University of Pennsylvania Working Papers in Linguistics

Volume 25
Issue 2 *Selected Papers from NWAV47*

Article 16

1-15-2020

How do Listeners Form Grammatical Expectations to African American Language?

Rachel Elizabeth Weissler
University of Michigan

Jonathan R. Brennan
University of Michigan

How do Listeners Form Grammatical Expectations to African American Language?

Abstract

Ideologies about standard language in the United States often posit Mainstream U.S. English (MUSE) as a morally superior variety (Hill 2008). Previous research has shown that this kind of hierarchical treatment of language varieties leads to negative perceptions of non-standard languages, which in turn makes them stigmatized, and ultimately perpetuates dialect discrimination. This kind of discrimination results in the mistreatment of users of non-standard varieties, which negatively affects the way those speakers can move through the U.S. context (Rickford 1999, Eckert and Rickford 2001, Schilling 2004, Rickford and King 2016). This study investigates how listeners alter their linguistic expectations when hearing speakers of standard and non-standard varieties of English through an Electroencephalography (EEG) experiment. We probe how social information influences syntactic processing to see if and when speakers of a standard dialect, MUSE, form grammatical expectations when processing MUSE and African American Language (AAL). Looking at online processing helps us better understand whether listeners have specific knowledge of the dialect that is not their own (dialect-specific hypothesis), or whether listeners more generally reduce expectations across the board when listening to a dialect or variant that they themselves do not speak (dialect non-specific hypothesis). In order to test the interaction between language variety and auxiliary usage, experimental sentences were constructed in order to reflect a variant that is grammatical in MUSE, a variant that is grammatical uniquely to AAL, and a grammatical variant that is ungrammatical in all varieties of English. The experimental stimuli from a bidialectal Midwestern black speaker of both MUSE and AAL, yielding a within subject 2 (language varieties) by 3 (grammatical features) design. The results do not cleanly favor the dialect-specific nor the dialect non-specific hypothesis. Rather, the evidence points a nuanced version of a mixture of both hypotheses. Through analysis of American English dialects, this work contributes to further understanding of how social information interfaces with online processing, and expectations that may be formed depending on the perceived identity of a voice. The impact of this work is paramount, as perceptions of stigmatized language varieties can lead to dialect discrimination that negatively affects the way those speakers are treated (Rickford 1999, Purnell, Baugh, Idsardi 1999, Eckert and Rickford 2001, Schilling 2004, Rickford and King 2016)

How Do Listeners Form Grammatical Expectations to African American Language?

Rachel Elizabeth Weissler and Jonathan R. Brennan

1 Introduction

Ideologies about standard language in the United States often posit Mainstream U.S. English (MUSE) as a morally superior variety (Hill 2008). Previous research has shown that this kind of hierarchical treatment of language varieties leads to negative perceptions of non-standard languages, which in turn makes them stigmatized, and ultimately perpetuates dialect discrimination. This kind of discrimination results in the mistreatment of users of non-standard varieties, which negatively affects the way those speakers can move through the U.S. context (Rickford 1999, Eckert and Rickford 2001, Schilling 2004, Rickford and King 2016). The current study investigates how listeners alter their linguistic expectations when hearing speakers of standard and non-standard varieties of English through an Electroencephalography (EEG) experiment. Specifically, we probe how social information influences syntactic processing to see if and when speakers of a standard dialect, MUSE, form grammatical expectations when processing MUSE and African American Language (AAL). This research has theoretical implications for psycholinguistics, syntactic theory, sociolinguistics, as well as direct social implications. Using EEG to study issues of language processing, perception, and social identity is beneficial for sociolinguists as it helps us gain better insight into online processing, allowing us to see neurological responses at each instantiation of speech produced, which helps us in turn to address central questions about the language faculty itself (e.g. How do we represent grammatical knowledge of dialects that we do or do not speak?).

