Do Listeners Care About the Speaker or the Input? Evidence from Structural Priming

by

Douglas J. Getty

B.S, Westminster College, 2016

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Douglas J. Getty

It was defended on

March 9, 2021

and approved by

Tessa Warren, Ph.D., Department of Psychology

Marta Ortega-Llebaria, Ph.D., Department of Linguistics

Thesis Advisor: Scott Fraundorf, Ph.D., Department of Psychology

Do Listeners Care About the Speaker or the Input? Evidence from Structural Priming

Douglas J. Getty, M.S.

University of Pittsburgh, 2021

Past findings suggest that the representation of non-native speech may be underspecified relative to native speech during comprehension (Hanulíková et al., 2012; Lev-Ari, 2015, Gibson et al., 2017). Across two experiments, I explore two possible reasons why this might occur: rational expectation for reduced speaker reliability or ad hoc adaptation to the particulars of the given input. To probe these explanations, I examined structural processing of native and non-native speech. In experiment 1, I show that structural priming is reduced for non-native speech relative to native speech. However, I also find that this pattern reverses in the second half of the experiment. In experiment 2, I show that a less proficient non-native speaker elicits *more* priming than a more proficient non-native speaker, inconsistent with the expectation explanation. I suggest that listeners' comprehension of non-native speech is governed by contextually optimized processing strategies rather than a global reliance on speaker information when processing non-native speech.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

Contemporary theories of language processing generally hold that comprehension is fast and accurate in part because it draws on experience with the linguistic environment (Jurafsky, 1996; Levy, 2008; MacDonald et al., 1994). But, variability in spoken language can pose a challenge by rendering prior experience unhelpful or even misleading (Liberman et al., 1967). Examples of such situations include comprehension of novel constructions (Fraundorf & Jaeger, 2016; Kaschak & Glenberg, 2004), new word meanings (Fang et al., 2017), or unfamiliar accents (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Kraljic & Samuel, 2005; Maye et al., 2008). In such situations, listeners need to either adapt to the unfamiliar features of the input and/or generalize features of their past experience to the input (Kleinschmidt & Jaeger, 2015). In the current set of studies, I examine how language processing changes when listening to native and non-native speakers and explore whether listeners generalize their past experience to aid in the comprehension of non-native speech or simply adapt their processing strategies to manage the input.

Specifically, the current set of studies examine the influence of non-native speech on processing of syntactic structure using a structural priming paradigm. A growing number of studies have investigated the specific interaction of non-native speech and structural priming (Chun et al., 2016; Chun & Kaan, 2020; Constantine et al., In Prep; Weatherholtz et al., 2014; Williams et al., 2020) however the task and design across these studies varies and the results are often contradictory. A key question when evaluating the influence of non-native speech on structural priming is whether the observed effects are due to listeners deploying knowledge about non-native speakers to guide processing, or whether the effects are due simply to comprehenders managing the perceptual characteristics of the input.

1.1 Non-Native Speech

Non-native speakers typically do not achieve native-like proficiency in an L2 (Birdsong & Molis, 2001), which has effects on many levels of spoken language. Reduced proficiency results in non-standard prosody, accented speech, higher rates of syntactic errors (Munro, 1993; Trofimovich & Baker, 2006). The combined effect of all of these deviations from "standard"[1] native speech is that listeners' comprehension processes must change to accommodate comprehension across many levels of the language input. Past research has indicated that knowledge about speaker proficiency may guide structure processing during overtly anomalous or ungrammatical speech input (Arnold et al., 2007; Caffarra & Martin, 2019; Gibson et al., 2017; Hanulíková et al., 2012; Kamide, 2012; Seifeldin et al., 2015), but we do not know whether proficiency information may influence processing of speech that is accented but otherwise well-formed grammatically. It could be the case that expectations of speaker proficiency guides comprehension globally during comprehension of speech, or it could be that proficiency only

---

[1] It should be noted that the definition of "standard" speech is relative across language communities. Any type of non-native speech may actually be "standard" to individuals who have a extensive experience with a given non-native style of speaking. "Standard" in this case refers to the assumption that most monolingual English speakers in the U.S. are unlikely to have extensive experience with non-native accents.

serves as a cue when comprehending speech that is anomalous. The case for each of these possibilities will be discussed further below.

Much work has examined listeners' adaptation to non-native accent (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004), but comparatively little work has examined the influence of non-native speech on comprehension of multi-word utterances. In the EEG literature, several studies have demonstrated that grammatical errors in non-native or non-standard speech elicit a smaller P600 component than grammatical errors in native speech (Caffarra & Martin, 2019; Hanulíková et al., 2012; Seifeldin et al., 2015). The P600 is thought to reflect syntactic repair processes, and so reductions in P600 amplitude suggest listeners may not rely on the standard repair processes for non-native speech. Non-native accent has also been found to influence predictive processing (Lev-Ari, 2015; Romero-Rivas et al., 2016), memory for lexical items (Lev-Ari & Keysar, 2012), reliance on plausibility heuristics (Gibson et al., 2017), processing of scalar implicatures (Fairchild & Papafragou, 2018), and interpretation of ironic statements (Caffarra et al., 2018). These findings suggest in that non-native accent has wide-ranging implications for multi-word processing beyond simply the need to adapt to an unfamiliar accent.

1.2 Adaptation to Non-Native Speech

Listeners might accommodate their processing to non-native speech in a couple of different ways. One possibility is that comprehenders rely on past experience (even if it is limited) or knowledge about characteristics of non-native speakers when processing an utterance from non-native speakers. Listeners may infer the speaker's socio-indexical characteristics (in our case, non-native speaker status) from a small set of phonetic cues (Kleinschmidt, Weatherholtz, & Jaeger, 2018), and then use that inferred identity to update expectancies of other unobserved features that are likely to co-occur with that speaker's identity. This top down speaker identity driven manner of processing can give additional information beyond just the language input itself, such as information about characteristics of the speaker or, often relatedly, the types linguistic input that is upcoming, thus aiding in comprehension. Alternately, comprehenders may not use any top-down knowledge about non-native speakers, and instead process speech from non-native speakers in a way that reflects adaptation to the specifics of the given input. This more bottom-up manner of processing assumes that comprehension can proceed just fine without knowledge about given speakers.

1.3 Top-Down Influences on Comprehension of Non-Native Speech

There are many sources of knowledge listeners might use to help guide the processing of unfamiliar speaker output. Listeners might have statistical knowledge about syntactic error rates (Gibson et al., 2017; Ryskin et al., 2018), knowledge of a language disorder (Arnold et al., 2007), or other types of general knowledge about the language proficiency of non-native speakers. Under the top-down approach to processing non-native speech, comprehension might lean on some representation of speaker proficiency that serves to guide processing of language from future non-native speakers. This approach conceives of language proficiency as a general variable that listeners track and utilize across all utterances of a given speaker. For example, while listening to a low proficiency speaker, listeners would use proficiency information as a processing cue to inform comprehension of every utterance, regardless of whether the utterance was anomalous.

Support for this type of generalization from past speakers to novel speakers has been demonstrated in speech perception. Niedzielski (1999) showed that Detroit citizens perceived different vowel sounds depending on whether they believed the speaker to be from Canada or Detroit (see Hay & Drager, 2010 for a very similar effect). Similarly, Strand, (1999) showed that listeners' perception of the fricatives /s/ and /ʃ/ is dependent on whether listeners believe the speaker to be a male or female. In both cases, these findings demonstrate that knowledge about speaker characteristics can be used to inform subsequent language processing at the phonetic level.

Some findings suggest that speaker characteristics are used as a cue to inform processing of language at other levels as well. In one classic study, participants listened to sentences such as "I might be pregnant because I feel sick" (Van Berkum et al., 2008). The sentences elicited a stronger N400 effect when the semantic of the sentence did not match the characteristics of speaker (e.g. if the speaker of the example sentence was male), suggesting that gender and the characteristics of gender are used to guide sense-making very early in the comprehension process. Consistent with a role for proficiency in guiding comprehension, Arnold et al. (2007) showed that listeners attribute disfluencies as signifiers of unfamiliar upcoming referents, but that this effect is attenuated when the speaker is reported to have object agnosia. Additionally, (Lev-Ari, 2015) showed that listeners engage in more contextually-licensed interpretations (rather than literal interpretations) when listening to instructions from non-native speakers. Both of these studies suggest that online language processing differs based on the identity of the speaker producing the utterance.

A number of findings suggest that listeners infer non-native speech to be more unreliable than native speech, and underspecifies the representation of non-native speech output to compensate. In a study based on the noisy channel model of comprehension, Gibson et al. (2017)

showed that listeners make more non-literal interpretations of implausible utterances when they are spoken by a non-native speaker than a native speaker. The authors interpret this to mean that listeners are relying on a prior belief that includes increased noise rate for non-native speakers. In a different study examining structural priming via grammaticality judgements, listeners rated quasi-grammatical sentences spoken by non-native speakers as less grammatical than the same sentences spoken by native speakers, even when the sentences have the same structure and content (Constantine et al., In Prep). That is, native listeners likely infer that non-native speakers have reduced syntactic proficiency and find the sentences less acceptable as a result. Similarly, listeners rate under-informative scalar implicatures (e.g. “Some people have two nostrils”) as more acceptable when spoken a non-native speaker (Fairchild & Papafragou, 2018). This effect was even present when the sentences were presented textually. Similiarly, the reduced P600 in response to grammatical errors can also be thought of as an increased expectation for unreliability (Caffarra & Martin, 2019; Hanulíková et al., 2012). These findings are all consistent with generalization from past to novel speakers (e.g. top-down modes of processing), in that the effects demonstrated could all plausibly be a result of listeners assuming that non-native speakers are more likely to be unreliable, which changes their judgments about the sentence.

However, all three studies utilize meta-linguistic tasks that do not probe the linguistic representation formed in the process of comprehending non-native speech. As such, it is not known whether listeners use information about speaker proficiency to guide language processing, or whether these effects only occur when asked to make meta-level language judgements. Of the small number of published studies that do examine representation of non-native speech via structural priming (Chun et al., 2016; Chun & Kaan, 2020; Weatherholtz et al., 2014), none manipulate proficiency of the speaker directly. Additionally, all the studies mentioned that use meta-linguistic tasks only examine processing of non-native speech on items which are anomalous in some way: either ungrammatical, implausible, or under-informative, so the question of whether the effects of anomalous input extends to grammatical items is unknown. The current study addresses these gaps by examining the representation of grammatical sentences spoken by native speakers, non-native speakers, and reduced proficiency non-native speakers.

1.4 Expectation Account

One account of the mechanisms underlying the above findings is the *expectation account* proposed by Lev-Ari (2015). This account states that listeners expect lower linguistic proficiency in non-native speakers, and thus adaptively rely less on the bottom-up speech input and more on top-down methods of comprehension, including context and speaker knowledge. Critically, this account predicts that processing is adjusted *from the outset as a function of the comprehender's expectation*, rather than simply as an adjustment during online processing. In a strong version of this account, reduced proficiency should *always* serve as a cue to guide processing, leading to less reliance on bottom-up input regardless of whether the input is anomalous or not. This account also predicts that decreases in proficiency of the speaker should lead to decreases in reliance on the speech input. In the context of structural priming, this would be manifest as decreased repetition of structure as proficiency decreases.

One question is why should this account of non-native language comprehension include processing of structure at all? First, as noted above, structure processing differs when hearing ungrammaticalities in non-native speech than in native speech (Caffarra & Martin, 2019; Hanulíková et al., 2012). a process which may occur as a result of listeners adjusting structural processing at the outset. Second, a number of studies suggest that native listeners assume non-native speakers produce more unreliable structure (Constantine et al., In Prep; Gibson et al., 2017). Relatedly, evidence from language instruction literature suggests that native listeners correct syntactic errors less often than discourse and vocabulary errors in non-native speakers (Chun et al., 1982). These studies suggest that native listeners consider a higher proportion of syntactic anomalies to be a feature of non-native speech. The idea behind the expectation account is anticipate upcoming unreliable cues in order to avoid being misled—if listeners in fact do expect unreliability at the syntactic level, it is plausible under this account that they should adapt their processing at the outset in order to comprehend more efficiently.

