Learning to Expect the Unexpected:

Speech Disfluencies, Expectancy, and the N400 Effect

A Senior Honors Thesis for the Cognitive and Brain Sciences Major, Department of Psychology

Barbara Storch

Tufts University

Advisor: Gina Kuperberg, M.D., Ph.D.

Thesis Committee: Gina Kuperberg, M.D., Ph.D., Ariel Goldberg, Ph.D., Nathaniel Delaney-Busch, Ph.D., Meredith Brown, Ph.D.

Abstract

Prior work has indicated that speech disfluencies are produced in a systematic fashion. Specifically, speakers are more likely to produce a speech disfluency before an unexpected word or phrase. This patterned production of disfluencies may lead listeners using disfluencies as a signal to "expect the unexpected" continuation (Arnold & Tanenhaus, 2011). Disfluencies have been shown to attenuate the N400 effect in previous studies (Corley, MacGregor, Donaldson, 2007). Additionally, prior research has shown that listeners can change the way they use disfluencies based on given prior knowledge about the speaker (Bosker et al., 2014). The present study aims to examine first whether people can learn if disfluencies are a reliable signal, and then to examine the nature of the attenuation of the N400 effect following disfluencies. We adopted a between-subjects design in which one speaker uses disfluencies reliably to signal an unexpected completion, one uses disfluencies in an unreliable, unsystematic fashion. The preliminary data reported here shows a statistically significant N400 effect of cloze, as well as a significant N400 effect of fluency. These results indicate a proof of concept of the classic N400 effect for cloze, and signals that listeners are utilizing disfluencies in a systematic fashion to aid in prediction, although the nature of that utilization is not yet clear. The present sample size is not large enough to fully examine both the interaction between expectancy and fluency as well as the between subjects manipulation, which is a limitation of the present results.

Introduction

Prediction is a central feature of language comprehension. Having patterns and systems in linguistic contexts enables listeners to create predictions in order to comprehend and process language more efficiently. Instead of needing to construct an entirely new conversational schema with every novel piece of information the speaker provides (this can be at the word level, or the sentence level), listeners can create predictions about what speakers will say next. Kuperberg & Jaeger (2016) discuss how conversation can occur fluently even in noisy environments *because* speakers are able to predict. Predictions are created by listeners based on the lexical and syntactic content of the latest information in the conversation, which creates a context-based internal representation. This internal representational state can in turn facilitate the processing of the information that does come in. In the present experiment, we focus on how speech signals like disfluencies (e.g. "um" and "uh") can influence how listeners make these predictions during comprehension.

Past research has indicated that speech disfluencies are produced in a systematic fashion. In particular, speakers are more likely to produce a speech disfluency before an unexpected piece of information than before a more highly expected completion or continuation (Beattie & Butterworth, 1979). In other words, instances of speech with lower degrees of expectancy, or predictability, are likely to have higher instances of hesitations immediately preceding them. For example, say we have a sentence that begins "The man placed the sheet on the …". The most expected completion, or the completion with the highest *cloze probability*, would be the word "bed." However, the completion "clipboard" is still sensical, although has substantially lower cloze probability (is less expected). So, prior research has shown that speakers are more likely to produce a disfluency before words like "clipboard" rather than "bed." Perhaps when a listener hears a disfluency before the final word, they conceptualize that the speaker is having trouble retrieving something; they understand that the speaker still intends to hold the floor, but needs an extra moment to figure out what they are trying to say (Clark and Fox Tree, 2002). The listener can then infer both that they should not interrupt and that the speaker is having a word production challenge, so what they are about to say will most likely be unexpected or challenging.

Because speakers tend to produce disfluencies in a systematic pattern, listeners could potentially utilize the disfluencies as a tool for guiding predictions, thus facilitating comprehension. As disfluencies are much more likely to be produced before an unexpected completion, the disfluencies may actually serve as a reliable signal to the listener to expect an unexpected continuation. This "expect the unexpected" pattern has in fact been shown to be reliably produced in the eye-tracking literature. Arnold & Tanenhaus (2011) presented participants with a visual display containing four objects, and then gave them two instructional sentences about two of the objects in the display (for example "put the grapes below the candle."). The second of the instructional sentences was either fluent or contained a disfluency prior to the second object named. The second object either referred to the same object referenced in the first sentence, or something new (but still in the visual display). They found that in disfluent conditions, eye movements revealed a preference for discourse-new objects, whereas for fluent conditions there was an eye movement preference towards given objects. Thus, participants seemed to be showing a preference for discourse-new, or unexpected, items once they heard the disfluency. Bosker et al. (2014) performed a similar eye-tracking experiment, but showed participants a visual paradigm with high and low-frequency objects where they were asked, either fluently or disfluently, to click on an object. They again found that in the disfluent conditions participants showed an anticipatory preference towards the low-frequency object. Both of these experiments show that listeners have a preference to anticipate something discourse-new, or unexpected, once they hear a speaker produce a disfluency.

