Language in Society, page 1 of 30.
doi:10.1017/S0047404518001112

Genre and linguistic expectation shift: Evidence from pop song lyrics

LAUREN SQUIRES

The Ohio State University, USA

ABSTRACT

Popular song lyrics constitute an exception to dominant, standard language ideologies of English: nonstandard grammatical forms are common, relatively unstigmatized, and even enregistered in the genre. This project uses song lyrics to test whether genre cues can shift linguistic expectations, influencing how speakers process morphosyntactic variants. In three self-paced reading experiments, participants read sentences from pop songs. Test sentences contained either 'standard' NP_{SG} + *doesn't* or 'nonstandard' NP_{SG} + *don't*. In Experiment 1, some participants were told that the sentences came from lyrics, while others received no context information. Experiment 2 eliminated other nonstandardisms in the stimuli, and Experiment 3 tested for the effect of stronger context information. Genre information caused participants to orient to the sentences differently, which partially—but not straightforwardly—mitigated surprisal at nonstandard *don't*. I discuss future directions for understanding the effects of context on sociolinguistic processing, which I argue can inform concepts like genre and enregisterment, and the processes underlying language attitudes. (Morphosyntactic variation, genre, invariant *don't*, language ideology, pop songs, experimental sociolinguistics, sentence processing)*

INTRODUCTION

This article is broadly concerned with the links between sociolinguistic variation, language processing, and speakers' knowledge of the links between speech forms and speech contexts. There has of late been a burgeoning interest in *sociolinguistic knowledge*: What do speakers know implicitly about sociolinguistic variation? (Labov, Ash, Ravindranath, Weldon, Baranowski, & Nagy 2011; Campbell-Kibler 2016). Considering knowledge of dialect variation (e.g. Wolfram 1982), more specific questions emerge: What happens when speakers encounter unfamiliar linguistic forms—do they store them as part of their knowledge, or discard them? If they store them, how do they categorize them along with existing knowledge? Are all variants processed 'equally'? Does social evaluation affect how a variant is stored in linguistic knowledge? Is linguistic knowledge 'indexed' with social information? If so, is this the case at every level of linguistic representation?

Is it different in kind across levels of structure? When is specific sociolinguistic knowledge recruited during language production and comprehension? And so on.

The overarching question for the present study is: How does social context contribute to how morphosyntactic dialect variants are processed and comprehended? This question relates to experimental work investigating the social evaluation of morphosyntactic variants (Squires 2013; Hesson & Shellgren 2015; Levon & Buchstaller 2015), but is about the inverse relationship—not how social perceptions are triggered by variants, but how the processing and comprehension of variants is affected by social information. While work on the influence of social/contextual information on speech (phonological) production and perception is now robust (see recently Sanchez, Hay, & Nilson 2015; Hay, Podlubny, Drager, & McAuliffe 2017), research into morphological or syntactic processing/perception remains limited to a few studies (Squires 2013, 2014a,b, 2016; Weatherholz, Campbell-Kibler, & Jaeger 2014; Seifeldin, Cantor, Boland, & Brennan 2015).

This study offers more evidence that sentence processing can be influenced by social factors, and also expands the range of social factors pertinent during processing. Here, the social factor at play is speech genre—namely, popular song lyrics. I argue that in Anglophone mass culture, the lyrics of popular music are construed so as to render them largely immune to the evaluative metrics of the dominant operating 'standard language ideology' (see Milroy & Milroy 1999; Milroy 2001). As discussed further below, in part due to the value placed on linguistic 'authenticity' in popular subgenres of music (such as country and hip-hop), song lyrics are *ideological exceptions*, within which nonstandard structures may go unnoticed or unstigmatized. The premise of the present research is that we should be able to find, experimentally, corollaries of this exceptionalism in the form of processing behavior. One additional aim of this article is thus to make explicit connections between now-commonplace concepts within sociolinguistics and linguistic anthropology and those within the newer interdisciplinary field of experimental sociolinguistics.

In what follows, I first offer relevant background on sentence processing, focusing on linguistic expectation as fundamental to processing. I then discuss the genre of pop songs, reviewing work showing that pop songs carry characteristics that mark them as ideologically 'exceptional', and I articulate how we might expect this exceptionalism to manifest in language-processing behavior. The discussion of three experiments follows, and then a concluding discussion raises possibilities for future directions.

