment 1, except that participants in Experiment 2 were instructed that they were going to listen to a non-native speaker of Dutch.

Results

Data from both experiments were combined in all analyses. The reported results follow the order of the experimental sessions: first the eye-tracking data are introduced, followed by the mouse click data, and finally the post-experimental questionnaire.

Eye fixations

Prior to the analyses, blinks and saccades were excluded from the data. Eye fixations from trials with a false mouse response were excluded from analyses (<1%). Only fixations on the pictures themselves were coded as a look toward that particular picture.

Mixed effects logistic regression models (GLMMs: Generalized Linear Mixed Models; Quené & Van den Bergh (2008)) as implemented in the `lme4` library (Bates, Maechler, & Bolker, 2012) in R (R Development Core Team, 2012) evaluated participants' eye fixations. The eye fixation data were evaluated in two time windows: one *pre-target time window* preceding article onset and one *post-target time window* following article onset. However, because the post-target time window is not informative of anticipation of linguistic content before it is mentioned, the analyses of the post-target time window are provided in Section Supplementary Materials.

Pre-target time window. The analysis of the data in the pre-target time window tested whether disfluency elicited prediction of low-frequency referents. The time window of interest was defined as starting from sentence onset and ending before article onset (i.e., all fixations while hearing *Klik op* and *Klik op uh*). In the pre-target time window no phonetic information about the target itself was available to the listener. Therefore, we did not analyze participants' looks to target, as is common in analyses of the Visual World Paradigm. Instead, we analyzed participants' looks to the low-frequency object, coded binomially, with intercepts which varied randomly by participants, items, and sentence templates.

If disfluencies guide prediction, then we predict that the probability of participants looking at low-frequency objects — prior to target onset — would be higher in the disfluent condition than in the fluent condition. Note that our Visual World data do not allow for distinguishing a preference for low-frequency referents from a dispreference for high-frequency referents. In the remainder of this paper, we interpret a higher probability of fixations on low-frequency objects as a preference for low-frequency referents, without excluding the possibility that the same result may be accounted for by a dispreference for high-frequency referents.

Visual World data are event related. The interest is in response patterns related to the presentation of a certain stimulus. However, the time at which certain words were presented differed considerably across the fluent and disfluent condition. For instance, in the fluent condition, the eye movements at 700 ms after sentence onset are informative of how listeners responded to the target word description. In contrast, in the disfluent condition, the eye movements at 700 ms after sentence onset are informative of how listeners responded to the filler *uh*. If we were to combine the data from the fluent and disfluent conditions in one large analysis, we would be comparing apples and oranges. Therefore, separate analyses were run per fluency condition, resulting in two separate statistical models.

In both models we included a fixed effect of Nativeness (0 = native; 1 = non-native), to test for differences between native and non-native speech. Furthermore, we incorporated two time components as continuous variables (cf. Barr, 2008a; Mirman, Dixon, & Magnuson, 2008): (1) a fixed effect of Linear Time (centered at sentence onset) tested for a linear time component (linear increase or linear decrease over time). Time values were divided by 200 in order to facilitate estimation. Also (2), a fixed effect of Quadratic Time (square of Linear Time as defined above) tested for a quadratic time component (i.e., first an increase followed by a decrease, or first a decrease followed by an increase). Interactions between the factor Nativeness and the two time components were included in both models. No random slopes were included in the models, because adding random slopes for the continuous predictors Linear Time and Quadratic Time led to models whose coefficients could not be reliably estimated (i.e., non-convergence).

Multicollinearity of the fixed factors was assessed for both models by calculating cross-correlations (correlations for both models <0.13, except for the correlation between the two time components which, naturally, was very high) and the collinearity number ($\kappa_{\text{fluent}} = 18.85$; $\kappa_{\text{disfluent}} = 19.63$), indicating a medium degree of collinearity among the relevant predictors. We also tested for a cubic time component, which significantly improved the fit of the model of the disfluent data (comparisons drawn by means of likelihood ratio tests, comparing the likelihood values of models; Hox, 2010; Pinheiro & Bates, 2000; Snijders & Bosker, 1999). However, the addition of a cubic time component did not lead to a qualitatively different interpretation of results. For the sake of intelligibility, we only present models without a cubic time component. Fig. 1 illustrates the combined linear, quadratic, and interaction effects of Nativeness, Linear Time, and Quadratic Time on the estimated proportion of looks to low-frequency objects. The two models are reported in Table 2.