This “expect the unexpected” pattern that listeners create upon hearing a disfluency has also been observed on the N400 ERP component. The N400 component is an negative-going event-related brain potential that is modulated by context and cloze probability: if an incoming word is highly supported by its context, the N400 amplitude is attenuated (Kutas & Federmeier, 2011). However, if an incoming word is not activated by the previous context, the N400 amplitude is larger. For example, a sentence such as “It was windy outside so the boy went to fly his kite” would produce an attenuated N400 amplitude, as “kite” (the predicted completion of this sentence) relative to some other less expected word like “balloon”. The context so far has created a highly-constraining, very strong prediction for “kite”. The word fits right in, and the listener does not have to do much work to integrate “kite” into their pre-existing context, because they have already begun the work of processing “kite” into their prediction before the completion arrived. In contrast, while still grammatically and linguistically sensible, the completion of “balloon” violates the strong prediction that the context of the sentence has created. The speaker therefore must work a little harder to integrate this unexpected new information, resulting in a larger, more negative-going, N400 peak. The N400 effect (the difference in amplitude between an expected and unexpected completion) is well documented across a variety of modalities (spoken language; written, visually presented language; signed language), and is a reliably-produced effect that can be understood as as the extent to which semantic expectations were met during language comprehension (Kuperberg & Jaeger, 2016).

The ERP literature, in addition to the eye-tracking literature, has also shown that listeners produce a disfluency-catalyzed "expect the unexpected" pattern, shown in this N400 effect. In particular, disfluencies produce an attenuation in a listener's N400 cloze effect. Corley, Macgregor & Donaldson (2010) found this attenuated N400 effect when a disfluency came before an unexpected word. This attenuation could result from having the "expect the unexpected" pattern, originally activated by the disfluency, fulfilled when the unexpected completion arrived. These results suggest there is perhaps a processing benefit that comes with having the "expect the unexpected" pattern fulfilled.

Additionally, the "expect the unexpected" pattern that is activated when a listener hears a speech disfluency can be mediated by the listener's knowledge of the speaker. Comprehenders seem to utilize disfluencies more during comprehension when they believe the disfluencies are more informative (Bosker et al., 2014). For example, Arnold, Kam, & Tanenhaus (2007) performed an eye-tracking study with familiar and unfamiliar pictures; however, in one experiment their participants were told that the speaker they would be listening to had agnosia. With this information, the preference towards an unexpected completion upon hearing a disfluency was greatly reduced. In other words, based on the prior information the listeners' had about the speaker, the "expect the unexpected" pattern was highly attenuated. Listeners changed what they would deem as challenging for a speaker to retrieve based on what they have been previously told about that speaker's abilities. Thus, the "expect the unexpected" bias is intertwined with the listener's' knowledge about what might be challenging for their particular speaker to retrieve. However, it is unknown whether listeners can implicitly adapt to the the informativeness of disfluencies when they are not given any speaker-specific information ahead of time. Can listeners still manage to reduce their "expect the unexpected" bias based not on prior information about a speaker's abilities, but rather their own experience with a particular speaker?

The current experiment first aims to address this general interaction between expectancy and fluency. Both expectancy and disfluencies relate to on-line processing of language, and how one can create and update expectations in real time. Thus, throughout the current experiment, it is hypothesized that that disfluencies followed by a low-cloze word would have an attenuated N400 amplitude as opposed the N400 amplitude of disfluencies followed by a high-cloze word. Thus, the overall N400 cloze effect will be attenuated (the difference between the two conditions will be smaller) with the presence of disfluencies. If so, this would indicate that disfluencies are, in fact, a reliable and useful signal for listeners' in creating predictions in conversation. In addition to the N400 effect, this experiment also aims to address the effects of disfluencies on the late positivity effect, an effect elicited by unexpected words completing strongly constrained sentences (Federmeier et al., 2007). The addition of this component is a novel aspect of the study, as it aims to examine the interaction of disfluencies, expectancy, and sentential constraint