PROCESSING, EXPECTATION, AND SOCIOLINGUISTIC KNOWLEDGE

A basic finding of decades of experimental research on sentence processing is that processing is disrupted—slowed down—by syntactic ill-formedness, structural unfamiliarity, and semantic implausibility. These effects have been measured most

commonly through reading times, eye gaze, and brain activity. Very few sentence processing studies have taken dialect variation as their focus, but it is reasonable to hypothesize that the processing of such variation may be subject to different influences than, for instance, syntactic alternatives that are not socially conditioned (see discussion in Squires 2014b).

The primary set of psycholinguistic studies testing dialect variation involves the 'needs washed', or NEED + PAST PARTICIPLE construction, common in some dialects of North American English. Kaschak & Glenberg (2004), Kaschak (2006), and Boland, de los Santos, Carranza, & Kaschak (2015) have shown that for participants unfamiliar with this dialect feature, the construction at first causes processing difficulty, but that difficulty dissipates through more exposure—similar to syntactic satiation (the more you hear/read something, the more grammatical it seems; see Luka & Barsalou 2005). The researchers have interpreted this effect as either participants learning the new construction in an abstract sense (Kashak & Glenberg 2004; Kaschak 2006), or simply relaxing their grammatical constraints during the processing task (Boland et al. 2015). In either case, the initial processing difficulty comes about because the structure lies outside of a person's existing linguistic knowledge.

But what is the envelope of what is known by speakers? Is all that is known reflected in production, or do speakers also have knowledge of structures that they do not produce? Specifically in terms of grammatical (morphosyntactic) features, earlier work like Labov (1973) and Wolfram (1982) posed these questions, but there has been little exploration of the issues since then. I have elsewhere provided evidence that processing behavior shows a distinction between grammatical forms that are totally unknown to a speaker and those that are not used by the speaker, but known ABOUT—that is, experienced even if not produced (Squires 2014a). This evidence was in the form of a cline of processing disruption: an unattested form elicited longer processing time than a nonstandard dialect form, which elicited longer processing time than a standard form. Other-dialect forms that are familiar to speakers may constitute part of their implicit sociolinguistic knowledge, and these forms may be treated differently in processing than completely unfamiliar ones.

In part as a reflection of this sociolinguistic knowledge, processing is not immune to context. Again, while not examining dialect variants, psycholinguistic studies have found evidence that the activation of linguistic knowledge is mediated by social information pertinent to a processing situation. Kamide (2012) and Yildirim, Degen, Tanenhaus, & Jaeger (2016) show talker-specificity effects for syntactic structural alternatives (see also Squires 2014b): participants connect specific structures with specific speakers. Van Berkum, van den Brink, Tesink, Kos, & Hagoort (2008) and Tesink, Petersson, van Berkum, van den Brink, Buitelaar, & Hagoort (2009) show that listeners integrate social beliefs about speakers, including their age and sex, with the semantic content of their sentences.

Recent studies of event-related potentials (ERP)—which indicate electrical brain responses to stimulus events—provide especially compelling neurolinguistic

evidence of the relevance of context to processing. Linguistic ERP effects are discussed as 'components' labeled by several factors: whether the brain activity consists of positive (P) or negative (N) voltage, distribution of the activity across areas of the brain, and/or the time-course of the activation in milliseconds. For instance, the P600 is the component that has been shown to mark encounter with a syntactic anomaly/ungrammaticality (P for positivity; 600 for milliseconds) (for an overview see Steinhauer & Connolly 2008). Hanulíková, van Alphen, van Goch, & Weber (2012) found that when Dutch-speaking listeners heard Dutch speech with grammatical gender errors, they did not exhibit a P600 effect when the speaker had a nonnative accent, yet there was a P600 effect when the speaker sounded like a native Dutch speaker. Similarly, Hanulíková & Carreiras (2015) found that speaker gender mediated the effects of grammatical gender-related errors in Slovak. These studies demonstrate that when listeners believe that a certain person or type of person is talking, their comprehension system adjusts itself for the input.