Inspection of the upper panel of Table 2 (data from the fluent condition) reveals that there were no effects of Nativeness or any time component: there was no preference for low-frequency objects in the fluent condition. Inspection of the lower panel (data from the disfluent condition) reveals that several predictors affected the likelihood of looks toward low-frequency pictures when listeners were presented with a disfluent sentence. The predictor Linear Time shows that there was an increase in looks toward low-frequency pictures across time. The predictor Quadratic Time reveals that there was a negative quadratic time component in the disfluent data. This indicates an increase in looks toward low-frequency pictures followed by a decrease. The interactions of the time components with the factor Nativeness counteract the effects of the time components. Only when listeners were presented with native disfluent speech did we find a preference for looking toward low-frequency pictures. Together these results demonstrate that native disfluencies elicit anticipation of low-frequency referents, but non-native disfluencies do not.

The analysis described above investigated whether there was an overall effect of disfluency on anticipatory eye movements. We also evaluated the looking behavior across the experiment. For this purpose, we tested another GLMM on the disfluent data which, in addition to the predictors introduced above, included the continuous predictor Order—testing for an increase or decrease in looks to low-frequency objects across the experiment (centered at the median). Interactions between Order and the other predictors were also included in the model. This new model estimated similar effects of Nativeness and the time components as reported in Table 2, confirming the overall preference for low-frequency objects in the native disfluent condition (and no preference in the non-native disfluent condition). Moreover, the model established an order effect in Experiment 1, but not in Experiment 2. A statistically significant interaction between Order × Linear Time, and an interaction between Order × Quadratic Time was

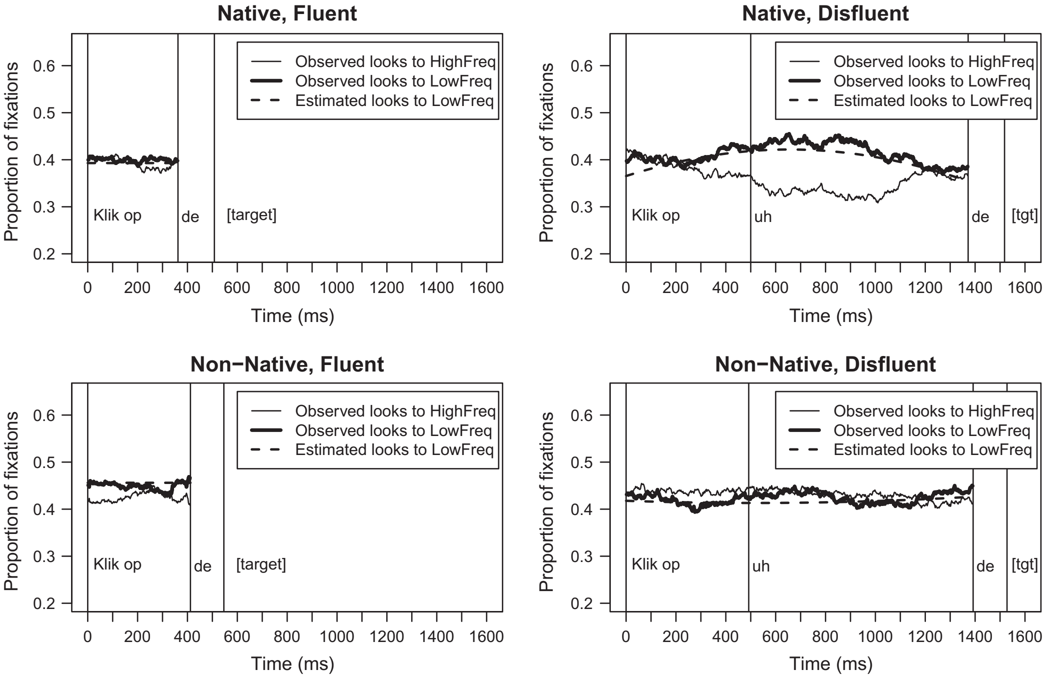

Fig. 1. Proportion of fixations, broken down by fluency and nativeness, in the pre-target time window. Time in ms is calculated from sentence onset. The thick line represents looks to low-frequency objects, the thin line looks to high-frequency objects, the dashed line represent the models' estimates of looks to low-frequency objects. Vertical lines represent the (median) onsets of words in the sentence.

Table 2
Estimated parameters of two mixed effects logistic regression models (standard errors in parentheses; pre-target time window from sentence onset to article onset) on the looks to low-frequency objects, with fixed effects of Nativeness (0 = native; 1 = non-native), and two time components (centered at sentence onset).