1.5 Good-Enough Processing Account

Alternatively, a *good-enough* processing account assumes that listeners optimize comprehension resources to the task goal (Ferreira et al., 2002; Karimi & Ferreira, 2016), and does not give a privileged representation to speaker identity. The good-enough processing perspective emphasizes that the language processor has resource constraints, that structural processing does not always be veridical to the input, that comprehension of meaning can occur via heuristics rather than algorithmic processing routines, and that the end result of comprehension processing only needs to be adequate to accomplish the task goal (Christianson, 2016). Unlike the expectation account, the more general good-enough processing account suggests that listeners do not universally rely on speaker specific information like proficiency; instead, listeners dynamically shift processing to accommodate new linguistic features. Critically, *this account does not make the claim that processing is based on the comprehender's expectations*. This account predicts that listeners *may* use shallow processing to comprehend non-native speech, but *can* process it deeply if prompted (e.g., if a low proficiency non-native speaker requires more resources). This approach suggests that non-native speech is comprehended using the same processes as speech from native speakers, and that proficiency becomes an important factor only when the processor comes across anomalous input.

1.6 Syntactic Processing and Structural Priming

Here, I compare and contrast these two accounts using a structural priming paradigm. *Structural priming* (also referred to as *syntactic priming*) is the tendency to re-use recently encountered syntactic structures (and/or comprehend them more readily), independently of lexical and semantic content (Bock, 1986; Pickering & Ferreira, 2008). Most structural priming studies have examined production-to-production priming (referred hereafter as “PP priming”), while the current study uses a comprehension-to-production priming paradigm (referred hereafter as “CP priming”). The sequence of events in a standard PP priming trial is as follows: participants listen to a prime sentence, repeat the sentence verbatim, and then describe an image (Bock, 1986; Bock & Griffin, 2000). CP priming uses this same sequence of events, but simply omits the repetition of the prime sentence (Bock et al., 2007; Jacobs et al., 2019). A recent meta-analysis suggests that priming effect sizes in CP priming paradigms are comparable to those in more traditional PP paradigms (Mahowald et al., 2016). For the current study, CP priming has several desirable properties: (a) it ensures that any reduction in the priming effect is due entirely to participants’ comprehension of the non-native speaker and not errors or uncertainty generated from their own repetition of the primes, (b) it removes the possibility that any differences between Native and

Non-native speakers are obscured by self-priming (again, due to the repetition of the prime), and (c) this paradigm places no communicative pressure on participants to moderate their alignment to the speakers for communicative benefit; thus, any differences between the Native and Non-native speaker should be free of effects driven by this pressure.

1.7 Current Study

To examine the influence of non-native talker identity on structural priming, the current set of studies employed the comprehension-to-production priming paradigm in two Qualtrics experiments modeled after Bock et al. (2007). In Experiment 1, listeners were exposed to a non-native and a native speaker in separate blocks, and priming in response to both speakers was measured. Here, in keeping with the expectation account, I expected that listeners will be less likely to repeat syntactic structures spoken by the non-native speaker than by the native speaker. In Experiment 2, listeners were exposed to two non-native speakers, but one non-native speaker will be less proficient, as evidenced by syntactic errors in the filler sentences. This served as a direct test of the assumption of the expectation account—namely, that listeners use information about speaker proficiency to guide their processing of non-native speech. Here, I hypothesized that listeners will be less likely to repeat syntactic structures spoken by the less proficient non-native speaker than by the non-native speaker.

2.0 Experiment 1

2.1 Method

2.1.1 Participants and Sample Size

I sampled participants (*N* = 150) from the University of Pittsburgh subject pool, with the goal of a final target sample size of at least 128 participants. Although asking for native English speakers only, many non-native English speakers participated. I defined native English speakers as individuals who indicate that English is the first language they learned and that they did not learn any other language concurrently. Participants were also required to be older than age 18 and have normal or corrected-to-normal hearing. After excluding participants who were non-native speakers, participants who had low comprehension accuracy, and participants who produced a low number of usable picture descriptions, I was left with our target sample of 128 participants.

The target sample of 128 participants (and 48 items) was derived from power estimates in a recent meta-analysis of structural priming effects (Mahowald et al., 2016). The authors simulated data from structural priming studies that examined interaction effects for varied combinations of subjects and items. Assuming a structural priming effect of 0.51 log-odds ratio (the estimate of the average lexically independent priming effect derived from the meta-analysis) and a moderately sized interaction coefficient of 0.5, they estimate that 128 subjects and 48 items results in 86% power to detect the interaction. For a small interaction coefficient of 0.2, even simulations with 400 subjects and 72 items did not yield 80% power.

The true interaction coefficient of structural priming with speaker accent is not known a priori. However, we might reasonably predict that it would be at least moderately sized, given that, relative to other possible moderators of structural priming, the non-native accent manipulation is easy to detect (Constantine et al., In Prep). Constantine et al. (In Prep) found relatively large differences in participants' ratings of how well the native and non-native speakers know English on a scale of 1-7 ($M_{native} = 6.53$, $M_{non\text{-}native} = 3.51$, $d = 2.86$). Thus, if non-native accent does not moderate structural priming, it is almost certainly not due to the fact that the manipulation is undetectable. Further, powering the experiment to detect even a small interaction is not feasible because the power analysis conducted by Mahowald et al. (2016) suggests that such an interaction would be virtually undetectable without an impractically large sample—and if the interaction is too small to detect with the current design, it may be of comparatively little theoretical or practical importance.

2.1.2 Materials

2.1.2.1 Stimulus Sentences

This study used a set of stimulus sentences combined and adapted from Bock and Griffin (2000) and Pickering & Branigan, (1998), the full list for which can be found in Appendix A. Participants were exposed to 48 dative prime sentences, such as (1a) or (1b). Dative structures can be formulated in one of two alternations: either the prepositional-object structure (PO, *1a*) or the semantically equivalent double-object structure (DO, *1b*). In a context with no preceding discourse, both versions of the dative alternation are semantically similar, making the dative an ideal alternation to study syntactic processing because there are few semantic constraints on selection. In addition, 96 filler sentences made up of transitive, intransitive, and predicate adjective structures were interspersed around the experimental prime-target pairs sentences.

(1a) Dative {PO}:	The man read a story to the old woman.
(1b) Dative {DO}:	The man read the old woman a story.
(2a) Filler {Transitive}:	The janitor cleans the floors daily.
(3a) Filler {Intransitive}:	The resourceful campers survived.
(3b) Filler {Predicate Adjective}:	The college textbooks were expensive.

2.1.2.2 Stimulus Images

Experimental images consisted of 48 line drawings, taken primarily from the list of images used in Bock and Griffin (2000). Not all images used in that study could be recovered, so additional images were added to ensure there are 48. These new images were normed (described below) to ensure they elicited dative structures reliably. The dative-eliciting images (Figure 1) consist of an agent passing an object to a recipient and can be reasonably described using a PO or a DO structure. Additionally, 96 filler images were designed to elicit intransitive constructions.

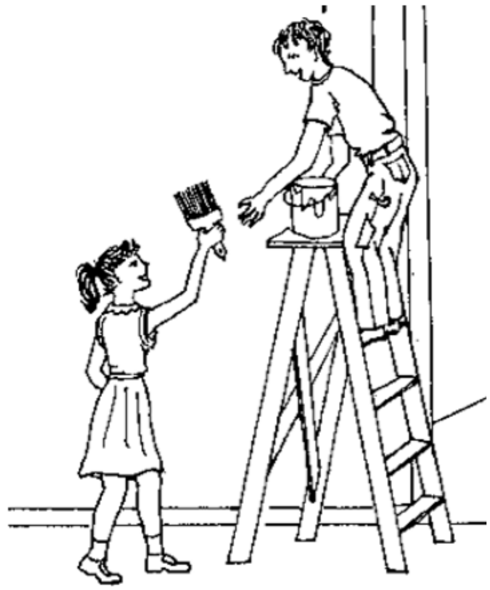

Figure 1: Dative-eliciting image.

Experimental images were selected from a broader set of 64 images based on the results of a norming study. In the norming study, participants were instructed to type a description of the images using one sentence. The only instructions they received was not to use pronouns to describe subjects and not to describe inferred events that are not explicitly depicted in the image. Based on the results of this norming study, 48 dative-eliciting images that were most likely to be described with a dative construction were selected for use in the experiment.

Experimental images were randomly paired with dative primes to create 48 prime-target pairs. I ensured that the content of the images had no clear relationship with the words or event in described in the prime sentence beyond a depiction of one item being passed between two entities.

2.1.2.3 Sentence Recordings

All sentences were recorded by age and gender-matched speakers. For Experiment 1, both speakers were females in their early 20's. The non-native English speaker's first language was

Mandarin Chinese, and the native English speaker was monolingual. In Experiment 2, I added an additional non-native speaker (female, early 20's) whose first language was Mandarin Chinese. Her accent was less pronounced, and so she "put on" a bit more of her accent during recording to ensure that her accent was detectable. When recording the sentences, measures were taken to ensure that the non-native speakers were intelligible while still retaining a readily detectable accent. Specifically, I worked with the non-native speakers to address questions about pronunciation and provide feedback on recordings where the identity of a word was particularly unclear. To record the sentences, the speakers used Praat recording software and a Shure microphone or their internal MacBook microphones. All speakers recorded each sentence twice, and then the cleanest version of each sentence (most intelligible and free of audio artifacts) was selected. All recordings were normalized to 70 dB.

2.1.2.4 Comprehension Questions

Comprehension was assessed on 48 filler sentences to ensure attention remained on-task, as well as to ensure that participants did indeed listen to the audio. Questions were two-answer forced choice, wherein both responses were approximately equiprobable if the previous sentence had not been heard. For example, if the filler sentence was "The truck ran the bicycle off the road", the comprehension question was "Who was run off the road? A = motorcycle, B = bicycle." Thus, participants needed to hear and attend to the preceding sentence in order to accurately answer the question.

2.1.2.5 Speaker Perception Questionnaire

Participants were asked a number of questions probing their perception of each speaker. These questions served as a manipulation check and to examine whether participants' explicit perceptions of each speaker might predict their propensity to be primed from that speaker. For each speaker, I examined participants' perceptions of each speaker's level of education, intelligence, kindness, likability, English proficiency, similarity to the participant, height, and wealth level. Each item was rated on a 7-point Likert scale, ranging from *1 = Not at all* to *7 = Extremely*. The middle value of 4 was "Average". Before answering the speaker perception questions, an audio clip was played to ensure that participants knew which speaker they were assessing.

2.1.2.6 Accent Exposure Questionnaire

Participants' language background and exposure to accents was assessed via a questionnaire adapted from the Bilingual Language Profile (BLP) questionnaire (Birdsong et al., 2012). Participants were asked if English was their first language, which other languages they learned concurrently, the age at which they began learning English, whether they spoke English at a native level of fluency. They were also asked about the length of time (in years) they have spent in various environments where English, Chinese, and Other languages are spoken. The environments included school, country, family, group of friends, and work environment. I probed Chinese exposure specifically, because past research suggests that familiarity with an accent may influence priming (Chun et al., 2016). Participants were also asked to estimate the proportion of their weekly conversations that are with people with an American English accent, Other English accent, Chinese accent, or some Other accent. I asked for a global estimate, as well as estimates within the social domains of family, friends, co-workers and class.

2.1.3 Experimental Design

This experiment used a 2 (Speaker: native vs. non-native) x 2 (Structure: dative vs. transitive) factorial within-subjects design. The task was adapted from Bock et al.'s (2007) comprehension-to-production paradigm wherein participants were first exposed to auditory prime sentences and then asked to describe target images. In each trial, participants heard dative prime sentences such as (1a, PO) or (1b, DO). They were then asked to describe corresponding dative-eliciting images (Figure 1). These images can be equivalently described using either alternation, but the alternation of the prime structure will bias participants to re-use that structure in their description, regardless of the image content.