The current experiment also aims to address listeners' longer-term expectations about a speaker's reliability without any a priori knowledge about the speaker and their tendencies. In other words, it examines how listeners use information they learn about the speaker over the course of the experiment to create predictions. The literature, as described previously, has shown that individuals can change how they "use" disfluencies to create predictions depending on their knowledge of the speaker (Arnold, Kam, & Tanenhaus (2007). However, an important aspect of the aforementioned study is that listeners were told in advance about the speaker's conversational differences. They did not have to gauge this knowledge for themselves, and learn about the speaker over time, as might be the case in many real-world settings. To address this limitation of the prior research, we performed a between-subject design where half of the participants hear a reliable speaker, who always uses disfluencies in front of the predicted lower-expectancy words and not before highly predictable words. But, half of the participants hear a speaker who uses disfluencies 50 percent of the time in front of an expected word, and 50 percent of the time in front of an unexpected word. Neither group is told any information about the speaker in advance. We hypothesize that the participants who hear the unreliable speaker will gradually learn that the speaker is unreliable, and not utilize the disfluency to create the "expect the unexpected" pattern. Rather, the participants in this group will begin to almost ignore the disfluencies, as they are not being used systematically and in a conversationally relevant way. Thus, listeners who hear the unreliable speaker will perhaps produce classic N400 cloze effects without the attenuating factor of the disfluency, by the end of the experiment. However, listeners who hear the reliable speaker may begin to produce very attenuated N400 effects, as they learn that their speaker consistently uses disfluencies before an unexpected completion.

Methods

Construction of Stimuli

Each participant heard 192 experimental sentences and 192 filler sentences. First, 128 highly constraining sentence stems were obtained (courtesy of Edward Wlotko). Then, a highly expected continuation and an unexpected but still plausible continuation was obtained for each stem. For example, a highly constraining stem such as "He put a clean sheet on the" would have the highly expected continuation of "bed" and the unexpected, yet plausible continuation of "sheet" (see Table 1 for further examples). Half of these unexpected words were then additionally placed inside different low-constraint stems, for a total of 64 low constraint, unexpected but plausible continuations. For example, the highly constraining, unexpected sentence "As the weather got warmer, Anne started wearing her lighter jacket instead of her heavy winter *stockings*" had a low constraining, unexpected sentence with the same critical word ("Most men cannot understand why women wear *stockings*"). These 64 low-constraining sentences all had the same critical word as 64 of the high-constraint unexpected scenarios. Finally, we obtained 192 fillers each with an expected and an unexpected continuation (see Table 2 for the overall stimulus design).

Stimulus Characteristics

The context sentences (which would manipulate constraint) and the critical words (which would manipulate cloze) were both part of an existing stimulus set. The average cloze probability of the expected completions was 85.2%, and the average cloze probability of all of the unexpected completions was 3.28%. The critical words were approximately matched for phonological neighborhood, word frequency, number of phonemes, and number of syllables, using data from the English Lexicon Project's online database (Balota et al., 2007). Critical words were selected on an additional number of phonological characteristics, in order to ensure that any experimental result would not be able to be attributed to the prosodic cues of the previous words. Critical words starting with vowels (especially if they were preceded by 'the' or 'a'), liquids, and glides were avoided. The average cloze probability of the expected completions for the fillers was 63%, and the unexpected completions were created based off of the existing normed stems within the guidelines of the phonological criteria described above.

Splicing

Each highly constraining sentence was recorded six times, and then spliced together to create the final stimulus set (please see Table 3 for a detailed schematic). The recording software used was Audacity, and the sentences were all recorded by a single speaker in a sound-proof room. Praat software (Boersma & Weenink, 2010) was used to splice the recorded sentences together. In the recording session, the speaker first read the context up until the critical word, and in place of the critical word said a filler word; this context was read twice, once fluently and once disfluently. In splicing, these sentences were used to extract the disfluency and the syllable before it for disfluent conditions. This would ensure that the disfluency used was the same across expectancy, and that contexts of the sentence weren't getting any subtle effects from either the upcoming disfluency and/or the upcoming critical word. The next two sentences were fluent sentences, one with the expected and one with the unexpected completion. In splicing, these two sentences were used to extract the sentence final critical words (the expected and the unexpected completion). The final two sentences were both disfluent, again one with the expected and one with the unexpected completion. In splicing, either the the disfluent-unexpected sentence or the disfluent-expected sentence was used as the sentence stem (Please see Table 3 for a full schematic of the splicing). Half of the filler sentences throughout the experiment were also spliced, to ensure that there was no systematic listener bias that came from hearing only the critical sentences being spliced.