	Estimates	*z* Values	Significance
Model of fluent condition			
Fixed effects			
Intercept	−0.607 (0.202)	−3.01	$p = .003^{**}$
Linear Time	−0.059 (0.058)	−1.01	$p = .310$
Quadratic Time	0.018 (0.031)	0.58	$p = .565$
Nativeness	0.431 (0.254)	1.70	$p = .090$
Nativeness × Linear Time	−0.087 (0.074)	−1.18	$p = .237$
Nativeness × Quadratic Time	0.049 (0.038)	1.29	$p = .196$
Random effects			
Participant intercept	1.050		
Item intercept	0.237		
Sentence template intercept	0.006		
Model of disfluent condition			
Fixed effects			
Intercept	−0.552 (0.142)	−3.88	$p < .001^{***}$
Linear Time	0.145 (0.007)	19.42	$p < .001^{***}$
Quadratic Time	−0.022 (0.001)	−21.32	$p < .001^{***}$
Nativeness	−0.220 (0.188)	1.17	$p = .241$
Nativeness × Linear Time	−0.159 (0.010)	−15.46	$p < .001^{***}$
Nativeness × Quadratic Time	0.025 (0.001)	17.47	$p < .001^{***}$
Random effects			
Participant intercept	0.320		
Item intercept	0.075		
Sentence template intercept	0.025		

Note: $* \ p < .05$; $** \ p < .01$; $*** \ p < .001$.

found: as participants in Experiment 1 advanced through the experiment, the preference for low-frequency referents upon hearing a disfluency increased. Significant interactions between Nativeness × Linear Time × Order, and between Nativeness × Quadratic Time × Order counteracted the order effect of Experiment 1: the order effect only held for the listeners hearing native speech. We return to this order effect, present in Experiment 1 but not in Experiment 2, in the Discussion section.

The graphs in Fig. 1 illustrate the preference for low-frequency objects in the native disfluent condition. The rise in looks to low-frequency objects in the native disfluent condition starts before the median onset of the disfluency *uh*. Two factors may account for the fact that the preference arose so early. Firstly, there was some variance in the onset of the disfluency across the three disfluent sentence templates, but this variance was not very large (maximal negative deviance from the median: minus 130 ms). Secondly, the early preference may be due to the disfluent character of the disfluent sentence template as a whole, including the prosodic characteristics of the content preceding the filled pause *uh* (i.e., the initial two words *Klik op*, ‘Click on’; see Table 1). This would suggest that, in addition to the filler *uh*, the prosodic characteristics of the words preceding the filler also contributed to the perceived disfluent character of the speech.

Closer inspection of Table 1 reveals that the contrast between the prosodic characteristics of the words *Klik op* in the fluent and disfluent condition was slightly smaller in the non-native speech materials than in the native speech materials. If the prosodic characteristics of the words preceding the filler *uh* indeed contributed to the perceived disfluent character of the utterance, then the non-native disfluency condition might have been perceived as ‘relatively less disfluent’ than the native disfluency condition. In turn, this difference in perceived disfluency may even be argued to be responsible for our finding that disfluency in native speech elicited anticipation of low-frequency referents, but disfluency in non-native speech did not. In order to test this idea, a post-test was designed.

Identification post-test. We designed two identification experiments. If the contrast between the words *Klik op* in the fluent vs. the disfluent condition was smaller in our non-native speech materials (compared to the contrast between fluent and disfluent *Klik op* in our native speech materials), then naive listeners would be expected to be less accurate in perceiving this contrast (as compared to the native contrast).

In total 96 new participants took part in the two identification experiments: 49 were presented with native speech ($M_{age} = 19.96$; $SD_{age} = 1.67$; 13m/36f), and 47 with non-native speech ($M_{age} = 21.04$; $SD_{age} = 3.90$; 11m/36f). All were native Dutch speakers and participated with implicit informed consent in accordance with local and national guidelines. None of these participants had taken part in the eye-tracking experiments.

In the eye-tracking experiments, 12 sentence templates had been used: 3 templates × 2 fluency conditions × 2 speakers. From each sentence template, the words *Klik op* ‘Click on’ were excised. This resulted in 6 native and 6 non-native speech fragments that were lexically identical (i.e., no filled pauses). Two presentation lists were designed that contained either the native *Klik op* speech fragments (each fragment repeated 5 times) or the non-native *Klik op* speech fragments (each fragment repeated 5 times; 30 items per list). The order of items in each list was randomized. Half of the participants in each identification experiment was presented with a reversed presentation list.

Prior to the experiment, participants were introduced to the two fluency conditions, used in the eye-tracking experiments, and saw written examples of the fluent and disfluent instructions. They were told about the excised *Klik op* speech fragments and were instructed to identify each item as either stemming from a fluent sentence or from a disfluent sentence (two-alternative forced-choice task; 2AFC). Participants were explicitly instructed about the (non-)native identity of the speaker they would listen to.

The mean accuracy in identification of *Klik op* speech fragments for participants that had been presented with the native speech materials was 61% (fluent: 72%; disfluent: 50%). For those participants presented with non-native speech materials, the mean identification accuracy was 60% (fluent: 68%; disfluent: 51%). We built a GLMM analyzing participants' responses from both identification experiments. The fixed effect Nativeness tested for a difference in accuracy between the two experiments, but no effect was found ($\beta = -0.245,\ SE = 0.500,\ p = .624$).