An example of a typical trial is shown in Figure 2. Each trial had a prime-target pair embedded within two randomly ordered filler sentences and one filler image. Comprehension questions occurred after one filler sentence in each trial. Participants typed their descriptions of all filler and target images. Although Bock and Griffin (2000) recorded participants' picture descriptions auditorily, spoken-to-written modality priming paradigms have been shown to be sensitive to priming in both transient priming (Cleland & Pickering, 2006) and cumulative priming experiments (Jacobs et al., 2019).

Filler: The fashion model wore the designer's favorite dress.

Question: What was worn? A = Dress B = Skirt

Prime: The graduate students are baking cake for the professors

Target Image:

The girl is handing the paintbrush to the boy.

Filler Image:

The old man is sleeping.

Filler: A janitor cleans the floors daily

Figure 2: Example of one trial. Each trial contains two filler sentences, one filler image, one prime sentence, and one target image. Participants are blind to boundaries between trials.

Participants completed one block of 24 experimental prime-target pairs from a non-native English speaker and one block of 24 prime-target pairs from a native English speaker (Figure 2). Assignment of primes to speaker, block number, and syntactic alternation was counterbalanced across 4 experimental lists.

To disguise that the purpose of the study was to measure language production, participants were told that they were participating in a recognition memory task (a commonly used practice dating back to Bock, (1986)). This was not entirely a ruse, as I did assess recognition memory for 24 filler items at the end of the main task to confirm Lev-Ari and Keysar's (2012) finding that memory for non-native speech is less detailed. To do so, 10 of the repeat filler sentences were an exact match of the original. The other 10 repeat sentences were foil sentences that are similar to the original but with slight rewordings (e.g. “The professor's office was messy” became “The professor’s office was dirty.”). The prediction is that participants should be more likely to falsely indicate that the foil sentences are an exact match in the non-native condition than in the native condition, indicating less-detailed memory representations.

2.1.4 Procedure

Written instructions at the beginning of the Qualtrics experiment explained to participants that they would be completing a recognition memory task wherein they will listen to sentences, view images, and be tested on their memory for the sentences after viewing all of the stimuli. Participants were instructed to complete the experiment in a setting that minimizes distractions and to ensure that they will have 1 hour to complete the study. The experimental list that each participant was exposed to was pseudo-randomized within Qualtrics such that equal numbers of participants were exposed to each list. After completing the experimental portion of the study, participants completed the speaker perception questionnaire and the accent exposure questionnaire. Upon completion, participants were shown a debriefing statement explaining the purpose of the study.

2.1.4.1 Image Description Coding

Participants' written descriptions of the target images were coded offline as either PO, DO, or Other, using a three-step coding process. First, to speed the coding by reducing the number of labels that need to be typed, picture descriptions were automatically coded as PO, DO, and Other by a random forest classifier that was trained on a hand-coded subset of ~600 picture descriptions. The classifier had a lower test accuracy of about 75% on the remaining (out-of-sample) structures. Still, this accuracy is far greater than chance (33%), and thus greatly facilitated coding time.

Second, trained coders compared the classifier generated labels to the picture descriptions, correcting the results of the classifier where appropriate, by applying criteria adapted from Bock and Griffin (2000). Dative *prepositional-object* sentences were required to have three noun phrases depicting the following: a subject phrase that depicts the subject of the action, a prepositional-object phrase beginning with *to* or *for* depicting the goal of the action, and an object phrase representing the theme (the item being transferred). These sentences must also have a double object equivalent. *Double-object* sentences were required to have the same three noun phrases as above and a prepositional-object equivalent, but without the requirement for a preposition. All coders were blind to the speaker condition and the structure of the preceding prime.

Last, an additional coder recoded a random subset of 10% of the trials to assess inter-rater reliability. Prior to coding, it was decided that if reliability is less than 90%, the full dataset would be double coded and disagreements would be arbitrated by a third coder. However, reliability was high, with a Cohen's kappa of .95 and .97 respectively for Experiments 1 and 2.

2.2 Results

2.2.1 Data Analysis

The primary hypothesis in both experiments was tested using logistic linear mixed effects regression (LMER) models. All data were analyzed using R statistical software, and the LMER models were built using the lme4 package (Bates et al., 2015). In all LMER models, the maximal random effects structure was tested (Matuschek et al., 2017) and then, if necessary, reduced until the model converges.

2.2.2 Priming Results

Of the 6000 picture descriptions collected, 34.6% (2,073) were coded as DO, 37% (2,222) were coded as PO, and 28.4% (1,705) were coded as Other. This rate of Other descriptions is typical during open-ended picture description tasks such as this, particularly when no lexical constraints are placed on participants' responses. Importantly, the percentage of Other responses was similar for both Speaker conditions (Native = 13.4%; Non-native = 15%), suggesting that participants' rate of invalid responses is not responsible for priming differences between the two conditions. Following the analytic procedure of previous experiments (e.g. Bock & Griffin, 2000), the Other descriptions were dropped from analysis, leaving 4,295 valid observations that were either a PO or DO response.

Priming of target descriptions was examined in two parts. First, I examined the critical Speaker x Prime effect in the full dataset. And, to preview a bit, I found that the effect of Speaker x Prime differed across blocks, so I then examined the Speaker x Prime effect in Block 1 and Block

2 separately. I parameterized my results in terms of the rate of producing the normatively less frequent structure; that is, the PO dative (Bock & Griffin, 2000). Note, however, that treating the DO dative as the target structure would only reverse the coefficient sign but otherwise result in an equivalent analysis, since the dataset contained only DO and PO productions. The probability that participants produce a PO for the image descriptions was analyzed using a model that includes fixed effects of Speaker Type, Prime Structure, and their interaction. This fixed effect structure was used in all models. The random effects structure varied and will be reported alongside each model below. Speaker and Prime were effects-coded in all models (Native = -0.5, Non-native = 0.5; DO = -0.5, PO = 0.5).

2.2.2.1 Overall Analysis

For the overall effect of priming, the model had a near-maximal random effects structure containing random intercepts for Participant and Item, with by-Participant slopes for Speaker and Prime and a by-Item slope for Speaker. In this model, there was a significant main effect of Prime ($B = 0.53$, SE = 0.09, $z = 5.392$, $p < .001$), with the odds ratio of producing a PO after a PO prime being 1.68 times greater than after a DO prime (Table 1). This replicates the classical priming effect, wherein participants are more likely to produce a PO description after a PO prime sentence than after a DO prime sentence. Additionally, the size of the priming effect observed here is well in line with the Mahowald et al. (2016) meta-analytic estimate of an odds-ratio of 1.67. There was no main effect of Speaker ($B = -0.11$, SE = 0.12, $z = -0.94$, $p = .345$), which is unsurprising because there is no a priori reason to expect different production of POs by speaker alone. However, the critical interaction between Speaker and Prime was also non-significant ($B = 0.03$, SE = 0.15, $z = -0.22$, $p = .825$).

Table 1: Model output for the overall analysis

	B	SE	z	*p*
(Intercept)	0.24	0.22	1.11	.27
Prime	0.53	0.09	5.39	<.001
Speaker	-0.11	0.12	-0.94	.35
Prime:Speaker	0.03	0.15	-0.22	.83

Due to the blocked design, the non-significant Speaker x Prime interaction in Block 1 could be due to order effects wherein processing in the first half of the experiment influences processing in the second half. If that were the case, we might see a Speaker x Prime interaction in the first half trials that reduces or disappears in later trials. To examine this data further, I added a fixed effect of Block to the model which I allowed to interact with both Prime and Speaker. The main effect of Prime stayed present and there was a significant three way interaction between Prime, Speaker and Block (B = 1.56, SE = 0.37, z = 4.126, p = <.001). This three-way interaction suggested that there is a relatively large interaction taking place across the blocks that warrants further analysis.

2.2.2.2 By-Block Analysis

To examine this three-way interaction, I constructed separate models for Block 1 and Block 2. In both models, the random effects structure contained Participant and Item intercepts, with a by-Participant slope of Prime and a by-Item slope of Speaker. Note that analysis on this subset treats the Speaker manipulation as a between-subjects factor.

In Block 1, I found a significant effect of Prime (Table 2) that is similar in size to the overall effect (B = .47, odds-ratio = 1.59). There was no main effect of Speaker; however, there was a significant negative interaction between Speaker and Prime, indicating that the priming effect in the Non-native speaker condition is smaller than the priming effect in the Native speaker condition. This effect can be seen in Figure 3. This reduction in priming from the Non-native speaker is consistent with the expectation account, which predicts that participants rely less on bottom-up structural information during comprehension of Non-native speakers.

Table 2: Model output for Block 1 and Block 2 in Experiment 1.

		B	SE	z	*p*
Block 1	(Intercept)	0.24	0.22	1.11	.27
	Prime	0.47	0.11	4.25	<.001
	Speaker	0.25	0.32	0.80	.42
	Prime:Speaker	-0.67	0.22	-3.01	<.01
Block 2	(Intercept)	0.04	0.19	0.23	.82
	Prime	0.57	0.12	4.82	<.001
	Speaker	-0.41	0.31	-1.32	.19
	Prime:Speaker	0.72	0.24	3.06	<.01

However, I found different results in Block 2. The Block 2 results do show a significant main effect of Prime (Table 2). Interestingly, however, the Speaker x Prime interaction is significantly *positive*, suggesting that the Non-native speaker results in *more* structural priming than the Native speaker. I discuss possible reasons for this below, but it is important to note that, due to the counterbalanced design, this means that the group of participants who were strongly primed by the Native speaker in Block 1 also showed strong priming when exposed to the Non-native speaker in Block 2, whereas the group of participants who were exposed to the Non-native speaker in Block 1 showed a small to non-existent priming effect throughout.

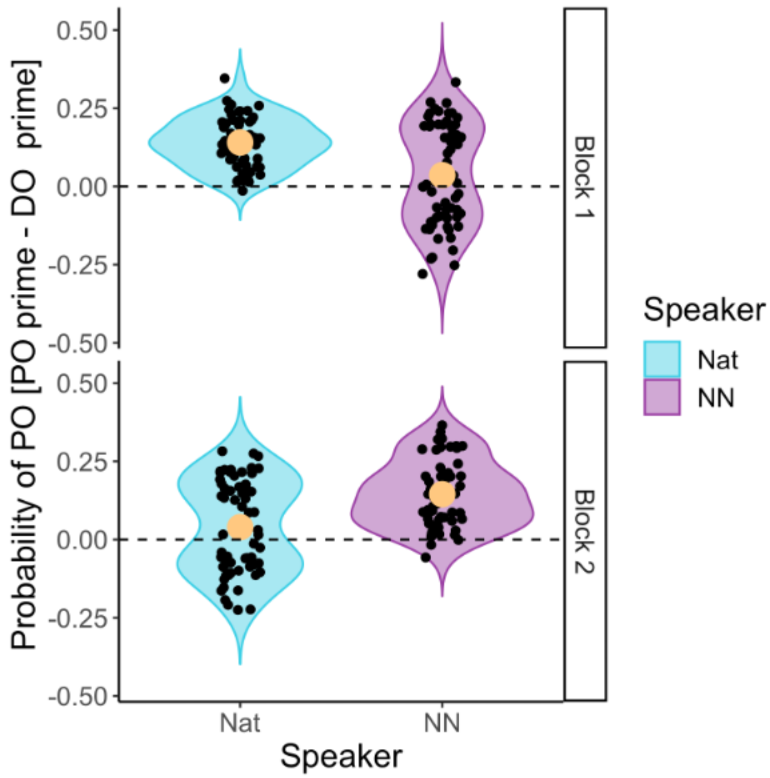

Figure 3: Black points represent model-estimated individual means (transformed from logits to probability). The cream colored points represents the group mean for each cell and are not direct model estimates. Values greater than zero reflect positive priming effects (increased probability of producing PO after PO prime).