Lists and Subsets

Utilizing the final set of recorded stimuli, the stimuli were distributed among four distinct lists. The four high-constraint conditions were distributed pseudorandomly across these four lists in a counterbalanced manner. List assignment for these stimuli was first constrained in order to also counterbalance the low-constraint conditions, which were only available for half of the stimuli. Each low constraint unexpected sentence was placed in the same list as the corresponding high constraining unexpected sentence. For example, the highly constraining "He put a clean sheet on the bed" as a fluent condition would be in the same list as the disfluent low constraint "In his haste, the young man had forgotten to sign the uhhh clipboard." The remaining condition of "He put a clean sheet on a the clipboard" was assigned to a different list, to minimize repetitions within lists.

List assignment was also constrained to minimize potential effects of repetitions; within the highly constraining condition and the filler items, there was some final-word repetition both within and between conditions. To address repetitions (there were some repetitions within high constraint, low constraint, and filler sentences, as well as some repetition across the conditions), the highly constraining repetitions were distributed first to ensure that there were not repetitions of those within the same list. Additionally, any low-constraint item that had the same critical word as a highly constraining item within the same list always appeared later in the experiment than the corresponding high-constraint item, since processing of low-constraint items should be less affected by word repetition effects than high-constraint items. After these steps, the rest of the high constraining words were distributed randomly throughout the lists to ensure the appropriate number of sentences was in each list. This process of list creation was done six times, for a total of 24 distinct lists (one for each participant in each group), to ensure that each participant saw a different composition and order of stimuli in a manner that still remained counterbalanced.

Finally, fillers were added to each list in a manner that would reflect the between-group manipulation of the reliable and unreliable speaker. A version with reliable fillers and a version with unreliable fillers was created for each list, such that each of the 24 participants in the reliable group would see the exact same experimental items as the corresponding participant in the unreliable group. Half of the total of 192 filler sentences contained disfluencies. For the reliable group, in the 96 disfluent sentences, all disfluencies preceded unexpected words. This is in-line with the literature showing that disfluencies are typically used by speakers in front of an unexpected completion, so that listeners may use them to create the “expect the unexpected” pattern. In the unreliable group, 48 disfluent sentences had a disfluency before an unexpected word, and 48 had a disfluency before an expected word. This would ideally influence the creation of the “expect the unexpected” pattern by not giving listeners’ a systematic input from which to create predictions. In total, there were 24 total lists, each of which had a reliable and an unreliable version; the versions contained the exact same experimental items, and the only difference was the disfluency manipulation described above.

Participants

For this preliminary reporting of data, a group of eight Tufts undergraduate students between the ages of 18 and 22 participated in the ERP study. Every participant had normal or corrected to normal vision, no history of hearing problems or current hearing problems, and no history of head trauma. All participants were right-handed, native English speakers with no history of, psychiatric disorders, learning disorders, or neurological disorders. Each participant provided informed consent according to the requirements of the Institutional Review Board of Tufts University, and received hourly compensation for their time.

Experimental Procedure

Participants sat in a quiet room with a single, dim light and listened to the stimuli over a pair of headphones. Volume was kept consistent for all participants except for one, who requested an increase in volume which was documented. The entire experiment was self-paced, and began with the showing of a 500ms fixation cross with the playing of the sentence. The participants were asked to keep their eyes steady on the fixation cross while listening to the sentences, in order to attempt to reduce any sort of head movement, eye movement, blocking, or blinking artifacts. After hearing the sentence, participants were occasionally asked to answer a Yes/No comprehension question; these comprehension questions had to do with the content of the previous sentence, but required enough reasoning to hopefully ensure that the participant was paying full attention. These comprehension questions were not used to collect any data other than information about the engagement-level of the participants. The participant was then shown a blink sign, and could decide when to continue to another sentence.

EEG Recording

A 29 electrode elastic cap was used to collect the data (Electro-Cap International Inc., Eaton, OH). Additional electrodes were placed on both the left and right mastoid area (behind each ear), below the left eye, and on the outer canthus of the right eye. For data collection, the electroencephalography (EEG) signal was referenced to the left mastoid. When setting up each participant, impedance was kept below 5 kΩ for all scalp electrodes, below 2.5 kΩ for each mastoid electrode, and below 10 kΩ for the two eye channels. An Isolated Bioelectric Amplifier System, Model HandW-32/BA (SA Instrumentation San Diego, CA) was used to amplify the EEG signal with a bandpass of 0.01 to 40 Hz; the signal was continuously sampled at 200 Hz by an analog-to-digital converter. The stimuli and behavioral responses were monitored simultaneously by a digitizing computer, as well as a researcher to ensure optimal focus and blinking throughout the experiment. Trials were rejected if they contained missing data, a blink, large head movements, or a "blocked" electrode. We also filtered the data with a low-pass butterworth filter at 30 Hz, and normalized the amplitude data for each participant using the average of 200 known 10-microvolt calibration pulses.