The identification post-test did not show any evidence for a difference in the perceived disfluency of the native vs. non-native *Klik op* speech fragments. Therefore, a slightly smaller prosodic contrast between the fluent and disfluent condition in Experiment 2 provides an unlikely explanation for the absence of a disfluency bias for non-native speech.

Mouse clicks

Across the two experiments, participants were very accurate in their mouse clicks (Experiment 1: 99.7%; Experiment 2: 100%) such that tests for effects of fluency or frequency on accuracy were not viable. The mouse reaction times (RTs) are given in Table 3 (calculated from target onset and for correct trials only). We performed Linear Mixed Effects Regression analyses (Baayen, Davidson, & Bates, 2008; Quené & Van den Bergh, 2004, 2008) as implemented in the `lme4` library (Bates et al., 2012) in R (R Development Core Team, 2012) to analyze the mouse click RTs (log-transformed) from both experiments, with intercepts which varied randomly by participants, items, and with a by-participant random slope for Order. Adding more random effects did not significantly improve the model (comparisons drawn by means of likelihood ratio tests comparing the likelihood values of models; Hox, 2010; Pinheiro & Bates, 2000; Snijders & Bosker, 1999).

The model tested for fixed effects of Nativeness (0 = native; 1 = non-native), Disfluency (0 = fluent; 1 = disfluent), and Frequency (0 = high-frequency; 1 = low-frequency). Interactions between these three fixed effects were also added as predictors. The number of degrees of freedom used for statistical significance testing of t values was given by $df = J - m - 1$ (Hox, 2010), where J is the most conservative number of second-level units ($J = 30$ experimental items) and m is the total number of explanatory variables in the model ($m = 11$) resulting in 18 degrees of freedom.

The model showed (1) a main effect of Nativeness ($p = .017$) revealing that participants listening to a non-native speaker responded slower than participants listening to a native speaker; (2) a main effect of Frequency ($p < .001$) revealing that participants were slower responding to LF targets relative to HF targets; and (3) an interaction between Disfluency and Frequency ($p = .036$) counteracting the negative effect of Frequency: when low-frequency targets were presented in disfluent context, participants were slightly faster in their response than when the low-frequency target was presented in fluent context.

Post-experimental questionnaire

Participants in both experiments had rated the naturalness, the accentedness, and the fluency of the speech stimuli on a scale from 1 to 9 (with higher ratings indicating more natural, more accented, more fluent speech). The average naturalness of the speech was rated 7.05, $SD = 1.73$ (native) and 6.12, $SD = 1.77$ (non-native), $t(83) = 2.44,\ p = .017$. The average accentedness of the stimuli was rated 1.44, $SD = 1.33$ (native) and 6.10, $SD = 1.90$ (non-native), $t(83) = -13.11,\ p < .001$. The fluency of the speech from both experiments was rated 5.88, $SD = 2.11$ (native) and 5.36, $SD = 1.82$ (non-native), $t(83) = 1.23,\ p = .221$. Finally, participants also rated the extent to which they regularly interacted with non-native speakers of Dutch in their daily lives: 4.00, $SD = 1.99$ (native) and 3.83, $SD = 2.13$ (non-native), $t(83) < 1$.

Surprise memory test

The original design of the present study included two surprise memory tests to test for longer term effects of native and non-native disfluencies on memory. Previously, surprise memory tests following ERP experiments (Collard et al., 2008; Corley et al., 2007; MacGregor, Corley, & Donaldson, 2009; MacGregor et al., 2010) have revealed that disfluency may have a beneficial effect on the recall accuracy of target words following disfluency (as compared to target words in a fluent context). This has been argued to be due to either attentional effects of disfluencies (disfluencies orient attention to the speech stream; e.g., Collard et al., 2008; Fraundorf & Watson, 2011) or the temporal delay inherent to disfluencies (providing listeners with additional time for comprehension processes; cf. the Temporal Delay Hypothesis in Corley & Hartsuiker (2011)).

We tested longer term effects of native and non-native disfluencies on memory by means of two surprise memory tests. If disfluencies in our eye-tracking experiments would have had longer term effects on the retention of following target words, we would expect that listening to disfluencies leads to higher recall accuracy of target words. Furthermore, comparison of retention of words in native and non-native speech could potentially reveal possible differences between the effects of native and non-native disfluencies on memory. However, participants' recall accuracy

Table 3
Mean reaction times of mouse clicks (in ms, calculated from target onset and for correct trials only) in both experiments (standard deviation in brackets).