2.2.3 Comprehension Question Accuracy

Comprehension questions were asked after 48 filler sentences (24 for each speaker) to ensure participants' attention remained on the task. Overall comprehension accuracy was extremely high (*M* = 99.02% across conditions) and did not differ by speaker (Native = 98.9%, Non-native = 99.2%). I expected at least a marginal decrease in accuracy for the non-native speaker, possibly reflecting difficulty understanding the non-native speaker or reduced attention to the non-native speaker condition. The fact that there are no differences here suggests that any the differences in priming observed above cannot easily be attributed to intelligibility or attention.

2.2.4 Speaker Perception Questionnaire

The primary speaker perception question of interest is whether participants rated the Non-native speaker as less proficient than the Native speaker. This measure serves as a manipulation check to ensure that participants did in fact perceive the speaker accent manipulation. The Native speaker's English proficiency was rated higher on average ($M = 6.19$, $SD = 1.02$) than the Non-native speaker's ($M = 4.45$, $SD = 1.49$), $t(123) = -12.243$, $p < .001$. This suggests that participants perceived a difference between the Native and Non-native speaker, and that they believed that the Non-native speaker was less proficient than the Native speaker.

For the other speaker perception items, participants generally rated the Native and Non-native speaker equivalently (Figure 4). One notable difference is the similarity item, where participants rated the Native speaker as more similar to them ($M = 5.18$, $SD = 1.26$) than the Non-native speaker (M = 2.88, SD = 1.4), $t(123) = -12.797$, $p < .001$. This suggests that listeners are picking up on the fact that the speakers are different on the native vs. non-native dimension and not on other spurious differences such as intelligence or education.

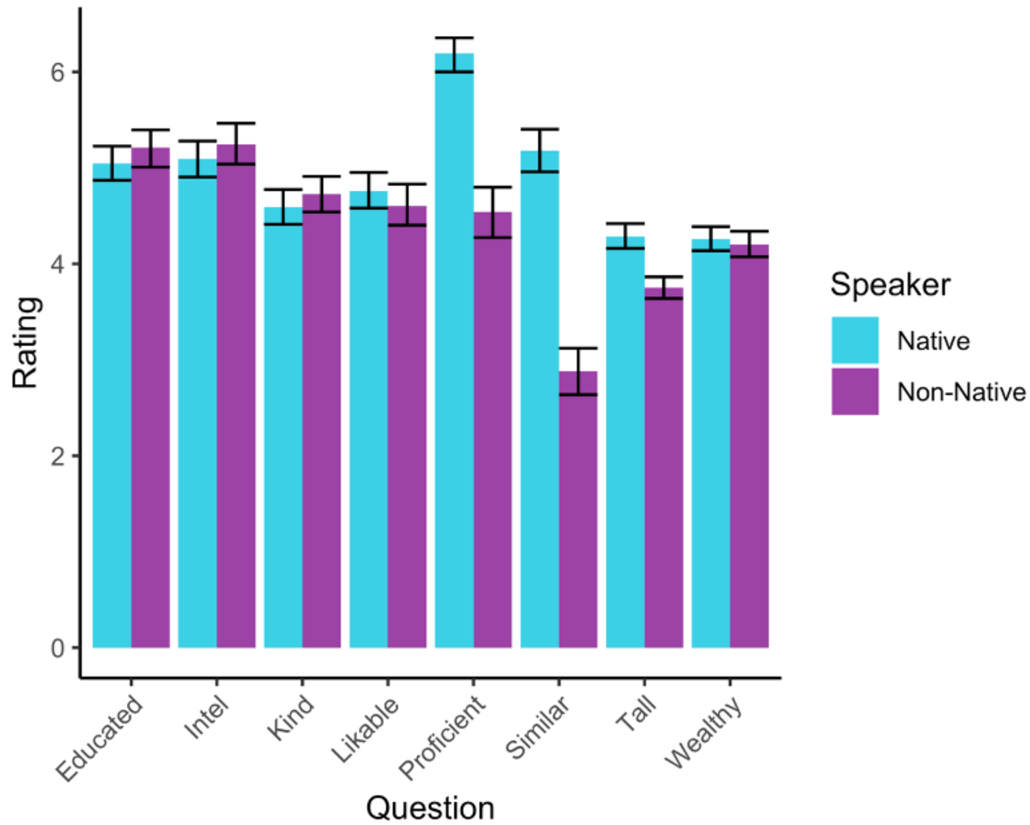

Figure 4: Speaker perception ratings for each item. Error bars represent bootstrapped 95% confidence intervals.

2.2.5 Recognition Memory Results

Overall, participants struggled on the recognition memory task, with an average accuracy of 68.5% (*SD* = 10.1). (where chance level responding would lead to 50% accuracy). Average recognition accuracy for the Non-native speaker (*M* = 69.1%, *SD* = 12.8%) was slightly higher than for the Native speaker (*M* = 67.8%, *SD* = 12.5%). Accuracy varied substantially across participants, with some participants scoring as low as 25% or as high as 100%.

To probe this data further, I analyzed recognition accuracy in a signal detection framework, wherein responses are parameterized such that the response given (i.e., *Old* or *New*) is the outcome variable, and the correct response is included as (effects-coded) fixed effect. This allows for an empirical and theoretical dissociation between a response bias (an overall tendency to respond either "New" or "Old") and the sensitivity of responses towards the true correct response (responding *Old* more when the sentence was veridically seen before). I also added a fixed effect of speaker (effects coded) and allowed it to interact with the correct response to test whether people were more likely to respond correctly conditional on Speaker. The random effects structure included random intercepts for Subject and Item, with a by-Subject slope for Correct and a by-Item slope for Speaker.

The full model results are in Table 3. Overall, there was no significant response bias towards responding either "New" or "Old", as indicated by the non-significant intercept parameter. There was a main effect of the Correct response ($B = 1.69$, $p < .001$), such that the odds of responding that an item was "Old" when it was truly old were 5.31 times higher than responding "New". There was not a significant main effect of Speaker ($B = .20$, $p = .18$), meaning that there were no significant differences in response bias between speakers. There was no significant interaction between the Correct response and Speaker condition ($B = -0.14$, $p = .38$), suggesting that there was no detectable effect of Speaker on responding correctly. It is difficult to make a strong conclusion from null results. But given the relatively large sample of individuals and number of items I would tentatively conclude that if Speaker does influence correct responding on the memory task, the effect is likely to be so small that it is not detectable here. This could be in part because the picture description task takes longer than a typical memory task, and thus the latency over which sentences must be remembered is relatively long.

Table 3: Results of signal detection regression model.

	B	SE	z	*p*
(Intercept)	-0.17	0.13	-1.29	.20
Correct	1.69	0.20	8.59	<.001
Speaker	0.20	0.15	1.32	.18
Correct:Speaker	-0.02	0.27	-0.06	.95

2.2.6 Accent Exposure Results

To assess the influence of prior exposure to accents on language processing in this experiment, I asked participants to estimate their exposure to accents (American English, Chinese, Other English, or Other Accent) on a weekly basis. Participants supplied global estimates and estimates for specific domains (Class, Work, Friends, Family). The distributions of estimates were very similar for across all domains, so only the global estimates will be presented for simplicity.

Somewhat unsurprisingly, participants had very little exposure to other accents overall (Table 3). Participants estimated on average that 90% of their exposure to speech was with people who spoke with an American English accent, while exposure to Chinese, Other English, and Other accents was generally quite low (all less than 5% on average). The distributions were heavily skewed, so median values suggest an even more extreme lack of exposure to varied accents throughout the week.

Table 4: Global estimates of exposure to accent types on a weekly basis.

	Mean	Median	SD
American English	92%	100%	15.0%
Chinese	1.9%	0%	5.68%
Other English	2.11%	0%	6.9%
Other Accents	3.96%	0%	9.57%

To test for the influence of accent exposure on priming, I added each global accent exposure measure to the LMER models tested above (the model in the overall analysis as well as both the models in the by-block analysis), and allowed it to interact with each fixed effect. This did not yield any meaningful relationships between prior accent exposure and structural priming, even after applying data transformations to account for skewness. To examine accent exposure from another angle, I computed entropy scores for each participant across the global measure of accent exposure, as well as within each subdomain[2]. I added these scores to the same models of priming, but found no relationship between language entropy and structural priming.

[2] Higher entropy scores represent more consistency in exposure to all four accent types, while low scores represent less consistency in accent exposure. For example, a participant who is exposed to 25% American English, 25% Chinese, 25% Other English, and 25% Other Accents would have a high entropy score, relative to an individual with 90% American English and 10% Chinese. Intuitively, we might expect that participants with higher entropy may be able to comprehend accented speech more effectively; possibly due to an increase in cognitive control. However, evidence for this is not present in this dataset.

2.3 Discussion

Experiment 1 was designed to examine the way that native speakers process language from a non-native speaker. I found that the effect of accent on priming was not uniform across the entirety of the experiment. In the first block, participants in the non-native condition showed decreased priming relative to those in the Native condition. In the second block, this pattern was reversed such that participants in the Native condition now showed reduced priming relative to participants in the non-native condition. Due to the counterbalanced design, this means that participants who were exposed to the non-native speaker first were primed less by the Native speaker in the second half, in an interesting carry-over effect. Conversely, the participants who were exposed to the Native speaker first showed a similarly high priming effect in the second half of the experiment. While I did not expect this carry-over effect, it is a striking feature of the data and will bear heavily on our discussion. However, I think of carry-over effect in the second block as a second-order influence of Speaker, whereas the results of the first block speak most directly to the first-order influence of speaker. As such, all discussion below will be in regards to the Speaker effect in the first block unless otherwise specified.

Lev-Ari's (2015) expectation account can explain the pattern of results in the first block as a reduced reliance on the features of the input in the non-native condition, relative to the Native condition. This account could potentially explain the carryover effect if I assume that the processing in the first half can lead to a global "task processing mode", wherein listeners process language in the task shallowly or deeply (depending on the accent they are exposed to). Then, I can assume whatever mode participants were engaged in during the first half of the experiment was carried over into the second half of the experiment. However, the idea of a processing mode carrying over to new speakers within a task, indiscriminately of accent, is not a feature of the current expectation account. This fact also undercuts one assumption of the expectation account—namely, that the shallower mode of processing is an adaptation to facilitate comprehension of an unreliable speaker. If this processing strategy is not deployed in a way that is specific to unreliable speakers, then it suggests that the expectation account is not a good explanation for the data.

A good-enough processing explanation may fit the data better. The good-enough processing perspective emphasizes that the language processor has resource constraints, that structural processing is not always veridical to the input, that comprehension can occur via heuristics rather than algorithmic processing routines, and that the end result of comprehension processing only needs to be adequate to accomplish the task goal (Ferreira et al., 2002). Often the goal is conversation, but the goal can also be specific to a given task. Under this perspective, the explanation for the observed results can be thought of in this way: the reduced priming effect from the non-native speaker in the first block occurs as a result of participants engaging in shallow heuristic-based processing rather than full algorithmic structural analysis. This account also explains why comprehension accuracy could be high across all conditions despite reduced priming from the non-native speaker. That is, comprehension processing is optimized towards the goal of getting surface level information that is necessary for the task, not for deep comprehension. Although not tested in Experiment 2, one prediction of this view is that more difficult comprehension questions might reduce the Speaker x Prime interaction.

This approach may help explain the carryover effect as well. Assuming the good-enough processing explanation is correct, the carryover effect may occur as a consequence of the process of optimizing language processing to the task at hand. For the Non-Native—Native ordered participants, shallow processing was entirely adequate to accomplish processing in the first half, so it would definitely be adequate to accomplish processing in the second half. For the Native—

Non-Native participants, deeper processing worked well in the first half because they did not have the resource constraints of accent decoding. As a result, they were able to learn the types of structures present in the task—only about 5 in total!—and then use this knowledge to more efficiently process the structures spoken by the non-native speakers in the second half.