Statistical Analysis

Following artifact rejection, EEG responses were time-locked to the critical words in the experimental sentences. The responses were averaged across each participant using the time window of 300-500ms to examine the N400 effect, and were referenced to the average of the left and right mastoid. Two main sections of the scalp were used to examine how the N400 varied across the scalp, averaged over the electrodes within those sections. A repeated-measures ANOVA was used to examine both the central mid-region and the parietal mid-region

In the repeated-measures ANOVA, the within-subjects variables were Expectancy (expected, unexpected completion) and Fluency (fluent, disfluent sentence). Alpha was set at 0.05 for all of the tests.

Results

The grand-average ERP-waveforms for the main effect of expectancy are shown are shown in Figure 1. The main effect of fluency is shown in Figure 2.

Behavioral Data

Throughout the experiment, participants were asked to randomly respond to Yes/No comprehension questions to ensure a high level of focus on the task at hand, as well as general understanding of the task. Each participant scored over 90% of the comprehension questions correctly ($M = 97.6$, $SD = 0.03$). This high proportion of correct answers allows inference that all participants were engaged in the task at an appropriate level throughout the experiment.

N400: 300-500ms

The N400 component was maximal over two mid-regions of the scalp: the central mid-region and the parietal mid-region. Repeated measure ANOVAs in both the central mid-regions and parietal mid-regions showed significant main effects of expectancy (central mid region: $F(1, 7) = 24.97$, $p = 0.002$; parietal mid region: $F(1,7) = 20.69$, $p = 0.003$). For the parietal mid-region only, there was also a significant main effect of fluency ($F(1,7) = 8.18$, $p = 0.024$). At this time, the interaction of expectancy and fluency did not reach significance, a result that was most likely influenced by the current small sample size. However, a Bayesian Repeated Measures ANOVA was also conducted. The Bayes factor of M for the interaction between expectancy and fluency compared to the null model was 4.880 (interaction) compared to 0.005 (null)for the parietal mid-region.

Discussion

As a proof of concept and a reliable replication of past work, the current data do show a significant N400 cloze effect for the central mid-region as well as the parietal-mid region. This effect, while not directly informative about the role of speech disfluencies on the N400 effect, does serve the function of showing that the stimuli themselves do reliably manipulate cloze; even with the current small sample size, the N400 cloze effect is robust and statistically significant for the central mid-region as well as the parietal mid-region.

This study in particular aims to address how individuals can change the way they utilize disfluencies to create predictions based on information they learn about a speaker over the course of the experiment through a between-groups design. While the study currently does not have a large enough sample size to examine the between-groups manipulation, we did perform analysis on a nested aim of this study, which was to examine how disfluencies modulate the “expect the unexpected” pattern. We expected the N400 amplitude to be attenuated when a disfluency preceded a high-cloze word, as opposed to the N400 amplitude of a low-cloze word preceded by a disfluency (Corley, MacGregor, & Donaldson, 2007). While the interaction between expectancy and disfluency did not reach statistical significance for the current sample size, the factor of fluency itself did reach statistical significance for the parietal mid-region. This significance again provides somewhat of a proof of concept that for the N400 effect, disfluencies do affect processing. While the nature of that influence is not yet clear, the significant result does imply that listeners are utilizing disfluencies somewhat systematically to make predictions.

Currently, although the interaction between expectancy and fluency is not at a level of statistical significance, for the parietal mid-region (the same location for which disfluency itself was a significant factor), the data was also preliminarily analyzed by estimating a Bayes factor utilizing the Bayesian Repeated Measures ANOVA. The Bayes Factor of M reveals that an interaction between disfluency and expectancy seems to be roughly four times more probable than the null hypothesis, or a complete lack of interaction. A Bayesian analysis addresses not the binary cutoff of significance that the p-value often implies, but rather the probability that the data is reflective of the alternative hypothesis rather than the null hypothesis. In other words, a Bayesian analysis addresses how likely it is that one's experimental manipulation is driving the results over the likelihood of the null hypothesis driving the results given the data that was collected (Jarosz & Wiley, 2014).