2 Background

2.1 Perception, Prediction, and Processing

Bountiful neurolinguistic evidence shows that people invoke prediction during sentence processing (Van Berkum et al. 2008, Hanulikova et al. 2012, Kutas et al. 2014). Electroencephalography (EEG) is one method used by neurolinguists to look at this processing. EEG measures electric potentials that are generated by tens of thousands of cortical neurons using electrodes placed on the scalp. Averaging the EEG signal that is recorded to multiple instances of a specific perceptual event reveals systematic voltage changes associated with the cognitive processes elicited by that event, called the Event-Related Potential (ERP). EEG is useful to linguists because it allows researchers to passively monitor neural activity which reflects implicit and on-line linguistic judgements, including the social expectations of listeners. It also illuminates for researchers when, in real time, expectations on the part of the listener may be violated during processing, indicating that prediction is taking place. A great deal of prior work has revealed ERP signatures for semantic and syntactic violations (Luck 2005). For example, if something is semantically unexpected in a sentence string, a listener will exhibit an increase of negative voltages over the central scalp that peaks around 400 milliseconds after word onset (the "N400"). This has been taken to indicate that the speech signal was processed as a semantic anomaly within 400ms of the onset of the stimulus. When a particular morpheme of a sentence violates a syntactic or grammatical expectation, studies reveal a positive voltage spike over the posterior scalp peaking around 600ms (the "P600"), although this late component can be variable in both onset latency and duration.

2.2 The Influence of Speaker Identity in ERPs

Van Berkum et al 2008 used the N400 response to investigate the influence of speaker identity and semantic anomalies in Dutch. The semantic anomaly sentences they used included a word that made the sentence syntactically well-formed but semantically implausible (e.g. "Dutch trains are sour and blue." vs. "Dutch trains are yellow and blue."). The speaker inconsistency sentences were otherwise well-formed sentences that were produced by unexpected actors; a male saying something coded as female, someone who sounded upper class saying something that was coded as lower middle class, and a young child saying something that would be coded as only appropriate for adults to say. An example of this kind of sentence is, "I like a glass of wine before bed," spoken in a child's voice (unexpectedly) versus spoken in an adult voice (more expectedly). They found that Dutch listeners showed N400s for semantic anomalies and also speaker inconsistencies, albeit smaller ones for the latter. These results show that speaker identity can be taken into account as early as 200-300 milliseconds after the beginning of the anomalous word. This research supports the hypothesis that people take in perceived speaker information rapidly when processing sentences.

Hanulikova et al. 2012 built on this work by testing how native listeners process grammatical errors that are frequent in non-native speech. Generally, they found that a P600 effect for grammatical violations made by the Dutch-Accented Dutch speaker but did not find such an effect for those same grammatical violations spoken by the Turkish-Accented Dutch speaker. The researchers attributed this result to listeners altering their grammatical expectations depending on how native they perceived the speaker to be of the language that they are using. In addition to using syntactic anomalous stimuli, Hanulikova and colleagues also included semantic anomalous sentences which showed that the N400 effect was present and equal for both accent conditions. This indicates that the semantic content of the foreign-accented speech conditions was processed on par with the native-accented speech. In this case, the grammatical expectations of listeners seemed to be altered based on speaker.

Previous pilot work in the University of Michigan Computational Neurolinguistics Lab extends this finding from Hanulikova et al. 2012 to speakers of non-standard varieties of English. Listeners heard sentences with auxiliaries present and absent in three different varieties of American English: MUSE, AAL, and Indian English (IE). The grammatical phenomenon of "auxiliary dropping" is a feature of African American Language (AAL) (e.g. My brother, {he is/he's/he} working today). The auxiliary absent condition, (e.g. "My brother, he working today,"), like many aspects of AAL, is a stigmatized feature, despite being grammatical in the language variety. Listeners showed that listeners showed a P600 response to the non-standard utterance from the MUSE speaker, but not for the AAL speaker or for the speaker of IE, which is a non-standard variety that does not share the stigmatized grammatical feature. Apparently, listeners altered their grammatical expectations for non-standard speakers, but not for the MUSE speaker.

2.3 Research Questions

While there is now evidence that listeners modulate their expectations with respect to grammar of other speakers, prior research does not indicate whether listeners have specific knowledge of the dialect that is not their own, or whether listeners more generally reduce expectations across the board when listening to a dialect or variant that they themselves do not speak. We present an EEG study designed to test between these hypotheses. To preview our results, we do not see evidence cleanly favoring one, or the other view. Rather, the evidence points a nuanced version of a mixture of both hypotheses.