3.0 Experiment 2

Even if good-enough processing explains these data, we still do not know exactly what properties of the stimuli cued participants to engage in good-enough processing. One possibility is that participants avoided a full algorithmic analysis because they knew that the non-native speaker is less proficient (and therefore less reliable) than the Native speaker. This explanation is in essence a combination of the good-enough and expectation accounts. Another possibility is a pure good-enough processing account: Because processing resources are devoted to speech processing when listening to an accented but otherwise proficient non-native speaker, fewer resources are devoted to algorithmic (and costly) structural processing. This explanation emphasizes that resources are optimized to the task at hand, and they can be allocated to a greater or lesser extent according to the specifics of the task.

These two accounts can be compared in Experiment 2. Experiment 2 will allow for a direct test of the hybrid "good enough" and expectation account outlined above. While Experiment 1 compared comprehension-to-production priming from a native and non-native speaker, Experiment 2 instead compares two types of non-native speakers: a more proficient non-native speaker and a less-proficient non-native speaker. The less proficient non-native speaker was made to produce syntactic errors in 30% of the filler sentences. This manipulation allowed us to examine whether comprehenders use information about Non-native speaker proficiency specifically. (e.g. probability of producing syntactic error on a given trial) to moderate the degree to which they are primed by the Non-native speaker. The expectation-based account predicts that reduced proficiency should obligatorily lead to less reliance on bottom-up input because that input is less reliable. On the other hand, the good-enough processing view emphasizes that listeners may typically rely on shallow heuristic-based processing to comprehend non-native speech, but *can* process it deeply if prompted (e.g., if a low proficiency non-native speaker requires more processing resources).

3.1 Method

3.1.1 Materials

All materials from Experiment 1 are the same for Experiment 2, except for the addition of a second non-native speaker and ungrammatical filler sentences. These changes are outlined in the section below.

3.1.1.1 Speaker Recordings

Both of the speakers for this experiment were females in their early 20s whose native language is Mandarin Chinese. Speaker 1 was the same non-native speaker from Experiment 1, and Speaker 2 was a new non-native speaker. To create the erroneous filler sentences, I introduced morphosyntactic errors on ~30% of the filler sentences from Experiment 1. Each erroneous sentence contained two or three morphosyntactic errors: the replacement of a gerund form with the bare lemma and a subject-determiner number agreement error (examples of each in 2a and 2b), and an additional bare lemma in the sentential complement of sentences like 2b. I used multiple errors to increase the probability that participants detected the errors, since a single small morphosyntactic error might not be reliably detected by native listeners. None of the experimental prime sentences contained such errors.

(2a) The janitor **[is clean] [a floors]** daily.

(2b) The union leader **[is assist] [a workers]** in **[organize]** the strike.

3.1.2 Participants

As in Experiment 1, the target sample size was 128 participants. I again collected data from 150 participants in total, but after excluding participants who were non-native speakers, participants who had low comprehension accuracy (as per the Experiment 1 criterion), and participants who produced a low number of usable picture descriptions, I was left with 114 participants. Because the interaction in Experiment 1 was larger than anticipated, this somewhat below target sample size will not reduce statistical power unless the interaction in Experiment 2 is significantly smaller.

3.1.3 Study Design

The design was identical to Experiment 1, except both Speaker conditions were non-native speakers. That is, participants were exposed to the more proficient non-native speaker in one block and the less proficient non-native speaker in the other block. However, in order to ensure that the priming effect differences cannot simply be attributed to intrinsic characteristics of the two speakers (e.g., differences in speech pattern or accent) rather than the proficiency manipulation per se, the speaker that was assigned to be more or less proficient (i.e., whose input contained the morphosyntactic errors) was counterbalanced across lists.

During piloting, a small number of participants indicated that they could not tell the difference between the two speakers. To ensure that participants knew the speakers were in fact different people, I created fictional biographical details for each speaker and then displayed them at the beginning of each block. The biographical details included their first name, home city in China, career goal, and hobby, which were different for both speakers. The exact text of the bios can be found in Appendix B.

3.1.4 Procedure

The only change to the procedure is that participants were given a slightly different cover story. Although I kept the memory task cover, I wanted to ensure that participants inferred that the syntactic errors were due to the non-native speakers' proficiency, and not something that was being directly manipulated by the experimenters. As such, participants were told that the study was designed to test memory for naturalistic sentences. Participants were led to believe that the sentences they were hearing were from other participants who had been assigned to the "Recording" condition, while they themselves had been assigned to the "Listening" condition. To increase the credibility of the existence of a "Recording" condition, participants were shown the following message at the beginning of each block "Please wait while we randomly select the speaker. It may take a few moments to prepare the audio files. The arrow button will appear as soon as the audio is ready." This was accompanied by a delay of 15 seconds before participants were allowed to proceed, in order to give the impression that a speech processing program was preparing and loading audio files for the participant. The rest of the experimental procedure is identical to Experiment 1.

3.2 Results

3.2.1 Data Analysis

All data analysis procedures were identical to Experiment 1.

3.2.2 Priming Results

Of the 5,472 picture descriptions, 41.8% (2,288) were coded as DO, 31.9% (1,745) were coded as PO, and 26.3% (1,439) were coded as Other. The Other descriptions were dropped from analysis, leaving 4,033 valid PO or DO descriptions. As in Experiment 1, the percentage of Other responses was similar for both Speaker conditions (More Proficient = 13.4%; Less Proficient = 12.9%).

Priming of target descriptions was again examined both across the entire dataset, and then within Block 1 and Block 2 separately. The probability that participants produced a PO for the image descriptions was analyzed using the same model structure from Experiment 1, with fixed effects of Speaker Type, Prime Structure, and their interaction. The random effects structure varied across analyses to facilitate model convergence or accommodate the data structure and will be reported alongside each model below.

3.2.2.1 Overall Model

For the overall effect of priming, the model had a near-maximal random-effects structure containing random intercepts for Participant and Item and by-Participant and by-Item slopes for Speaker and Prime. In this model, there was a significant main effect of Prime (B = 0.67, SE =

0.09, $z = 7.26$, $p < .001$), such that the odds of producing a PO after a PO prime were 1.99 times greater than after a DO prime (Table 5). There was a marginal main effect of Speaker ($B = -0.18$, $SE = 0.1$, $z = -1.73$, $p = .09$). This suggests there was a slight tendency for participants in the Less Proficient speaker condition to produce fewer POs overall than participants in the More Proficient speaker condition. The critical Prime x Speaker interaction was not significant ($B = 0.23$, $SE = 0.15$, $z = 1.47$, $p = .14$), suggesting that, overall, participants' priming was not influenced by the speaker proficiency manipulation. To examine the influence of block order, I added a Block factor that fully interacted with both fixed effects. As in Experiment 1, there was a significant three-way interaction between Speaker, Prime, and Block ($B = -1.21$, $SE = .31$, $z = -3.80$, $p < .001$).

Table 5: Model Output for overall analysis.

	Beta	SE	z	p
(Intercept)	-0.44	0.18	-2.43	<.05
Prime	0.67	0.09	7.26	<.001
SpeakProf	-0.18	0.10	-1.73	.08
Prime:SpeakProf	0.24	0.16	1.53	.12

3.2.2.2 By-Block Analysis

However, in Experiment 1, we saw differential effects of our manipulation blocks—is the same true here? In Block 1, I found a significant effect of Prime (Table 6) that is similar in size to the overall effect ($B = .63$, odds-ratio = 1.88, $p < .001$). There was no main effect of Speaker. However, there was a significant *positive* interaction between Speaker and Prime, indicating that priming effect in the Less Proficient speaker condition is larger than the priming effect in the More Proficient speaker condition. This pattern is opposite what the expectation account would predict and will be discussed further below; to preview, however, I suggest that this strongly implies that the expectation account does not explain this data. That is, it is not correct to say that listeners process more shallowly due to the expectation that non-native speech is unreliable. Post-hoc analyses exploring this interaction found that this interaction is driven by a relatively large priming effect in the Less Proficient speaker condition ($B = 1.05$; odds ratio = 2.87, $p < .001$), whereas the priming effect in the More Proficient speaker condition is marginal but non-significant ($B = 0.26$; odds ratio = 1.30, $p = .09$). This difference between the two speaker conditions can be seen in Figure 5.

Table 6: Model output for Block 1 and Block 2.

		Beta	SE	z	p
Block 1	(Intercept)	-0.19	0.23	-0.83	.400
	Prime	0.63	0.12	5.14	<.001
	SpeakProf	-0.10	0.27	-0.38	.71
	Prime:SpeakProf	0.78	0.24	3.23	<.01
Block 2	(Intercept)	-0.69	0.21	-3.25	<.01
	Prime	0.66	0.12	5.43	<.001
	SpeakProf	-0.21	0.36	-0.57	0.57
	Prime:SpeakProf	-0.35	0.23	-1.53	0.13

But what about the results in Block 2? Do participants who hear ungrammatical sentences from the Less Proficient speaker in the second half show this same pattern? Or, do we see carryover effects comparable to Experiment 1? The Block 2 results do show a significant main effect of

Prime (B = .66, odds ratio = 1.93; Table 6) and no main effect of Speaker. The Speaker x Prime interaction was not significant in the second half. Post hoc analyses of each Speaker condition in Block 2 showed a significant priming effect in both conditions, although the priming effect in the More Proficient condition (B = 0.83; odds ratio = 2.29, p < .001) was numerically larger than the priming effect in the Less Proficient condition (B = 0.49; odds ratio =1.63, p < .01). This pattern of results is numerically consistent with the carryover results in Experiment 1. This also adds to the evidence suggests that reduced speaker proficiency does not always lead to reductions in priming.

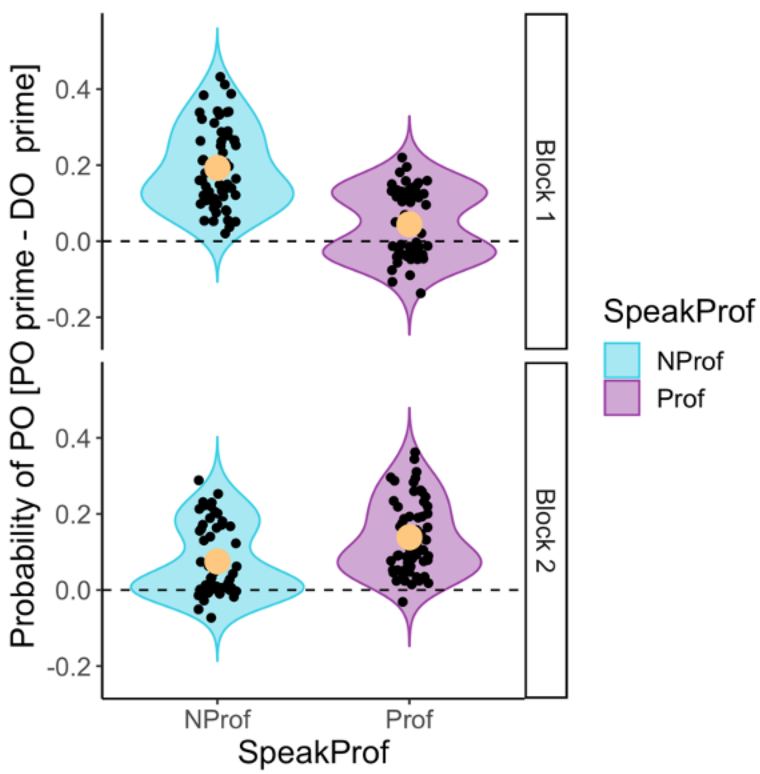

Figure 5: Black points represent model-estimated individual means (transformed from logits to probability). The cream-colored points represents the group mean for each cell and are not direct model estimates. Values greater than zero reflect positive priming effects (increased probability of producing PO after PO prime).