Adaptation based on the social properties of a talker is essentially also the finding of Seifeldin and colleagues (2015), who exposed listeners to variable copula deletion—a central feature of African American English dialects. When listeners thought they were listening to an African American speaker, they did not show the same P600 effect as when they thought they were listening to a White speaker. However, the mitigation also occurred when listeners thought they were hearing an Indian (South Asian) speaker, whose dialect would not include copula deletion. Having social information about the speakers facilitated processing the less-familiar grammatical structures: the social information elicited a general accommodation to nonstandard (or simply unexpected) structures. Seifeldin and colleagues' work highlights the important of EXPECTATION in sentence processing. There is mounting psycholinguistic evidence about the general importance of expectation/prediction in language processing (for a review, see Kuperberg & Jaeger 2016). Listeners (or readers) take stock of contextual factors, which inform our perceptual expectations, which then influence how we process the raw linguistic material that we encounter. Things that are expected are processed more easily, while things that provide new and/or unexpected information produce SURPRISAL, and typically a processing cost (Jaeger & Weatherholtz 2016).

Expectation and surprisal are probabilistic, functions of linguistic factors like prior experience (frequency) and recency of exposure (priming). But increasingly it is clear that what have traditionally been considered nonlinguistic factors also shape linguistic expectations. The articles reviewed above point to the relevance of social speaker factors such as age, sex/gender, and race/ethnicity. The present project adds to our understanding of the kinds of contextual information that may shape linguistic expectations—moving from the social properties of speakers to the social properties of context itself. I use pop songs as a case study for speech genre as a form of sociolinguistic context, to test whether information about speech genre can modulate speakers' expectations for morphosyntactic variants.

ENGLISH POP SONGS AS A SPEECH GENRE

Linguistic variation recurs across contextual dimensions aside from speaker. *Speech genre* is one such contextual dimension, and of course has been a longstanding subject of theoretical attention (e.g. Bakhtin 1986; Briggs & Bauman 1992; Giltrow & Stein 2009). I use a basic definition of genres as *recognizable categories of speech*. Hymes (1974:61), for instance, lists 'categories such as poem, myth, tale, proverb, riddle, curse, prayer, oration, lecture, commercial, form letter, editorial, etc.'. Hymes notes that genres are often associated with *speech events*—he gives the example of a sermon in a church service—but genres are also (usually) detachable from the events themselves, able to be called upon in other events or circumstances (hence the use of a sermonic tone in political speeches); on detachability, see also Briggs & Bauman 1992.

While genres connect to other contextual elements—setting, participants, and event structure—they also have internal linguistic markers. Genres are patterns of co-occurrence between social/speech context and linguistic forms (Ferguson 1994; Biber & Conrad 2009), with recognizability as a key metalinguistic property (Biber 1993). As corpus analyses have shown, genres are partially distinguished by feature co-occurrence. Some of these features cross linguistic varieties/dialects (like pronouns or tense categories), yet it is also the case that language varieties can in and of themselves be generic markers. For instance, the academic written English of publishing is by and large Standard English (with differences arising on national dimensions), while one finds a greater mix of dialects in English literary genres like poetry and fiction. Likewise, the spoken genre of interpersonal conversation is one in which native dialect is typically favored (depending on the interlocutors), while the genre of an academic lecture again favors Standard English. Individual features or constructions may also be *enregistered* (Agha 2007) as part of a genre: for instance, the (not otherwise socially marked) opening phrase 'In a world…' as it connects to the genre of movie trailers, or 'Once upon a time…' as it connects to the genre of children's stories.

Coupland (2011) reminds us to consider genre as not only IN THE TEXT (with text meant broadly to encompass speech as well), but also IN A SPEAKER'S MENTAL MODEL OF THE SPEECH SITUATION. Coupland cites Hanks (1987:670) and his characterization of genres as 'orienting frameworks', mediating between linguistic and social structures. So what is the mental reality of a sociolinguistic genre? Since people infer genre from linguistic material, it is reasonable to hypothesize that things also work the other way around: generic category may inform one's expectations of linguistic material.