3 Methods and Predictions

3.1 Participants

Twenty-nine Ann Arbor residents (17 female, 12 male, mean age = 22 years, range 18-45 years) participated in the current study. All participants were right-handed based on a Handedness Survey (Oldfield 1971), had normal or corrected-to-normal vision, and no history of any neurological disorders. Participants were all native speakers of varied American English dialects, including native AAL speakers, and that English variety was their first language. All participants gave their informed consent and were compensated fifteen dollars per hour for their time and participation. All experimental protocols are in compliance with and underwent review by the Institutional Review Board at the University of Michigan, IRB # HUM00075912.

3.2 Materials, Stimuli Construction, and Creation

In order to test the interaction between language variety and auxiliary usage, experimental sentences were constructed in order to reflect a variant that is grammatical in MUSE, a variant that is grammatical uniquely to AAL, and a grammatical variant that is ungrammatical in all varieties of English. We focus on the grammatical auxiliary "be," which must be overt in combination with the progressive aspect in MUSE but may be omitted in AAL. Alternative auxiliaries like "will" are disallowed in both varieties. Thus, three conditions were created: (1) Auxiliary Present, (2) Auxiliary Absent, and (3) Ungrammatical Variant "will." Examples are shown in the right-hand column of Table 1.

All experimental sentences were produced by a bidialectal Midwestern African American male who speaks both MUSE and AAL. The choice to use one speaker was motivated by previous research indicating that multidialectal speakers can be assigned different racial, regional, and even attitudinal impressions depending on the guise or language variety employed (Purnell, Baugh, Idsardi 1999, Lambert et al. 1960).

The stimulus creation process is schematized d in Table 1. The leftmost column shows examples of three sentences that the speaker actually produced, the middle column illustrates how the recordings were spliced together, and the rightmost column displays the result audio file which corresponds to the three conditions described above. The speaker recorded a total of 658 sentences, with 108 sentences produced in both AAL and MUSE, in both the auxiliary present condition and the auxiliary absent condition. Importantly, to create the ungrammatical condition, the speaker was not asked to produce ungrammatical sentences, as this may yield incongruent prosody. Thus, he produced 108 grammatical sentences with the "will" construction across both varieties (i.e. "The clown we hired, he'll blow up balloons tomorrow"). These sentences were then spliced back into the critical region of the host sentences, to create the ungrammatical condition sentences. In order to construct carefully controlled sentences with similar acoustic contexts to the other naturally produced sentences, every stimulus was spliced so that each of the target regions were always edited. Thus, as seen in Table 1, even the auxiliary present condition was edited and spliced, in order to maintain consistency in splicing all conditions. Each of the "will" construction sentences included a voiced or voiceless bilabial, alveolar, or velar stop that followed the "will" construction region, which then resulted in sentences that sounded seamless in construction and prosody, despite being ungrammatical. Each sentence had the critical region an average of six words into the sentence.

	Sentences Produced	**Sentence-Splicing**	**Result Sentences**
(1)	He got sick, so he's coughing a lot.	"He got sick, so he's" from (2) spliced into (1) before "coughing"	He got sick, so he's coughing a lot.
(2)	He got sick, so he's coughing a lot.	"He got sick, so he" from (2) spliced into (1) before "coughing"	He got sick, so he coughing a lot.
(3)	He got sick, so he'll cough all day most likely.	"He got sick, so he'll" from sentence (3) spliced into (1) before "coughing"	He got sick, so he'll coughing a lot.

Table 1: Stimulus Creation Process.

The speaker was recorded in a sound-attenuated booth at the University of Michigan, using an AKG C4000B condenser microphone and an Edirol UA-25 audio interface, at a sampling rate of 44,100 Hz. The speaker was paid $20 per hour for his time and participation. Experimental sentences were distributed across two lists, each with 108 sentences, resulting in 216 sentences across all conditions, distributed using a Latin Square design.

3.3 Procedure

Participants arrived to the lab and signed a consent form, followed by a Handedness Survey. After being fit with the EEG cap, participants were seated about 100 cm in front of a computer screen and electrolyte gel was applied to minimize impedances between each electrode and the participant's scalp. Electrodes were also placed on the inside of the right wrist, and above and below the left eye to monitor heartbeat and eye blinks respectively. The participant was then fitted with two in-ear earphones (Etymotic Inc. EA-2). Sound levels were set to 45 dB above each individual's hearing threshold, assessed using 1000 Hz tones (300 ms, 100 ms fade in/out). This was followed by a two-minute quality check test, in which the participant was instructed to sit still and stare at a fixation cross on the screen while listening to 120 1000 Hz tones. EEG data were visually inspected to ensure low noise in the data before moving on to the main experiment.