An estimated Bayes factor (null: alternative) reported that the data were in favor of the alternative hypothesis of an interaction of fluency and expectancy. This shows that there is a higher probability, at the moment, of the interaction between fluency and expectancy driving the results. These data, although not conclusive, do perhaps suggest that the predicted "expect the unexpected" result will in fact be produced; disfluencies will serve to lead speakers to create predictions about an upcoming unexpected word, which will attenuate the typical N400 cloze effect.

At present, although there is not enough evidence to examine the group effects for reliability, we would expect a large fluency and expectancy interaction for the reliable group.

Hypothetically, over the course of the experiment, the unreliable group would end up producing just a classic N400 cloze effect rather than the disfluency-catalyzed attenuated N400 effect. This would be because the listeners learn over time that the speaker is not producing disfluencies in a systematic way, so the listeners can't (or shouldn't) use the disfluencies to create systematic "expect the unexpected" predictions based off of the disfluencies. As participants continue to be run, there will be opportunity to look at these data at the group level, in order to see if participants do in fact learn about how listeners are using a speakers' reliable or unreliable disfluency across the course of an experiment, and then use that information to modulate their predictions.

Limitations and Future Directions

An immediate limitation of this study is the current small sample size; however, the study will continue to be run with the ultimate goal of a sample size of 48. This increase in sample size will result in an increase in statistical power, allowing for more robust interpretations of the expectancy and frequency interaction, as well as allowing for an exploration of the between groups interaction of the reliable and unreliable speaker.

At the present moment, however, we can conclude that our results are in-line with the classic production of the N400 effect, and are hopefully moving in the direction of showing an "expect the unexpected" interaction in the fluency and expectancy interaction.

If the results do indeed confirm the "expect the unexpected" interaction, this will provide a replication of an effect of relatively recent interest in the speech prediction ERP literature, and will extend the current knowledge base by using a design that examines the effects of disfluencies on multiple ERP signatures of language production (the N400 effect and the late positivity). Additionally, ideally, viewing the between groups effects will offer further information as to how disfluencies are used in creating predictions; can a listener learn over the course of the experiment about a speaker's tendencies, and choose to "ignore" irrelevant pragmatic information such as these disfluencies. In sum, this information will enable us to examine further how exactly disfluencies and expectancy interact to aid listeners' in making predictions about language. This has implications in the larger field of pragmatics, as it relates to listeners learning and then being able to use information about a speaker to help facilitate their comprehension.

References

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

Arnold, J. E., & Tanenhaus, M. K. (2011). Disfluency effects in comprehension: How new information can become accessible. *The processing and acquisition of reference*, 197-217.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. Behavior Research Methods, 39, 445-459.

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), 201-211.

Bögels S, Kendrick KH, Levinson SC (2015) Never Say No … How the Brain Interprets the Pregnant Pause in Conversation. PLoS ONE 10(12): e0145474. doi:10.1371/journal.pone.0145474

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native ‘ums elicit prediction of low-frequency referents, but non-native ‘um’s do not. *Journal of memory and language*, *75*, 104-116.

Clark and Fox Tree (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*(1). 73-111.

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*(3), 658-668.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research*,*1146*, 75-84.

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.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*,*7*(1), 2.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience*, *31*(1), 32-59.