3.2.3 Speaker Perception Questionnaire Results

As in Experiment 1, I examined whether participants indicated that they perceived a difference in proficiency between the two Speakers. If our proficiency manipulation is successful, the less proficient speaker should be rated lower than the more proficient speaker. However, the mean proficiency for the more proficient speaker (Figure 6; M = 5.11, SD = 1.33) was approximately the same as for the less proficient speaker (M = 5.16, SD = 1.39), and a paired sample *t*-test found no significant difference between ratings for the two speaker conditions ($t(111) = 0.55$, $p = .58$). There are a few reasons why this may have occurred. One, participants may not have detected the grammatical errors at all. This seems unlikely, because the errors were designed to be noticeable, were intentionally placed at the start of the less-proficient block, and pilot participants did report that the errors were noticeable. Or, despite our best efforts at ensuring participants realized that the speakers were distinct, it is possible that they did not remember which speaker had made the grammatical errors when they completed the questionnaire, and thus rated both speakers similarly. Finally, participants may have detected the grammatical errors, but not penalized the speaker for them in their rating. Perhaps participants did not consider the grammatical errors as a factor in their rating and rated the speakers' proficiency based on other features of their speech instead (e.g. accentedness, speech rate, intelligibility). In any case, the fact that the grammatical errors are not reflected in participants' proficiency ratings makes the influence of the grammatical errors on priming all the more compelling. This perhaps suggests that listeners assessments of the speaker's proficiency is not what drove the priming effects, but rather the priming effects were driven simply by adaptations to syntactic errors in the environment.

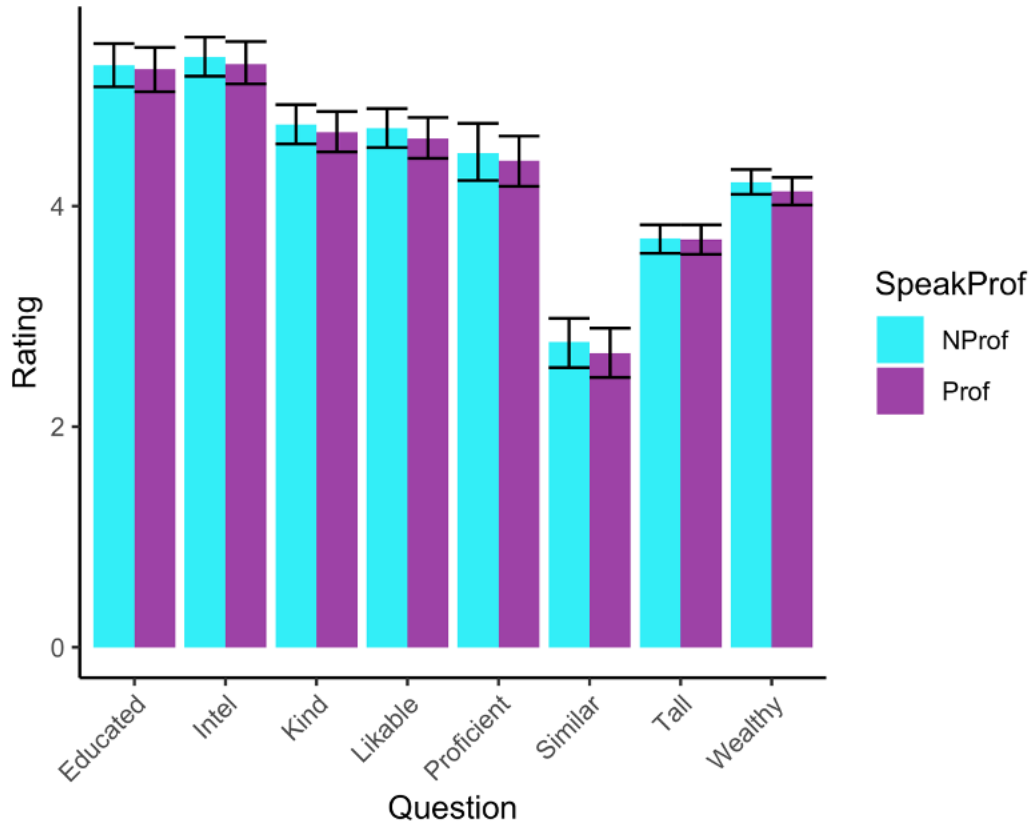

Figure 6: Speaker perception ratings for each item. Error bars represent bootstrapped 95% confidence intervals.

3.2.4 Comprehension Accuracy Results

As in Experiment 1, overall comprehension accuracy was extremely high (98.7%) across all conditions and did not differ by speaker (More Proficient = 99%, Less Proficient = 98.4%). The lack of difference is not particularly surprising. Given that comprehension accuracy was very high also for the Native speaker in Experiment 1, there was no reason to expect that adding ungrammatical fillers would reduce comprehension. But, this high accuracy does demonstrate that participants on the whole were attending to the sentence recordings.

3.2.5 Recognition Accuracy Results

The pattern of recognition accuracy results in Experiment 2 was almost identical to the pattern in Experiment 1. Overall, participants' recognition accuracy was low (M = 68%, SD = 10.4%); however, there was a fairly wide range of accuracy scores, with some participants scoring as low as 25% or as high as 100%.

These data were analyzed using the same model structure as in Experiment 1. Overall, there was no response bias towards responding either "New" or "Old", as evidenced by the intercept parameter (B = -0.03, p = .84). There was a main effect of true Correct response (B = 1.74, p < .001), such that the odds of responding that an item was "Old" when it was truly old were 5.7 times higher than responding "New". There was no main effect of Speaker (B = .04, p = .81), meaning that there was no significant difference in in response bias between the two levels of Speaker proficiency. There was also no significant interaction between the Correct response and Speaker condition (B = 0.19, p = .52), suggesting that there was no detectable effect of Speaker proficiency on responding correctly. Indeed, recognition accuracy for the More Proficient speaker (M = 67.6%, SD = 14.7%) was nearly identical to that for the Less Proficient speaker (M = 68.3%, SD = 13.5%). Given the low overall accuracy and the null results across two experiments, it is difficult to conclude anything about the influence of non-nativeness and proficiency on memory representations of speech, so I will not discuss these results further.

3.2.6 Accent Exposure Results

The accent exposure data again revealed that participants in Experiment 2 had very little exposure to other accents overall (Table 7). The same analyses were conducted on this data as in Experiment 1, but again, no influence of accent exposure was found.

Table 7: Global estimates of exposure to accent types on a weekly basis.

	Mean	Median	SD
American English	92%	100%	15.0%
Chinese	1.9%	0%	5.68%
Other English	2.11%	0%	6.9%
Other Accents	3.96%	0%	9.57%

3.3 Discussion

The expectation account posits that listeners assume that non-native speakers are less reliable speakers, and so the speech input is not processed as deeply. In Experiment 2, I tested this idea by manipulating the reliability of the speaker directly. The expectation account predicts that lower speaker proficiency should lead to reductions in priming because listeners shallowly process the speech input as an adaptive processing mode.

I found that proficiency does seem to influence priming, but not in the manner predicted by the expectation account. Instead, in Block 1, I found that participants in the Less Proficient speaker condition showed *more* priming, suggesting listeners had a stronger structural representation of utterances from the Less Proficient speaker. In contrast, participants in the More Proficient speaker condition showed a non-significant priming effect.

In contrast to the expectation account, the principles of good-enough processing can explain the full set of results seen here. The good-enough processing account could explain the increased priming from the Less Proficient speaker as follows: in the Less Proficient speaker condition, participants are led to expect syntactic errors in addition to the processing burden of speech processing. To compensate, they devote more processing resources to each sentence. On all of the prime sentences, these increased processing resources are deployed on well-formed sentences, leading listeners to fully process the structure of the primes and subsequently have a strengthened representation of the sentence structure. In contrast, participants in the More Proficient condition do not expect syntactic errors and are simply optimizing their processing resources to the processing of the accent. They find an optimal level of processing resources that allows them to understand the accented speech well enough to do the task and does not require full structural analysis of the sentences.

Although the carryover effect was not as strong in the analyses presented here[3], the good-enough processing account does explain the numerical pattern towards the carryover effect. That is, in the analysis of the effect of proficiency on priming, participants in the Less Proficient-More Proficient order exhibited numerically more priming in the second half of the experiment than participants in the More Proficient-Less Proficient order.

---

[3] Analyses of the priming effect on each list does show a strong resemblance to the carryover effect seen in Experiment 1, but these results are not presented here for the sake of space and sanity.

One interesting feature of this data is that the More Proficient speaker condition in Experiment 2 is basically equivalent to the non-native speaker in Experiment 1. That is, the More Proficient speaker is non-native speaker who does not produce grammatical errors, same as in Experiment 1. Thus, the non-significant priming effect in the More Proficient speaker condition can be seen as a replication of the non-significant priming effect of the non-native speaker in Experiment 1.

4.0 General Discussion

In this set of experiments, I set out to examine what knowledge and strategies listeners use to accommodate the variability of spoken language. Modern theories of language processing emphasize the role of past experience in language processing (Jurafsky, 1996; Levy, 2008; MacDonald et al., 1994); over time, listeners may track richly detailed statistical information about language and speakers that they can use to guide processing (Kleinschmidt, 2019; Sumner et al., 2014). But, what happens when this past experience is no longer helpful? Listeners may have varying levels of familiarity with different words, constructions, dialects, accents, and idiolects, and they may have to rely on different processing strategies when their prior experience with these speech characteristics is low. One form of knowledge listeners might use when listening to an unfamiliar speaker is an expectation about the speaker's proficiency, which could allow listeners to adapt their processing and avoid being misled by unreliable features of the speech (e.g., ungrammatical sentences, mislabeled referents). Using structural priming as a measurement tool, I wanted to test whether listeners use knowledge about speaker proficiency to guide their comprehension processes while listening to non-native speakers. Across two experiments, I found some evidence that speaker proficiency influences structural processing, although not in the way that some accounts suggest. I also found consistent task-related priming effects that raise interesting questions about the mechanisms of structural priming.

I found no clear directional relationship between speaker proficiency and the comprehension processes used. In the first block of Experiment 1 the native speaker elicited more priming than the non-native speaker, but in the first block of Experiment 2 the less proficient non-native speaker elicited more priming than the more proficient native speaker. These first block results are, overall, inconsistent with the prediction of the expectation account (Lev-Ari, 2015) that listeners always rely less on bottom-up information when comprehending non-native speech. Instead, these results imply that the relationship between proficiency and priming is more complicated and is likely to be a function of the task and characteristics of the speech input.

I suggest that one takeaway from this set of experiments is that listeners' comprehension of non-native speech reflects contextually optimized processing strategies rather than an intrinsic reliance on top-down comprehension when processing non-native speech, and they are better explained by the more general framework of good-enough processing (Ferreira et al., 2002). Perhaps one underlying principle guiding processing is that listeners direct resources towards repairing the parts of the input that deviate (or are likely to deviate) from what they have prior experience with (e.g. Caffarra & Martin, 2019; Gibson et al., 2017; Hanulíková et al., 2012; Ivanova et al., 2017). For example, in Experiment 1, listeners may have spent resources on speech processing because the speech sounds produced by the non-native speaker were different than what they have experience with. In Experiment 2, listeners may have prepared to spend resources on structural repair because of the increased likelihood of syntactic errors, but instead were met with well-formed prime sentences. A process like this could in theory be guided by features of the stimulus itself or by top-down sources (such as distributional knowledge about non-native speakers or judgements of proficiency). However, given the carry-over effects seen in both experiments, these processing repair strategies were not deployed in a speaker specific manner, suggesting that these strategies were initiated in response to the stimuli themselves, and simply carried over to subsequent speakers. In other words, although the initial Block 1 effects here seem to be driven by the proficiency of the speaker (though, notably, in a manner that is non-monotonic with respect to changes in speaker proficiency), because priming in Block 2 does not change according to the novel speaker, it is difficult to make the claim that Block 1 processing was due to speaker-specific processing. Instead, the more plausible explanation is that initial structural priming is governed by the features of the input (which happen to vary by speaker), and then subsequent priming becomes entrenched to the initial mode of processing.