The participant was told that they would be listening to sentences spoken by a few different people for 15-20 minutes, answering occasional yes/no comprehension questions, and they would have periodic breaks. The stimuli were presented by using E-Prime software. After each stimulus, there was a one in four chance of seeing a comprehension question. For example, if the stimulus sentence was, "The clown we hired, he blowing up balloons for the kids," the following question might appear on the screen: "Did the clown blow up balloons for the kids?" The participant would click "y" for yes or "n" for no on the keyboard depending on the answer. This was to keep their mind on a task separate from the true one of the main experiment. Participants listened to 216 items across all conditions. After the task, the participant was asked to fill out a Post Experiment Questionnaire which details parts of their language exposure history, although those results are not included in this paper. The total time from start up to clean up took around 1-1.5 hours per participant.

3.4 EEG Recording

EEG was recorded with an elastic cap with 61 actively-amplified electrodes and one ground electrode (acti-Cap, Brain Products GmbH). Electrodes were distributed equidistantly across the scalp according to the Easycap M10 layout. The electrode impedances were kept at 25kΩ or below. Data were recorded at 500 Hz between 0.1 and 200 Hz referenced to an electrode placed on the right mastoid (actiCHamp, Brain Products GmbH). The electro-oculogram (EOG) was recorded from electrodes placed above and below the left eye.

3.5 ERP and Statistical Analyses

Data processing was conducted using Fieldtrip toolbox in MATLAB (Oostenveld et al 2011). Raw EEG data were re-referenced to the left and right mastoid electrodes, high-pass filtered at 0.1 Hz. The data were then each divided into time-locked epochs spanning 0.3 – 1 sec around the onset of the gerund, "-ing" or "'in," in each sentence of the six conditions (i.e. if the sentence was "The clown we hired, he'll blowing up balloons tomorrow," we identified the epoch onset with the end of the first syllable of "blowing" to 10 milliseconds after). The time-locking point is indicated in bold in Table 2. Ocular artifacts were removed using Independent Component Analysis (Jung et al 2000), and remaining artifacts were identified and removed following visual inspection and inter-rater reliability. At most, 20% of epochs were marked as artifactual from each participant, which left over 170 trials per participant. Signals from the electrodes with supra-threshold impedance or exceptional noise were replaced using surface spline interpolation (Perrin et al 1987). Each epoch was low-pass filtered at 40Hz. Baseline correction was applied with a window of -0.1 – 0 sec. To test for the P600 effects we compute the average voltage potential, per subject, across central posterior sensors in a time window from 600 to 900ms. We submitted these values to a 2-way repeated measures ANOVA, with language variety (*MUSE, AAL),* and auxiliary in the model (*present, absent, ungrammatical)* as predictor variables.

3.6 Predictions

Predictions for the study are shown in Table 2. It was predicted that if listeners form dialect-specific expectations, the presence of the ungrammatical "ll" feature (row 3) should elicit a P600 response when hearing both MUSE and AAL, whereas auxiliary deletion (row 2) should elicit a P600 in MUSE, but not in AAL. Alternatively, if listeners form non-specific predictions, meaning that listeners group all non-standard dialects into an "other" category with relaxed grammatical expectations, neither row 2 or 3 should show a P600 for AAL speech.

	Example	Acceptability	dialect-specific prediction	non-specific prediction
1	I don't know why she is blushi**ng** so hard.	MUSE, AAL	No P600	No P600
2	I don't know why she __ blushi**ng** so hard.	AAL	P600 for MUSE	P600 for MUSE
3	I don't know why she'll blushi**ng** so hard.	None	P600 for both	No P600 for AAL

Table 2: The predictions for this experiment.