	Native speech		Non-native speech	
	Fluent	Disfluent	Fluent	Disfluent
High-frequency target	774 (244)	792 (214)	870 (277)	892 (236)
Low-frequency target	849 (267)	832 (260)	954 (271)	962 (301)

did not reveal any effect of disfluency, nor of the (non-)native identity of the speaker (cf. Supplementary Materials). We did observe a main effect of frequency: participants in both experiments were significantly more accurate recalling low-frequency objects as compared to high-frequency objects.

Discussion

The first eye-tracking experiment revealed that listeners may attribute disfluency to speaker trouble with lexical retrieval. When participants in Experiment 1 were presented with native disfluent speech, they were more likely to fixate low-frequency objects relative to high-frequency objects. This effect was observed in the pre-target time window, indicating anticipation of low-frequency referents upon encountering a disfluency. This anticipation effect persisted into the post-target time window (cf. Supplementary Materials), where participants in Experiment 1 were less likely to look at high-frequency targets (i.e., more likely to look at low-frequency competitors) when they had heard a disfluency precede the target description.

Taken together, our results suggest that listeners are sensitive to the increased likelihood of speakers to be disfluent when formulating low-frequency referents (Hartsuiker & Notebaert, 2010; Kircher et al., 2004; Levelt, 1983; Schnadt & Corley, 2006). Moreover, this sensitivity guides them to use disfluency as a cue to predict reference to a low-frequency object. This finding extends our understanding of the comprehension system. It has previously been shown that listeners may predict discourse-new (Arnold et al., 2004; Barr & Seyfeddinipur, 2010) or unknown referents (Arnold et al., 2007) upon hearing a disfluency. Our findings show that listeners may also use disfluency to infer that the speaker is encountering difficulty in lexical retrieval (i.e., formulation). This finding emphasizes the flexibility of the language architecture, particularly of the predictive mechanisms available to the listener.

Note that in our experiments, in the filler trials, the two fluency conditions were not counterbalanced across low-frequency and high-frequency targets. Instead, filler trials with a low-frequency target were always disfluent and filler trials with a high-frequency target were always fluent. In this fashion, across experimental and filler trials a natural fluent:disfluent ratio was achieved that resembled the ratio in spontaneous speech. The presence of a disfluency bias in Experiment 1 may, conceivably, be explained as a consequence of this fluent:disfluent ratio. For instance, this distribution could potentially have led listeners to learn, in the course of the experiment, to associate the presence of a disfluency with a low-frequency target. However, when we tested models that corrected for possible order effects, the estimates from the simpler models were upheld. Furthermore, the same fluent:disfluent ratio was used in Experiment 2 where no disfluency bias was observed. These arguments suggest that the observed disfluency bias in Experiment 1 cannot *solely* be explained by participants learning to associate disfluency with low-frequency referents.

Nevertheless, we did find evidence of independent order effects in the pre-target and post-target time windows. We found that, as the session progressed, listeners in Experiment 1 increasingly anticipated reference to a low-frequency target when hearing disfluent native speech. This order effect may be accounted for by listeners learning to associate disfluency with low-frequency referents in our particular experimental design. Since a comparable distribution of disfluencies is encountered in spontaneous speech (e.g., Hartsuiker & Notebaert, 2010), this behavior is likely to generalize to spoken communication in general.

The effects observed in the eye-tracking data were confirmed by the mouse click reaction times. A beneficial effect of disfluency on RTs was observed, but only for low-frequency targets (i.e., in line with the disfluency bias). However, these data are not in line with data from Corley and Hartsuiker (2011). In this study, the authors tested disfluency effects on auditory word recognition. In their experiments, participants viewed visual arrays of two objects and listened to instructions to press a button corresponding to one of them. A beneficial effect of disfluency on word recognition of both low-frequency and high-frequency objects was found. The authors explain the benefits of disfluencies on word recognition in terms of the temporal delay inherent to disfluencies which provides listeners with additional time for comprehension processes.

Comparing our study with Corley and Hartsuiker (2011), we find several methodological differences (e.g., covariation of frequency and visibility, different frequency contrast). Moreover, the stimuli in Corley and Hartsuiker (2011) contained many more filled pauses than our stimuli. Instead of comparing disfluent sentences with fluent sentences, they compared early vs. late *um*'s (i.e., 'Now press the *um* button for the [target], please' vs. 'Now press the button for the *um* [target], please'). As a consequence of these differences, participants in their experiments were presented with, what may be argued to be, a rather unnatural disfluency distribution. The relative abundance of disfluencies occurring before high-frequency words, and before 'button', may have led listeners to attenuate their disfluency-driven predictions. However, this remains a somewhat speculative account. Only new experiments may reveal whether or not listeners, within the time-course of one experimental session, indeed adjust to atypical speaker-specific disfluency patterns.