To test this idea, I use English-language popular song lyrics, a genre that has the fundamental properties of recognizability and enregistered varietal features. Pop songs constitute a casual (though not necessarily 'natural') genre in which nonstandard features of various kinds do not seem to negatively affect songs' reception. In being exempt, at least to some extent, from the pressures of standard language

ideology (Milroy 2001), song lyrics seem to be treated ideologically differently than other more formal genres (and even many informal ones). Features that in other contexts garner explicit stigmatization appear robustly in pop songs of multiple subgenres—songs that are enjoyed, consumed, and even embodied by diverse audiences.

The ideological exceptionalism of pop songs would seem to stem from the commodification of 'authentic' vernacular language within subgenres that derive from reified cultural groups. Characterizing popular music as a fundamentally *vernacular* genre, Coupland (2011) points out that it is a mistake to treat song lyrics as 'pure' linguistic texts; they are performances, and the language within them is often therefore stylized. As such, it is possible that song lyrics overrepresent vernacularity relative to singers' actual dialects. There are clearly intentional cases of artists performing dialect variation for purposes of genre-fitting (Trudgill 1983; Eberhardt & Freeman 2015; Duncan 2017). Such performances only support the notion that dialect features are enregistered components of the genre. As Coupland (2011:595) puts it, '[d]ialectal vernacularity… is one meaningful resource in popular music performance'.

Through exposure to pop music, I suggest, speakers form sociolinguistic links between the genre (more precisely, the subgenre) and the forms likely to occur within it. Moreover, as Murphey (1992) notes, song lyrics have a way of anchoring themselves in memory, which he calls the 'Song-stuck-in-my-head phenomenon' (SSIMHP). There is no evidence that lyrics with nonstandard content are immune to this phenomenon (this would make a fascinating study). Indeed, in discussions about using song lyrics as teaching tools for English second-language learning (Murphey 1990, 1992), one finds competing arguments: On the one hand, lyrics' extreme memorability makes them useful targets for certain structures. On the other hand, some consider their language to be too nonstandard, including their 'grammatical ill-formedness', to be appropriate (Terhune 1997, cited in Kreyer & Mukherjee 2007:32). I suggest that, especially for speakers with more or less monodialectal Standard English repertoires, song lyrics may be a primary genre through which they encounter forms outside of their own productive knowledge, and by which those forms might even 'get in their heads'.

The generic features of song lyrics have been investigated in several corpus analyses looking across subgenres like rock, hip-hop, R&B, metal, pop, and country. As texts, lyrics tend to be repetitive with small vocabularies, giving them a low type-token ratio (Murphey 1992; Kreyer & Mukherjee 2007; Werner 2012; Bértoli-Dutra 2014). Lyrics have an especially high proportion of function words (pronouns, articles, determiners) as well as specific keywords related to their emotive themes. For example, Murphey (1992) finds that twenty-five percent of his total corpus is comprised of function words and the word LOVE.

Among these broad findings are mentions of dialect features, including nonstandard spellings representing phonetic pronunciations (Kreyer & Mukherjee 2007). In terms of morphosyntax, Werner cites the presence in lyrics of *ain't*, copula

deletion, invariant *don't*, and double negation as markers of both Americanness and African Americanness, but more generally of a 'colloquial or informal style' (2012:28). Of course, several researchers have discussed the role of specific place- or race-marked varieties in the music of specific artists, including Trudgill (1983), Simpson (1999), Beal (2009), Eberhardt & Freeman (2015), and Duncan (2017).

This work all points to a permissiveness of the pop song genre: lyrics not only accommodate but in some cases privilege 'nonstandard' forms, including vernacularisms, regionalisms, and idiolectal or poetic uses. In other words, the dominant ideology that attaches positive attributes to 'standard' and negative attributes to 'nonstandard' features (almost by definition) seems to be put on pause, so to speak, when listening to pop music. This is not paradoxical; rather, it is the logical outcome of how subgenres of pop music have come from certain 'vernacular' communities, and have been commodified and transported out of those communities (Coupland 2011; Eberhardt & Freeman 2015; Duncan 2017). Linguistic markers of the communities themselves travel with the subgenres, becoming enregistered as 'authentic' components of them.

One nonstandard grammatical feature that is common in song lyrics—perhaps because it is a general vernacularism, crossing multiple dialects—is invariant *don't* (mentioned in Werner 2012), wherein the negative auxiliary *don't* may appear regardless of the number of the subject. For a speaker with this feature, all of the following cases in (1) would be allowed for. These examples come from recording artist Justin Bieber, where the 'nonstandard' instance is (1a), with a third-person singular subject.