4 Results

Figure 1 shows the grand average of ERPs across the scalp at the time-locked point, at 600-900ms. The MUSE data show the auxiliary present (blue lines) and absent (red lines) conditions do not elicit a P600, but the ungrammatical auxiliary (green lines) does elicit a P600 of about 1 microvolt. There is not any evidence for a P600 in the AAL condition; this is shown in the right-hand panel of Figure 1. A 2-way repeated measures ANOVA showed a main effect of language variety, MUSE versus AAL, $F(1,30) = 58.25$, $p < 0.01$, but no main effect of auxiliary, $F(2, 60) = 7.70$, $p = 0.624$, and no significant interaction, $F(2,60) = 16.80$, $p = 0.337$. This replicates the AAL result from the pilot work in our lab which did not show an effect for varied auxiliaries. While the pattern for AAL is consistent with "non-specific" dialect predictions, the MUSE results present a more nuanced picture. These are discussed in more detail below. These results go against both the dialect-specific prediction and the non-specific prediction. Possible reasons for this are discussed below.

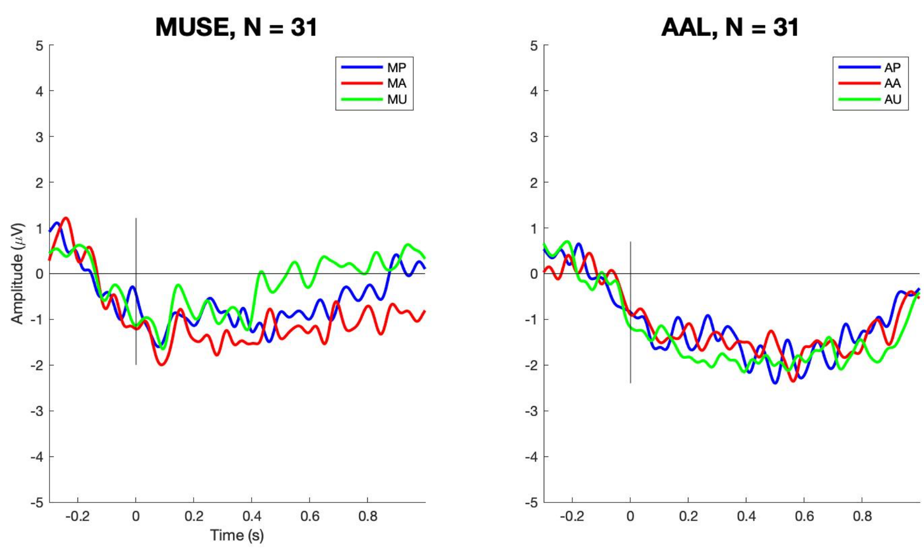

Figure 1: Event-related potentials from the centro-posterior electrodes, time-locked to the onset of the "-ing/-in" segment, are shown separately for the MUSE stimuli (left) and the AAL stimuli (right). Red traces indicate the variant that is grammatical in both dialects; blue also indicates the variant that is grammatical in AAL but not MUSE. Green indicates the variant that is ungrammatical in both MUSE and AAL.

5 Discussion

EEG signals collected during the experiment indicate that listeners expectations reflect a kind of dialect-specific expectations with respect to speaker, but not with respect to auxiliary. For the MUSE speaker, the data show an increased positivity for the "ungrammatical" condition, relative to the "auxiliary present" condition, and an increased negativity for the "auxiliary absent" condition. The 2-way ANOVA indicates that participants are processing these two dialects differently, however the statistics do not support connecting this difference in processing to the auxiliary variation; although, there is a visually-apparent pattern such that the MUSE results show a P600 when listeners heard an ungrammatical stimulus (e.g. "The clown, he'll blowing up balloons at the party."), but no P600 was observed for either of the other auxiliary conditions (e.g. "The clown, he's blowing up balloons at the party," and, "The clown, he blowing up balloons at the party.").

One explanation for the results is that the listeners are doing a certain kind of dialect work with the bidialectal speaker. It is possible that the use of one speaker across both language varieties might have caused some confusion on behalf of the listeners. Despite pretesting the speaker's voices, the phonology and prosody of AAL was sometimes unavoidably present in some of the MUSE condition speech. This could have been ambiguous or unclear to the listener. While the AAL results do not show a difference in effect across the auxiliaries, the MUSE speech does show a classic P600 for the ungrammatical auxiliary. This result suggests the possibility of dialect work going on, on behalf of the listener, granting acceptability for the auxiliary absent condition while still granting that the ungrammatical condition was unallowable. This leads to the question of what it means to be a "Black-Accented Standard" speaker in the United States, and how are those speakers received? Overall, the results from this study are consistent with the specific result from Hanulikova et al. 2012 which showed that listeners altered their expectations based on "foreign-soundingness," in this case, "non-standard-soundingness." of the speaker. We now see that these effects may reflect some kind of dialect-specific knowledge as well. Due to hearing MUSE syntax and AAL prosody bound up in single phrases, there may have been a mismatch of expectation versus input on behalf of the listener.