4.1 Reconciliation with Related Work

Across the small number of prior studies that have examined the influence of non-native speech on structural priming, the direction of this effect has been mixed. Chun et al. (2016) found that listeners showed more priming to non-native speakers (opposite to the results of the first block in Experiment 1 presented here), and that this was likely a function of familiarity, where decreased familiarity leads to increased priming. Similarly, Chun and Kaan (2020) found that L2 listeners showed more priming from L2 accents that were unfamiliar. Although this contrasts with the current results, I believe this can be reconciled on the basis of task-related differences. Specifically, both of these prior studies required participants to fully repeat prime sentences, and the authors rejected trials from analysis where participants did not successfully repeat the prime sentences. Repeating the primes requires a different level of task engagement than the current studies—which did not require participants to repeat primes—and it may lead to stronger structural activation via self-priming (see Jacobs et al., 2019). Both factors together can account for the strengthening effect of accent familiarity on priming in these studies. In contrast to the results of Chun et al. (2016) and Chun and Kaan (2020), Weatherholtz et al., (2014) find that listeners syntactically align more to speakers whose accent is more standard. However, the Weatherholtz et al. (2014) study examines alignment after hearing a lengthy monologue, whereas the other two studies use a trial-to-trial paradigm similar to the current study. Given these contrasting findings, as well as the somewhat complicated pattern of results in the current study, it is likely that the influence of non-native accent on structural priming in both experimental and real-world contexts is a complex function task goal, accent familiarity, and social factors.

4.2 Mechanisms of Structural Priming

Although the primary goal of these studies was not to examine the underlying mechanisms of structural priming, these results do speak to this somewhat. Specifically, the residual-activation view of priming (e.g. Pickering & Branigan, 1998) suggests that structural priming arises as a result of brief activation of structural information that decays rapidly. This type of priming is relatively automatic and obligatory, and it could be viewed as an epiphenomenon rather than a functional component of the language processor. By contrast, the implicit-learning view (Chang et al., 2006; Reitter et al., 2011) suggests that priming may arise as a result of learning (possibly statistical information) about the environment, and that priming thus reflects an important function of the processor.

The traditional residual-activation view cannot explain the moderations of the priming effect seen here without relying on theory-external mechanisms. There is no reason that residual activation should lead to varied effects of priming across tasks seen here, as long as structure is always being activated. The residual-activation view *could* explain the results of the two experiments here if we assume that processes upstream of structure processing modulate the extent to which listeners rely on structure. For example, as I propose above, if listeners engage in heuristic-based processing and no structural parse, priming should go down because the structural nodes are never activated in the first place.

The implicit-learning view would see these results in a slightly different way. Implicit learning can be context-specific (Kaan & Chun, 2018) and therefore give a privileged status to the specific features of the environment or input. An implicit learning account thus sees the moderations of the priming effect here as an adaptation to the environment in a way that should benefit comprehension. In particular, the implicit learning account is consistent with the carryover effects observed being a function of adaption to the particulars of the task environment, rather than adaption to particular speakers.

These results also raise interesting questions about additional factors that play into structural priming. The carryover effects seen in both studies imply that the choice of structure for any given language production may be driven by environmentally-oriented production mechanisms (preceding linguistic context) and by internally-oriented production mechanisms (personal structural preference, possibly dictated by distributional characteristics). That is, listeners in the Native—Non-Native order in Experiment 1 showed consistent priming throughout both blocks, suggesting that their choice of structure was susceptible to the influence of the structure of the preceding prime (environmental influence) even in the non-native block. In contrast, participants in the Non-native—Native order were on average less-susceptible to the effect of primes, suggesting that their production decisions even in the Native block were guided more by internally-oriented mechanisms. This suggests that production decisions may become routinized as a function of the task and preceding language input. This analysis shares similarity to the account of priming by Myslín and Levy (2016), who proposed that structural priming may arise as a function of rational expectation for repetition. Under this account, features of the language environment itself (e.g. accent or ungrammaticalities) may moderate the degree to which listeners rely on environmental cues to expect repetition (or in this case, to influence production).

One view of structural priming is that re-using structures is a function of more general linguistic alignment, which facilitates conversation (Pickering & Garrod, 2004). One a priori possibility is that structural priming is always reduced when comprehending non-native speech, which could result in impairments to conversation between native and non-native speakers. This could add difficulty to tasks where native and non-native speakers need to coordinate their efforts on a common goal. However, the results of the current experiments suggest that there is not a universal decrease in structural priming from non-native speakers, and so reduced alignment at the structural level is not necessarily a source of difficulty inhibited conversation between native and non-native talkers. This conclusion comports with other investigations into non-native accent and structural priming that show both positive and negative priming effects (E. Chun et al., 2016; Constantine et al., In Prep; Weatherholtz et al., 2014).

4.3 Limitations

This set of studies does have a number of limitations that merit discussion. First, the prime sentences used in this set of experiments are all plausible, grammatical datives. As a result, listeners have very little in the way of other sources of top-down information to use to parse these sentences that is not also contained in the bottom-up input. A stronger test of the expectation account would include, for example, implausible prime sentences which would allow listeners to bring in prior knowledge about plausibility to comprehend these sentences and rely less on the literal structural information. Thus, these results should not be taken as a strong test of the expectation account.

Second, both experiments had quite a high rate of picture descriptions coded as neither PO nor DO (approximately 30% in both experiments). Rates of Other descriptions this high are not abnormal in open-ended description tasks, but discarding so many trials does lead to a less than optimal use of participant and experimenter effort. Nevertheless, the Other structures were equally common across all conditions in each experiment and are thus not driving the differences across conditions. Future iterations of this work may benefit from adding a verbal constraint on participant responses (e.g., participants must use the verb "give" or some other dative verb in their description) in order to mitigate data loss issues.

Third, the comprehension questions were not difficult and were not asked after the prime sentences. The questions were constructed this way to unobtrusively ensure participants' attention remained on task without artificially requiring deep processing of the prime sentences (see Swets et al., (2008) for an example of comprehension goals influencing depth of sentence processing). But because the questions were easier, we cannot know just how deeply participants processed or understood the primes. However, given the simplicity of the prime sentences, it is hard to imagine that participants did have a hard time understanding them.

Lastly, I only tested the influence of one non-native accent: L1 Mandarin speakers. But, accents differ in their level of prestige (Cargile et al., 2010), the experience listeners have with them, and the negative biases they may induce in listeners (Purnell et al., 1999). These differences may well be germane to syntactic comprehension because factors such as liking (Balcetis & Dale, 2005), prestige (Lev-Ari, 2016), familiarity with accent (Chun et al., 2016), and similarity (Weatherholtz et al., 2014) have been demonstrated to moderate syntactic priming. The current study (unsuccessfully) attempted to determine the relationship between some of these factors and priming of non-native speech by measuring exposure to accents and listener's perceptions of the speakers. However, there did not appear to be an influence of these factors on structural priming, possibly owing in part to the fact that participants language background was not particularly diverse. Future work should try to distinguish the unique effects of each of these factors on the processing of non-native speech, possibly by experimentally manipulating them.

4.4 Conclusion

The current set of studies demonstrate that listeners rely on contextually-optimized processing strategies to process non-native speech, possibly in the service of simply repairing the parts of the input that are anomalous (or likely to be anomalous). Input variability (such as accent and ungrammaticality) does appear to play a role in influencing language processing but not in a manner that suggests that linear decreases in speaker proficiency do not uniformly lead to shallower processing. Additionally, these results are not consistent with views of language processing that give a privileged representation to speaker identity. Instead, the specific features of the language input appear to be driving comprehension processes, and the listener adapts their comprehension as needed when the input is unfamiliar.

Appendix A Sentence Stimuli

Appendix A.1.1 Dative Primes: PO and DO alternations

The corrupt inspector offered a deal to the bar owner.

The corrupt inspector offered the bar owner a deal.

The graduate students are baking a cake for the professors.

The graduate students are baking the professors a cake.

The lifeguard tossed a rope to the struggling swimmer.

The lifeguard tossed the struggling swimmer a rope.

The governess made a pot of tea for the princess.

The governess made the princess a pot of tea.

The foundation is giving several million dollars to the university.

The foundation is giving the university several million dollars.

A rock star sold some cocaine to an undercover agent.

A rock star sold an undercover agent some cocaine.

The legislature is sending a bill legalizing capital punishment to the governor.

The legislature is sending the governor a bill legalizing capital punishment.

The management company is renting three suites of offices to the CIA.

The management company is renting the CIA three suites of offices.

The team owner told an offensive joke to the columnist.

The team owner told the columnist an offensive joke.

The cheerleader saved a seat for her boyfriend.

The cheerleader saved her boyfriend a seat.

The dictator bought a Rolls Royce for the terrorist leader.

The dictator bought the terrorist leader a Rolls Royce.

The waitress took a tray of appetizers to the customers.

The waitress took the customers a tray of appetizers.

The credit card company mailed an application to the student.

The credit card company mailed the student an application.

The indulgent mother promised a puppy to her daughter.

The indulgent mother promised her daughter a puppy.

The judge awarded a hundred thousand dollars to the plaintiff.

The judge awarded the plaintiff a hundred thousand dollars.

The clerk issued an office key to the new typist.

The clerk issued the new typist an office key.

The ambitious father taught the alphabet to his 3-year-old son.

The ambitious father taught his 3-year-old son the alphabet.

The little girl read a short story to the old woman.

The little girl read the old woman a short story.

The driver sheepishly handed his license to the police officer.

The driver sheepishly handed the police officer his license.

The bored teen passed a note to the cute guy.

The bored teen passed the cute guy a note.

The toddler timidly fed a carrot to the rabbit.

The toddler timidly fed the rabbit a carrot.

The cocktail waitress served a martini to the tired executive.

The cocktail waitress served the tired executive a martini.

Mozart wrote a song for his wife.

Mozart wrote his wife a song.

The tenant owed 6 months' rent to the landlord.

The tenant owed the landlord 6 months' rent.

Appendix A.1.2 Intransitive Fillers

A flying saucer landed.

The young electrician fell down.

The vacationing family stayed together.

The angry customer stormed out.

The unhappy artist sighed.

The new television network struggled.

The reluctant volunteer slacked off.

The clerks giggled.

The real estate agent blundered.

The duck hunters whispered.

The successful business man retired.

The young man shaved too often.

The defeated king barely escaped.

The company's problems multiplied.

The old women gambled every day.

The computer crashed.

The lost child cried.

The resourceful campers survived.

The clock isn't running.

The tightrope walker fell.

The moon is shining brightly.

The university went broke.

The audience didn't laugh.

Mister Rogers smiles frequently.

Fred Astaire and Ginger Rogers danced.

The horse galloped gracefully.

The bus driver sneezed suddenly.

The young couple strolled arm in arm.

The delicate vase shattered.

The kidnapped child escaped.

Appendix A.1.3 Predicate Adjective Fillers

The textbooks were expensive.

The dinner was delicious.

The yellow curry was not too hot.

The man's favorite band was loud.

The computer was broken.

Her red purse was very spacious.

The professor's office was messy.

The puppy's ears were floppy and adorable.

The barista's coffee was burnt and bitter.

Mariah Carey's voice is amazing.

The scientific process is often hard.

The speaker at the meeting was too quiet.

The couch from Ikea wasn't cheap.

All the siblings are finally together.

The candle flame was small but hot.

Her rounded glasses are stylish.

The big green tractor was efficient.

The college students were indebted after graduation.

White wine is not always good.

The doctors and nurses were surprised.

City busses are usually dirty and always loud.

Pittsburgh's skyline is beautiful.

His tweets were annoying.

Her roommate is really messy.

Trader Joe's candy is too sweet.

The garden was very green.

The new iPhone isn't much faster.

The trash cans on the street are full.

The kitten's fur was very soft.

The old elevator is slow.