(1) a. My mama don't like you and she likes everyone
('Love yourself'; Sheeran, Blanco, & Bieber 2015)
b. You know I try but I don't do too well with apologies
('Sorry'; Bieber, Michaels, Tranter, Moore, & Tucker 2015)
c. Bring the doubters on—they don't matter at all
('Never let you go'; Bieber, Austin, & Cox 2010)
d. When you don't want me to move
('What do you mean?'; Bieber, Boyd, & Levy 2015)
e. Your love doesn't go unrecognized
('We were born for this'; Bieber & Hook 2014)

Note that, as with other linguistic variables, the nonstandard and standard forms can be in alternation even for the same speaker, as in (1e) above.

The presence of both (1a) and (1e) in the lyrics of the same recording artist may be attributable to different songwriters across tracks, and it is certainly possible that Bieber is not a native user of invariant *don't* (I suspect the variation has to do with the scripting of invariant *don't* in certain moments of song but not others for reasons of style, meter, or both). In any case, the listener/consumer of the song in (1a) has

now been exposed to invariant *don't*, and if they begin to sing the song (perhaps through the unwitting SSIMHP), they internalize it (the degree to which this happens is an empirical question). For a hit song like those of Bieber, the exposure to audiences will likely cross those who both have invariant *don't* as a part of their native dialect, and those who do not.

Because of my prior work with the processing of invariant *don't* (Squires 2013, 2014a, 2014b), this is the feature I use as a test case in the present study. Interestingly, Kreyer & Mukherjee (2007) found that *don't* was the twenty-third most frequent word in their corpus of lyrics. *Doesn't* was not in the top words; *don't* was the only negative contraction and the only form of *do* in the list. Logan, Kositsky, & Moreno (2004) also found that the word *don't* was among the top ten most frequent words for country, rap, and rock subgenres of their corpus. *Doesn't* and *didn't* were not among the top words for any subgenre. The abundance of *don't* in these corpora may partly be due to the fact that second- and first-person pronouns are more common in lyrics than third-person ones (Murphey 1992), but I assume it is also to do with at least some use of *don't* as compared to *doesn't* for singular third-person subjects. If this is true, the fact that invariant *don't* occurs across subgenres highlights that invariant *don't* is used across vernacular English dialects, cross-cutting racial/ethnic associations (though it may also be racialized through its association with AAE; see Werner 2012). The cross-variety status of invariant *don't* is beneficial for the present research, since its presence alone does not index a specific subgenre of music or a specific category of social persona.

The experiments presented below attempt to get at the mental substance of the 'ideologically exceptional' property of song lyrics. When a listener is consuming song lyrics, is she altering her expectations for what kind of linguistic forms she might encounter, such that she becomes less surprised by a variety of features? In exploring speech perception, Gibson (2010:147) found that playing vowel tokens with background music caused a 'perceptual style shift' in how participants categorized the vowels relative to a nonmusical condition. Similarly, Experiment 1 was motivated by the question of whether expectations for *grammatical forms* can be shifted along with the expectation for song lyrics.

EXPERIMENT 1

Methods

Experiment 1 used a moving window self-paced reading design, prepared and administered using Paradigm software's built-in self-paced reading tool.[1] In this setup, at the start of each trial the participant sees a continuous line corresponding to the length of the sentence, centered on screen. When the participant presses the response button, the first word in the sentence appears (in its sentence position, so on the left-hand side of the screen). Each time the participant presses a button, the previous word disappears and the next word appears, in a left-to-right fashion.

Two experimental manipulations tested the interaction of grammatical form and genre information. *Standardness* was a within-subjects manipulation: each target sentence contained either standard *doesn't* or nonstandard *don't*. *Context* was a between-subjects manipulation: participants were randomly assigned to either the NoContext condition, in which no genre or other contextual information was given before reading the sentences, or the Context condition, in which participants were told that the sentences they were to read came from popular song lyrics. Filler sentences were a mix of standard and nonstandard sentences, as elaborated on below. After filler sentences, participants occasionally responded to a comprehension question designed to keep them attending to the words.