6 Conclusions

The way in which we process speech based on expectations tells us something about normativity. When listeners are processing "anomalies," we, as researchers, are seeing disruptions in that normativity. Looking at the results of this study, we see two kinds of violation expectation. Firstly, we see a clear difference in how listeners are processing MUSE and AAL produced by a bidialectal speaker, which indicates to us that people can discern language varieties even intralinguistically (from the same speaker). Many speakers who do not control AAL will not be able to produce grammatical utterances in AAL, and the focus on online grammatical processing allows us insight into speaker's passive knowledge of dialect knowledge. Secondly, it seems as though that when normative expectations for a black voice were violated (Black speaker using MUSE), listeners are unwilling to grant dialect fluidity to the listener. The results from this study highlight the impact of doing neurolinguistics work on dialects, and within a sociolinguistic frame. Language lives in the world, and by investigating bidialectal speaker, we opened a nuanced perspective to language variation and perception. Ultimately, perceptions of stigmatized languages and language varieties leads to language discrimination, which affects the way speakers, people, are treated in their day to day lives. Through a multi -method neurolinguistic and sociolinguistic approach, we can better understand how the human language faculty is capable of recognizing and processing dialects.

References

Eckert, Penelope, and John R. Rickford. 2001. *Style and sociolinguistic variation*. New York: Cambridge Uni versity Press.

Hanuliková, Adriana., Petra.M. van Alphen, Merel M. van Goch, Andrea Weber. 2012. When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience* 24: 878–887.

Hill, Jane. 2008. *The Everyday Language of White Racism*. Massachusetts: Wiley-Blackwell Publishing.
Jung, Tzyy-Ping, Scott Makeig, Marissa Westerfield, Jeanne Townsend, Eric Courchesne, Terrence J. Sejnowski. 2000. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology* 111:1745–1758.
Kutas, Marta, Kara D. Federmeier, Thomas P. Urbach. 2014. The 'negatives' and 'positives' of prediction in language. *The Cognitive Neurosciences* 5: 649-656.
Lambert, William R., Richard Corrin Hodgson, Richard C. Gardner, Samuel Fillenbaum. 1960. Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology* 60: 44-51.
Luck, Steven. J. 2005. *An introduction to the event-related potential technique.* Massachusetts: MIT Press.
Oldfield, R. C. 1971. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9:97–113.
Oostenveld, Robert, Pascal Fries, Eric Maris, Jan-Mathijs Schoffelen. 2011. Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* 1(1-9).
Perrin, F., J. Pernier, Oliver Bertrand, Marie-Hélene Giard, J.F. Echallier. 1987. Mapping of scalp potentials by surface spline interpolation. *Electroencephalography Clinical Neurophysiology* 66: 75–81.
Purnell, Tomas, William Idsardi, John Baugh. 1999. Perceptual and Phonetic Experiments on American Eng lish Dialect Identification. *Journal of Language and Social Psychology* 18: 10-30.
Rickford, John R. and Sharese King. 2016. Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language* 92: 948-988.
Rickford,John. 1999. The Ebonics controversy in my backyard: A sociolinguist's experiences and reflections. *Journal of Sociolinguistics* 3: 257-266.
Schieffelin, Bambi, Kathryn Ann Woolard, Paul V Kroskrity. 1998. *Language ideologies: practice and theory*. New York: Oxford University Press.
Schilling, Natalie. 2004. Constructing Ethnicity in Interaction. *Journal of Sociolinguistics* 8: 163-195
Van Berkum, Jos. J. A., Danielle van den Brink, Cathelijne M. J. Y. Tesink, Miriam Kos, and Peter Hagoort. 2008. The neural integration of speaker and message. *Journal of Cognitive Neuroscience* 20: 580–591.

Computational Neurolinguistics Lab
Department of Linguistics
University of Michigan
Ann Arbor, MI 48109-1220
racheliw@umich.edu
jobrenn@umich.edu