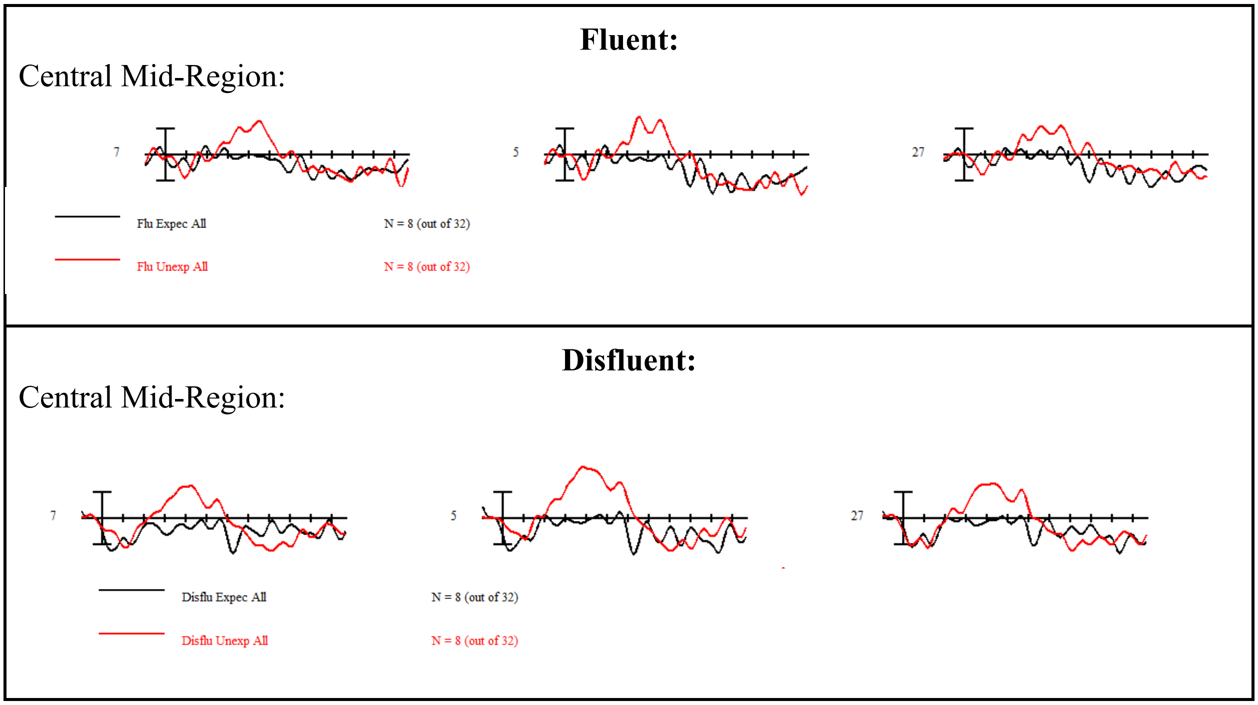
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP).*Annual review of psychology*, *62*, 621.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners.*Neuropsychologia*, *48*(14), 3982-3992.

Scontras, G., Badecker, W., Shank, L., Lim, E., & Fedorenko, E. (2015). Syntactic Complexity Effects in Sentence Production. *Cognitive science*, *39*(3), 559-583.

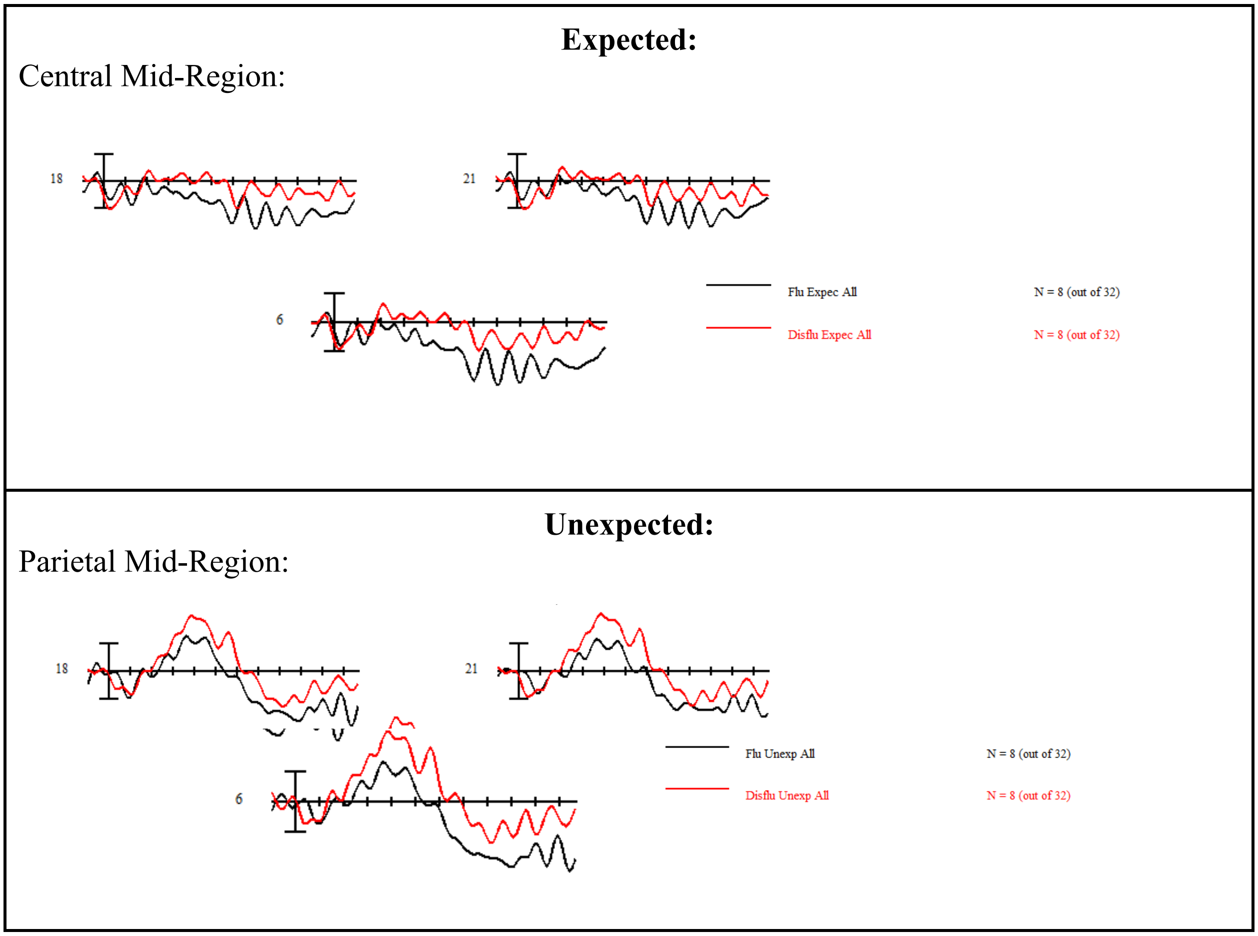
Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356-366.