A note is in order about the use of written stimuli. Pop songs are a somewhat unique genre in terms of linguistic modality: while one could say that they are a 'spoken' genre, it is of course more precise that they are a 'sung' genre. I chose written presentation for the following reasons. First, self-paced reading is a well-established methodology. Second and more importantly, vocal stimuli raise the confounding issue of social identity being encoded in voice, as there is no socially neutral voice (and since the forms of interest are grammatical, entire sentences are necessary; compare to Gibson's (2010) playing of isolated vowel tokens in music). Third, from a theoretical perspective, written stimuli set what is perhaps the STRICTEST bar for shifting expectations: notions of Standard English are largely based on writing, and writing is where English is policed most strongly (Milroy 2001; Squires 2010; Curzan 2014). Finally, it is not the case that song lyrics are never encountered in written form: lyrics have long been a part of album liner notes (Kreyer & Mukherjee 2007), and the internet now abounds with written iterations of lyrics in YouTube videos, in written identity performances such as on instant messenger (Childs 2016), in lyrics databases like AZlyrics.com, and on websites like Genius.com, which crowd-sources the exegesis of lyrical content.

Hypotheses

The hypotheses were as follows:

H1. STANDARDNESS MAIN EFFECT: Reading times for nonstandard *don't* will be slower than reading times for standard *doesn't*.

H2. STANDARDNESS X CONTEXT INTERACTION EFFECT: The effect of standardness will be smaller for Context participants than for NoContext participants.

Hypothesis 1 stems from previous research showing that in self-paced reading, NP_{SG} + *don't* causes a processing slowdown for American-English-speaking readers. Hypothesis 2 is based on the idea that expectations for linguistic material can be shifted by social context information. To paraphrase Seifeldin and

colleagues (2015), knowing the sentences come from song lyrics might 'loosen grammatical expectations', making the nonstandard form less surprising.

Participants

110 total participants from a Midwestern research university participated for either linguistics subject-pool credit or English-course extra credit. Fourteen participants were excluded due to either nonnative speaker status or experiment error. A total of ninety-six participants' data were used: forty-eight in the NoContext condition and forty-eight in the Context condition. Participants self-reported demographic information in the questionnaire following the experiment, giving the sample properties below.

- SEX/GENDER: thirty-five male, fifty-nine female, two 'other'
- RACE/ETHNICITY: seventy-one White, not Hispanic; seven Black/African American; two Hispanic/Latino; four Multiracial/Mixed-race; eleven Asian American; one American Indian, Alaska Native, Hawaiian, or other Pacific Islander
- PARENTS' EDUCATION LEVEL: twenty-seven no college degree; sixty-nine two-year college degree or higher

Design

Participants were presented with 144 total sentences across four blocks, as summarized in Table 1. The forty test sentences contained *don't/doesn't.* The 100 filler sentences were divided into two types: eight each contained one of eight other grammatical nonstandardisms, while thirty-six contained no nonstandardisms. These filler nonstandardisms were used to distract from *don't/doesn't* as a feature of interest. It is also true that song lyrics as a genre contain multiple nonstandardisms, not simply *don't/doesn't*, thus adding nonstandardisms was thought to lend more realism to the operationalization of genre.

Before the experimental trials began, eight neutral practice sentences (not from song lyrics) were presented, which also did not contain any nonstandardisms. In addition, because of the format in which Paradigm outputs self-paced reading data, a block introduction sentence was included at the beginning of each block. This sentence was nine words long, in the form 'This begins the first/second/third/fourth group of sentences you'll read'.

Two experiment lists were prepared.[2] For list A, half of the items were randomly assigned to be either standard or nonstandard, and items were pseudo-randomly assigned to one of four blocks (manual manipulation was done to ensure no more than two of the same type of sentence occurred in a row). Items in the nonstandard condition in list A were in the standard condition in list B and vice versa. Blocks 1 and 2 in list A were blocks 3 and 4 in list B. Lists A and B were used for the NoContext participants. These lists were then manipulated by adding the context information to

TABLE 1. *Overview of sentence types presented in Experiment 1.*

SENTENCE TYPE	NONSTANDARD GRAMMATICAL FEATURE	STANDARD ALTERNATION	TRIALS
Test	*don't*	*doesn't*	40
Filler	*gonna*	*going to*	8
	wanna	*want to*	8
	gotta	*got to*	8
	tryna	*trying to*	8
	demonstrative *them*	*those*	8
	they was	*they were*	8
	you was	*you were*	8
	there's + PL	*there are*	8
	none	–	36
Practice	none	–	8
Block introduction	none	–	4

the instructions (described below) to create two more lists, lists C and D, for the Context participants. Thus, each of the two sentence lists were shown to both participant groups.