In the original design of the two experiments, we had included two surprise memory tests to study longer term effects of disfluency on memory. However, no disfluency effects on the retention of target words were observed. The surprise memory tests reported in previous studies (e.g., Corley et al., 2007; MacGregor et al., 2010), evaluated participants' recall accuracy of stimuli presented in ERP experiments, whereas our memory tests investigated recall of stimuli presented in eye-tracking experiments. Owing to this difference, the lack of a disfluency effect may be attributed to several factors. For instance, the memory tests reported by Corley and colleagues differed from our tests in the duration of experimental sessions, the total number of trials, and the linguistic content of the speech stimuli. Any of these factors may be responsible for the null result obtained here. Our data only warrant the conclusion that disfluencies, in native speech, affect the prediction of target words, but no support was found for disfluency facilitating the retention of target words.

The data from Experiment 2 allowed for a comparison between the processing of native and non-native disfluencies. We hypothesized that the disfluency bias would be either attenuated when listening to a non-native speaker, or it would be enhanced. Our data only support the first hypothesis: when listeners were presented with non-native speech (Experiment 2), the disfluency bias for low-frequency referents was attenuated. This observation is in line with previous research that also found attenuation of comprehension processes when listening to a non-native speaker (e.g., attenuation of the P600 effect in response to grammatical gender violations; Hanulíková et al., 2012). Furthermore, it extends the reported attenuation of the disfluency bias when people listen to a speaker with object agnosia (Arnold et al., 2007, Experiment 2) to a much more common situation, namely when people listen to a non-native speaker.

Because the non-native stimuli were produced by imitation of the native speech stimuli (following the method of Hanulíková et al., 2012), the native and non-native stimuli were phonetically highly similar (cf. Table 1). The principal difference between the native and non-native stimuli was the presence of a foreign accent in the non-native speech (average accent rating of 6.1 on a 9-point scale). Furthermore, the identification post-test showed that the slightly smaller prosodic contrast between the fluent and disfluent condition in Experiment 2 could not explain the absence of a disfluency bias for non-native speech. Taken together, this suggests that the attenuation of the disfluency bias in Experiment 2 may be attributed to hearing a foreign accent.

What mechanism is responsible for attenuation of the disfluency bias when listening to non-native speech? Because of prior experiences with the different distribution of disfluencies in non-native speech, listeners may modify their expectations about the linguistic content following disfluencies. Listeners may consider non-native disfluencies to be worse predictors of the word to follow and, therefore, the effect of non-native disfluencies on listeners' predictive strategies is attenuated. Similarly, Brunellière and Soto-Faraco (2013) have proposed that prior experience with the variable phonology of non-native speakers leads native listeners to have less specified phonological expectations when listening to non-native speech.

We may speculate as to the nature of the adjustments to the listener's predictions about the linguistic content following a non-native disfluency. It may be that listeners abandon all predictions entirely because non-native disfluency can be due to a wide range of cognitive troubles in speech production (e.g., conceptualization, lexical retrieval, retrieval of grammatical gender, morpho-phonological encoding, etc.). Listeners may, therefore, consider the distribution of non-native disfluencies to be too arbitrary to make any kind of reliable prediction. Alternatively, listeners may adjust their predictions to specifically exclude reference to low-frequency objects or include reference to high- and low-frequency objects to the same extent. Both hypotheses are in line with our results and further research may investigate the specific nature of listeners' adjustments when encountering disfluent non-native speech.

However, as noted by one of the reviewers, the observed interaction between the disfluency bias and the (non-)native identity of the speaker could also 'simply' be due to adverse listening conditions: listening to non-native speech is hard (Van Wijngaarden, 2001; Quené & Van Delft, 2010) and slow (Munro & Derwing, 1995), thus preventing listeners from using predictive strategies. Note, however, that the disfluency bias was elicited by the sentence template *Klik op uh*, 'Click on uh'. It is unlikely that the recognition and comprehension of these monosyllabic words seriously challenged participants in Experiment 2, since the sentence template was repeated in each trial of the experiment ($N = 60$) and was introduced prior to the experiment in the instructions to participants. Moreover, the effect of adverse listening conditions on listeners' predictive strategies is debated. Instead of a negative effect on prediction, others have argued that adverse listening conditions may encourage listeners to rely more on top-down listening strategies (cf. McQueen & Huettig, 2012), such as prediction (Pickering & Garrod, 2007). Indeed, there is ample evidence of predictive processing in listening to non-native speech (e.g., N400 effects in Brunellière & Soto-Faraco, 2013; Hanulíková et al., 2012). Therefore, we do not think that the absence of a disfluency bias in Experiment 2 is due to adverse listening conditions. Rather, we argue that listeners adjusted their predictive strategies based on prior experience with variable non-native disfluency distributions.