Appendix A.1.4 Transitive Fillers: Passive and Active Alternations

A compromise is being suggested by the chairperson.

The chairperson is suggesting a compromise.

The referee was punched by one of the fans.

One of the fans punched the referee.

The returning astronauts were welcomed by a brief ceremony.

The brief ceremony welcomed the returning astronauts.

The building manager was mugged by a gang of teenagers.

A gang of teenagers mugged the building manager.

A passerby was jostled by the drunk.

The drunk jostled a passerby.

The jogger wasn't tripped by a chain.

A chain didn't trip the jogger.

The car's windshield was struck by a brick.

The brick struck the car's windshield.

The embassy staff isn't being evacuated by the government.

The government isn't evacuating the embassy staff.

The film critic was charmed by the new children's movie.

The new children's movie charmed the film critic.

The mayor was observed by a reporter leaving the mobster's home.

The reporter observed the mayor leaving the mobster's home.

The players are being assisted by a union leader in organizing the strike.

The union leader is assisting the players in organizing the strike.

The designer's favorite dress was worn by a bald fashion model.

The bald fashion model wore the designer's favorite dress.

The file was dropped by a clerk into the wastebasket.

The clerk dropped the file into the wastebasket.

A corpse was found by some hunters behind the ice cream plant.

The hunters found a corpse behind the ice cream plant.

An innocent bystander was grazed by the assassin's bullet.

The assassin's bullet grazed the innocent bystander.

The floors are cleaned by a janitor daily.

A janitor cleans the floors daily.

The potholes are being repaired by a crew from the city.

The crew from the city is repairing the potholes.

The news was dominated by an article about the hurricane.

The article about a hurricane dominated the news.

The bicycle was forced off the road by a truck.

The motorcycle forced the bicycle off the road.

The Cavaliers were beaten by the Warriors in four games.

The Warriors beat the Cavaliers in four games.

The valley's stillness was shattered by a gunshot.

The gunshot shattered the valley's stillness.

The chess master was outsmarted by the computer.

The computer outsmarted the chess master.

A medieval manuscript was misplaced by the museum after the exhibit.

The museum misplaced a medieval manuscript after the exhibit.

Thousands of acres of forest were destroyed by the fire in less than a week.

The fire destroyed thousands of acres of forest in less than a week.

Appendix B Fictional Biographical Information from Experiment 2

First name: Zehan

Hometown: Shenzhen, China

Career goal: High school teacher

Hobby: Running

First name: Suzy

Hometown: Guangzhou, China

Career goal: Social work

Hobby: Playing music

Bibliography

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), 914–930. https://doi.org/10.1037/0278-7393.33.5.914

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. The Journal of the Acoustical Society of America, 133(3), EL174–EL180. https://doi.org/10.1121/1.4789864

Balcetis, E. E., & Dale, R. (2005). An Exploration of Social Modulation of Syntactic Priming. Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27), 7.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Birdsong, D., Gertken, L. P., & Amengual, M. (2012). Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL.

Birdsong, D., & Molis, M. (2001). On the Evidence for Maturational Constraints in Second-Language Acquisition. Journal of Memory and Language, 44, 235–249. https://doi.org/10.1006/jmla.2000.2750

Bock, K. (1986). Syntactic persistence in language production. Cognitive Psychology, 18(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bock, K., Dell, G., Chang, F., & Onishi, K. (2007). Persistent structural priming from language comprehension to language production☆☆☆. Cognition, 104(3), 437–458. https://doi.org/10.1016/j.cognition.2006.07.003

Bock, K., & Griffin, Z. M. (2000). The Persistence of Structural Priming: Transient Activation or Implicit Learning? 16.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. Cognition, 106(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Caffarra, S., & Martin, C. D. (2019). Not all errors are the same: ERP sensitivity to error typicality in foreign accented speech perception. Cortex, 116, 308–320. https://doi.org/10.1016/j.cortex.2018.03.007

Caffarra, S., Michell, E., & Martin, C. D. (2018). The impact of foreign accent on irony interpretation. PLOS ONE, 13(8), e0200939. https://doi.org/10.1371/journal.pone.0200939

Cargile, A. C., Maeda, E., Rodriguez, J., & Rich, M. (2010). "Oh, You Speak English So Well!": U.S. American Listeners' Perceptions of "Foreignness" among Nonnative Speakers. Journal of Asian American Studies, 13(1), 59–79. https://doi.org/10.1353/jaas.0.0062

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. Psychological Review, 113(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. Quarterly Journal of Experimental Psychology, 69(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603

Chun, A. E., Day, R. R., Chenoweth, N. A., & Luppescu, S. (1982). Errors, Interaction, and Correction: A Study of Native-Nonnative Conversations. TESOL Quarterly, 16(4), 537. https://doi.org/10.2307/3586471

Chun, E., Barrow, J., & Kaan, E. (2016). Native English speakers' structural alignment mediated by foreign-accented speech. Linguistics Vanguard, 2(s1). https://doi.org/10.1515/lingvan-2016-0027

Chun, E., & Kaan, E. (2020). The effects of speaker accent on syntactic priming in second-language speakers. Second Language Research, 026765832092656. https://doi.org/10.1177/0267658320926563

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. The Journal of the Acoustical Society of America, 116(6), 3647–3658. https://doi.org/10.1121/1.1815131

Cleland, A., & Pickering, M. (2006). Do writing and speaking employ the same syntactic representations? Journal of Memory and Language, 54(2), 185–198. https://doi.org/10.1016/j.jml.2005.10.003

Constantine, R., Getty, D. J., & Fraundorf, S. H. (In Prep). Syntactic Adaptation to Native versus Non-Native Speech.

Fairchild, S., & Papafragou, A. (2018). Sins of omission are more likely to be forgiven in non-native speakers. Cognition, 181, 80–92. https://doi.org/10.1016/j.cognition.2018.08.010

Fang, X., Perfetti, C., & Stafura, J. (2017). Learning new meanings for known words: Biphasic effects of prior knowledge. Language, Cognition and Neuroscience, 32(5), 637–649. https://doi.org/10.1080/23273798.2016.1252050

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. Current Directions in Psychological Science, 11(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Fraundorf, S. H., & Jaeger, T. F. (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. Journal of Memory and Language, 91, 28–58. https://doi.org/10.1016/j.jml.2016.05.006

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't Underestimate the Benefits of Being Misunderstood. Psychological Science, 28(6), 703–712. https://doi.org/10.1177/0956797617690277

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. 24(4), 10.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. Linguistics, 48(4). https://doi.org/10.1515/ling.2010.027

Ivanova, I., Branigan, H. P., McLean, J. F., Costa, A., & Pickering, M. J. (2017). Do you what I say? People reconstruct the syntax of anomalous utterances. Language, Cognition and Neuroscience, 32(2), 175–189. https://doi.org/10.1080/23273798.2016.1236976

Jacobs, C. L., Cho, S., & Watson, D. G. (2019). Self-Priming in Production: Evidence for a Hybrid Model of Syntactic Priming. Cognitive Science, 43(7). https://doi.org/10.1111/cogs.12749

Jurafsky, D. (1996). A Probabilistic Model of Lexical and Syntactic Access and Disambiguation. Cognitive Science, 20(2), 137–194. https://doi.org/10.1207/s15516709cog2002_1

Kaan, E., & Chun, E. (2018). Syntactic Adaptation. In Psychology of Learning and Motivation (Vol. 68, pp. 85–116). Elsevier. https://doi.org/10.1016/bs.plm.2018.08.003

Kamide, Y. (2012). Learning individual talkers' structural preferences. Cognition, 124(1), 66–71. https://doi.org/10.1016/j.cognition.2012.03.001

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. Quarterly Journal of Experimental Psychology, 69(5), 1013–1040. https://doi.org/10.1080/17470218.2015.1053951

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. Journal of Experimental Psychology: General, 133(3), 450–467. https://doi.org/10.1037/0096-3445.133.3.450

Kleinschmidt, D. F. (2019). Structure in talker variability: How much is there and how much can it help? Language, Cognition and Neuroscience, 34(1), 43–68. https://doi.org/10.1080/23273798.2018.1500698

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. Psychological Review, 122(2), 148–203. https://doi.org/10.1037/a0038695

Kleinschmidt, D. F., Weatherholtz, K., & Jaeger, T. F. (2018). Sociolinguistic Perception as Inference Under Uncertainty. Topics in Cognitive Science, 10(4), 818–834. https://doi.org/10.1111/tops.12331

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? Cognitive Psychology, 51(2), 141–178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Lev-Ari, S. (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.01546

Lev-Ari, S. (2016). Selective Grammatical Convergence: Learning From Desirable Speakers. Discourse Processes, 53(8), 657–674. https://doi.org/10.1080/0163853X.2015.1094716

Lev-Ari, S., & Keysar, B. (2012). Less-Detailed Representation of Non-Native Language: Why Non-Native Speakers' Stories Seem More Vague. Discourse Processes, 49(7), 523–538. https://doi.org/10.1080/0163853X.2012.698493

Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the Speech Code. Psychological Review, 74(6), 431–461.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The Lexical Nature of Syntactic Ambiguity Resolution. Psychological Review, 101(4), 676–703. http://dx.doi.org/10.1037/0033-295X.101.4.676

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. Journal of Memory and Language, 91, 5–27. https://doi.org/10.1016/j.jml.2016.03.009

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. Journal of Memory and Language, 94, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. Cognitive Science, 32(3), 543–562. https://doi.org/10.1080/03640210802035357

Munro, M. J. (1993). Productions of English Vowels by Native Speakers of Arabic: Acoustic Measurements and Accentedness Ratings. Language and Speech, 36(1), 39–66. https://doi.org/10.1177/002383099303600103

Myslín, M., & Levy, R. (2016). Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. Cognition, 147, 29–56. https://doi.org/10.1016/j.cognition.2015.10.021

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. Journal of Language and Social Psychology, 18(1), 62–85. https://doi.org/10.1177/0261927X99018001005

Pickering, M. J., & Branigan, H. P. (1998). The Representation of Verbs: Evidence from Syntactic Priming in Language Production. Journal of Memory and Language, 39(4), 633–651. https://doi.org/10.1006/jmla.1998.2592

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. Psychological Bulletin, 134(3), 427–459. https://doi.org/10.1037/0033-2909.134.3.427

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences, 27(02). https://doi.org/10.1017/S0140525X04000056

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and Phonetic Experiments on American English Dialect Identification. Journal of Language and Social Psychology, 18(1), 10–30. https://doi.org/10.1177/0261927X99018001002

Reitter, D., Keller, F., & Moore, J. D. (2011). A Computational Cognitive Model of Syntactic Priming. Cognitive Science, 35(4), 587–637. https://doi.org/10.1111/j.1551-6709.2010.01165.x

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. Neuropsychologia, 85, 245–255. https://doi.org/10.1016/j.neuropsychologia.2016.03.022

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. Cognition, 181, 141–150. https://doi.org/10.1016/j.cognition.2018.08.018

Seifeldin, M., Cantor, M., Boland, J., & Brennan, Jonathan. (2015). They dropping copulas: Salient cues in the integration of speaker identity and syntax. CUNY Conference on Human Sentence Processing, Los Angeles.

Strand, E. A. (1999). Uncovering the Role of Gender Stereotypes in Speech Perception. Journal of Language and Social Psychology, 18(1), 86–100. https://doi.org/10.1177/0261927X99018001006

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.01015

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. Memory & Cognition, 36(1), 201–216. https://doi.org/10.3758/MC.36.1.201

Trofimovich, P., & Baker, W. (2006). LEARNING SECOND LANGUAGE SUPRASEGMENTALS: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. Studies in Second Language Acquisition, 28(01). https://doi.org/10.1017/S0272263106060013

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(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. Language Variation and Change, 26(3), 387–420. https://doi.org/10.1017/S0954394514000155

Williams, R., Schleicher, K., & Ivanova, I. (2020). Structural Alignment to Non-native Accented Speech [Poster Presentation]. Psychonomics.