Figure 1. *The main effects of expectancy*.



The main effect of expectancy showed a significant N400 effect for both experimental regions.

Figure 2. *The main effects of disfluency.*



The main effect of fluency showed a significant N400 effect over the parietal mid0region, but not a significant effect for the central mid-region.

Table 1. *Stimulus Examples*

Constraint:	**Expectancy:**
Low Constraint: In his haste, the young man had forgotten to sign the	*Unexpected:* clipboard
High Constraint: He put a clean sheet on the	*Unexpected:* clipboard
High Constraint: He put a clean sheet on the	*Expected:* bed
Low Constraint: Helen reached up to dust the	*Unexpected:* dresser
High Constraint: The groom took the bride's hand and placed the ring on her	*Unexpected:* dresser
High Constraint: The groom took the bride's hand and placed the ring on her	*Expected:* finger

One third of all sentences were low constraint with exclusively low cloze completions. Two thirds of the sentences were highly constraining, half of which had high cloze completions and half of which had low cloze completions; each low constraint-low cloze completion had a corresponding high constraint-low cloze completion.

Note: for all recorded sentences, the disfluency was placed before the final word

Table 2. *Experimental Design*

Total Items: 384			
Experimental Total: 192		Fillers Total: 192	
Low Constraint	High Constraint	"Reliable" (Fillers)	"Unreliable" (Fillers)
32 **disfluent** unexpected	32 **disfluent** expected	96 **disfluent** before unexpected	48 **disfluent** before expected
32 fluent unexpected	32 **disfluent** unexpected	96 fluent before expected	48 **disfluent** before unexpected
	32 fluent expected		48 fluent expected
	32 fluent unexpected		48 fluent unexpected
		Percent of Reliably Used Fillers: 1	*Percent of Reliably Used Fillers: 0.5*
Total Low Constraint: 64	Total High Constraint: 128	Total "Reliable" Fillers: 192	Total "Unreliable" Fillers: 192

The complete stimulus had half experimental words and half fillers; the fillers are the location of the between group manipulation. For the reliable group, all disfluencies were before unexpected completions, reflecting the systematic way in which they are typically used. For the unreliable group, half of the disfluencies were before unexpected words and half were before expected words, so that listeners had no reliable pattern of disfluencies to make predictions about.

Table 3. *Splicing Scheme*

Recorded Sentences 1. She had to leave the movie when the baby started **to** bacon 2. She had to leave the movie when the baby started **to uhhh** bacon 3. She had to leave the movie when the baby started to **cry** 4. She had to leave the movie when the baby started to **smell** 5. She had to leave the movie when the baby started to uhhh cry 6. **She had to leave the movie when the baby started** to uhhh smell
Spliced Sentences **She had to leave the movie when the baby started to cry** **She had to leave the movie when the baby started to smell** **She had to leave the movie when the baby started to uhhh cry** **She had to leave the movie when the baby started to uhhh smell**

The above shows the components taken from each recorded sentence to create the final spliced sentences.