Note that these adjustments are stereotype-dependent: on the basis of a foreign accent, listeners draw inferences about the L2 proficiency of the non-native speaker. Apparently, listeners bring stereotypes to bear for speech comprehension when perceiving certain voice characteristics (cf. Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). This raises the question whether the effect of such stereotypes (e.g., of non-native speakers) on speech comprehension may be modulated. For instance, how would listeners respond to hearing a non-native speaker whom they know to be a very proficient L2 speaker? The absence of a learning effect in Experiment 2 seems to argue against modulation, but it remains to be seen whether the attenuation of the disfluency bias when listening to non-native speech is a gradual process, affected by the inferred proficiency of the non-native speaker.

In conclusion, the present study exposes the adaptable nature of the comprehension system. When listening to native speech, listeners may attribute disfluency to difficulty in lexical retrieval for word formulation. Listeners take into account all information available in the speech signal—even performance characteristics, such as disfluency—to achieve successful comprehension of spoken language. Our study also extends our knowledge about the role of speaker characteristics in comprehension. Hearing a foreign accent reduces listeners' use of disfluencies to predict upcoming words. Presumably, listeners realize that patterns of disfluency may be highly variable, and hence not helpful for predicting upcoming words, in non-native speech.

Funding

This work was carried out at the Utrecht institute of Linguistics OTS (UiL OTS), Utrecht University, The Netherlands, and was supported by Pearson Language Testing by means of a grant awarded to Nivja H. de Jong ('Oral fluency: production and perception').

Acknowledgments

We are grateful to Pim Mak and Holger Mitterer for helpful communication about the present study. We also thank Theo Veenker from the UiL OTS lab for technical support, Iris Mulders for help with participant recruitment and eye-tracking assistance, our native and non-native speaker for their time and patience, and Farhad Sadri Mirdamadi and Indira Wakelkamp for testing participants. Finally, we also thank three reviewers for their helpful comments on an earlier version of this paper.

Appendix A. Appendix to Bosker, Quené, Sanders, and De Jong, *Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not.*

Items used in both experiments, together with picture naming norms, retrieved from Severens et al. (2005). Frequency in counts per million. Name agreement in percentage of participants in the Severens et al. (2005) study that used the dominant name. Age of acquisition given in years, and word length in syllables.

	Dutch name	English translation	Freq	FreqGroup	Name agreement	Age of acquisition	Word length
1	accordeon	accordion	1	low	94	8	4
2	neushoorn	rhinoceros	1	low	90	7	2
3	iglo	igloo	1	low	100	7	2
4	eenhoorn	unicorn	1	low	100	8	2
5	stethoscoop	stethoscope	1	low	93	n.a.	3
6	gieter	watering can	1	low	97	5	2
7	ventilator	fan	2	low	94	n.a.	4
8	pompoen	pumpkin	2	low	94	7	2
9	naaimachine	sewing machine	2	low	95	n.a.	4
10	typemachine	typewriter	1	low	90	n.a.	4
11	ananas	pineapple	2	low	100	7	3
12	schommel	swing	2	low	100	5	2
13	sneeuwman	snowman	1	low	89	n.a.	2
14	cactus	cactus	3	low	100	6	2
15	palmboom	palm tree	3	low	95	8	2
16	stofzuiger	vacuum	3	low	97	5	3
17	zaag	saw	3	low	100	7	1
18	trechter	funnel	2	low	90	n.a.	2
19	eekhoorn	squirrel	3	low	97	7	2
20	eskimo	eskimo	3	low	100	n.a.	3
21	tandenborstel	toothbrush	4	low	95	n.a.	4
22	dolfijn	dolphin	4	low	100	6	2
23	krokodil	alligator	5	low	100	6	3
24	puzzel	puzzle	4	low	98	4	2
25	weegschaal	scale	5	low	100	7	2
26	aardbei	strawberry	5	low	100	5	2
27	vleermuis	bat	6	low	100	n.a.	2
28	slak	snail	5	low	95	5	1
29	vlieger	kite	6	low	100	6	2
30	kruiwagen	wheelbarrow	5	low	97	6	2

(*continued*)

	Dutch name	English translation	Freq	FreqGroup	Name agreement	Age of acquisition	Word length
31	brug	bridge	52	high	97	5	1
32	regen	rain	55	high	89	4	2
33	bus	bus	58	high	100	5	1
34	ster	star	61	high	100	5	1
35	maan	moon	65	high	95	5	1
36	schoen	shoe	68	high	100	n.a.	1
37	kom	bowl	70	high	95	5	1
38	vis	fish	73	high	97	n.a.	1
39	sigaret	cigarette	74	high	97	6	3
40	baby	baby	79	high	92	4	2
41	trein	train	81	high	100	5	1
42	telefoon	telephone	84	high	100	5	3
43	tand	molar	89	high	87	5	1
44	bloem	flower	94	high	100	4	1
45	koning	king	100	high	97	5	2
46	neus	nose	101	high	97	4	1
47	fles	bottle	112	high	100	4	1
48	bank	bench	114	high	92	n.a.	1
49	trap	stairs	116	high	97	4	1
50	boom	tree	137	high	100	4	1
51	muur	wall	147	high	94	5	1
52	stoel	chair	151	high	100	5	1
53	hond	dog	168	high	100	3	1
54	kerk	church	205	high	97	6	1
55	auto	car	208	high	100	4	2
56	voet	foot	225	high	100	4	1
57	tafel	table	247	high	100	7	2
58	arm	arm	266	high	95	5	1
59	deur	door	376	high	100	4	1
60	hand	hand	1028	high	97	4	1

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2014.05.004.

References

Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review, 14*, 1177–1182.

Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research, 32*, 25–36.

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say -thee uh- you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 914–930.

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*, 578–582.

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 28–55.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Barr, D. J. (2008a). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457–474.

Barr, D. J. (2008b). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109*, 18–40.

Barr, D. J., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes, 25*, 441–455.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-39. <http://CRAN.R-project.org/package=lme4>.

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*, 201–211.

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical–semantic competition hypothesis. *Cognition, 96*, B45–B54.

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer [computer program]. Version 5.3.18. <http://www.praat.org/>.

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*, 123–147.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*, 274–296.

Brunellière, A., & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and Language, 125*, 82–93.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177–208.

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

Corley, M., & Hartsuiker, R. J. (2011). Why um helps auditory word recognition: The temporal delay hypothesis. *PLoS One, 6*, e19792.

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

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*, 292–314.

Davies, A. (2003). *The native speaker: Myth and reality*. Clevedon: Multilingual Matters.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics, 13*, 1–24.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review, 15*, 850–855.

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709–738.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition, 29*, 320–326.

Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language, 65*, 161–175.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*, 787–814.

Hanulíková, A., Van Alphen, P., Van Goch, M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*, 878–887.

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

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York: Routledge.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*, 151–171.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 824–843.

Kahng, J. (2013). Investigating utterance fluency and cognitive fluency in second language speech. In *Presentation at the new sounds 2013 conference*. Montréal, Canada.

Kircher, T. T. J., Brammer, M. J., Levelt, W. J. M., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: Engagement of left temporal cortex during pauses in speech. *NeuroImage, 21*, 84–90.

Kutas, M., DeLong, K., & Smith, N. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Using our past to generate a future* (pp. 190–207). Oxford University Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

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

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2009). Not all disfluencies are equal: The effects of disfluent repetitions on language comprehension. *Brain and Language, 111*, 36–45.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Investigating the consequences of disfluent silent pauses in speech for listeners. *Neuropsychologia, 48*, 3982–3992.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19–44.

Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception & Psychophysics, 3*, 427–438.

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America, 131*, 509–517.

Meyer, A. S., Roelofs, A., & Levelt, W. J. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language, 48*, 131–147.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*, 475–494.

Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology, 91*, 167–180.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*, 289–306.

Oostdijk, N. (2000). The spoken Dutch corpus project. *ELRA Newsletter, 5*, 4–8.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105–110.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer Verlag.

Quené, H., & Van Delft, L. E. (2010). Non-native durational patterns decrease speech intelligibility. *Speech Communication, 52*, 911–918.

Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication, 43*, 103–121.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*, 413–425.

R Development Core Team (2012). R: A language and environment for statistical computing. <http://www.R-project.org/> (ISBN 3-900051-07-0).

Riazantseva, A. (2001). Second language proficiency and pausing. *Studies in Second Language Acquisition, 23*, 497–526.

Schnadt, M. J., & Corley, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the twenty-eighth meeting of the cognitive science society* (pp. 750–755).

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

Severens, E., Lommel, S., Ratinckx, E., & Hartsuiker, R. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119*, 159–187.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*, 510–532.

Skehan, P., & Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. In S. Van Daele, A. Housen, F. Kuiken, M. Pierrard, & I. Vedder (Eds.), *Complexity, accuracy, and fluency in second language use, learning, and teaching* (pp. 207–226). Brussels: University of Brussels Press.

Snijders, T. A., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications Limited.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*, 71–79.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 443–467.

Van Berkum, J. J. A., Van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*, 580–591.

Van Wijngaarden, S. J. (2001). Intelligibility of native and non-native dutch speech. *Speech Communication, 35*, 103–113.

Veenker, T. J. G. (2012). The ZEP experiment control application. Computer software, version 1.2. <http://www.hum.uu.nl/uilots/lab/zep/>.

Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication, 50*, 81–94.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition, 99*, 63–72.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*, 1272–1288.