Materials

To create the stimuli, I gathered song lyrics by searching lyrics database websites for words that would bring up the nonstandard constructions of interest. For the test sentences, I searched for *don't*, and gathered lines in which NP_{SG} + *don't* occurred in the song. In the standard condition, the sentences were then modified to include the verb form *doesn't.* Likewise, for all nonstandard grammatical features used for the filler sentences, I searched for the words or phrases likely to yield these forms (e.g. *gonna, you was, them*) and used sentences where the nonstandard form was in the lyrics. The filler sentences without nonstandard features were chosen as mundane-sounding sentences from the same set of songs.

I attempted to find sentences that could conceivably be produced across subgenres (e.g. country, rap, rock) and did not sound obviously poetic. Effort was also made to minimize reference to themes or topics that would immediately give the sentences away as lyrics or as lyrics from a specific subgenre (on the repetitive thematic content of lyrics; see Murphey 1992; Bértoli-Dutra 2014).

For the test sentences, there were other more specific restrictions. I chose sentences with full noun-phrase subjects preceding *don't.* The majority of lyrics I found in the database contained pronominal subjects (true of pop lyrics in general; Werner 2012), but I avoided these because (a) pronouns are function words and thus likely to be read much more quickly than nouns, (b) in the context-free setting of the experiment, personal pronouns would have no referent, and (c) in the lyrics, personal pronouns were usually the first word in the sentence,

making the crucial experimental word (*don't*) only the second word in the sentence. To give participants more time to read into each sentence before the target word, all final sentences contained *don't* in the third slot, following a determiner and a noun. A reviewer points out that the nature of the subject noun could conceivably affect participants' interpretation of the subsequent verb form; that is, a noun indexing a social meaning like 'country' or 'African American' could prime participants to expect a nonstandardism further down the line in the sentence. On whole, I do not believe the subject nouns in the test sentences are socially marked in a way that raises concerns—or at least, what social marking they may have is diverse across the set (some examples: *train*, *ink*, *motherland*, *cat*, *industry*, *hype*).

Some additional manipulation was done to both test and filler sentences to make them sound more grammatical in isolation, and to eliminate other nonstandard features. For instance, some of the lines were extracted from embedded clauses, so conjunctions or subordinate clause markers were removed; temporal or spatial adverbs were occasionally added or removed; other nonstandardisms were made standard; auxiliary or copular verbs were added. Across conditions, sentences had a minimum of four and a maximum of nine words.

Procedure

Participants were assigned to one of the four lists. After the researcher obtained their consent, participants were seated at an individual computer testing station. The experimenter summarized the task instructions verbally, then began the experiment on the computer.

Detailed experiment instructions to the participants differed by experimental group. The following instructions, given in (2) below, were split across screens of the experiment. The underlined portions occurred only in the instructions of the Context participants.

(2) In this experiment, you will be silently reading a series of sentences.
The sentences come from song lyrics from a number of genres.
These genres include pop, country, rap, rock, folk, hip-hop, alternative, and R and B.

Each sentence will appear one word at a time.
You will use the middle green button on the response pad to advance through the sentences/lyrics.
Once a word appears, read it as quickly as possible, then press the green button to move on to the next word.

As you go through the experiment, you will be asked questions about some of the sentences.
Use the red (left) and blue (right) buttons on the response pad to answer the questions.
Be sure to pay attention to the words in the sentences/lyrics!

The following sentences are for practice.
They do not come from song lyrics.

(after practice trials)

Good job!
Now you are ready to begin the experiment.
Remember, these sentences are from song lyrics.

The forty comprehension questions, presented after filler sentences, took the following format: 'Which of the following words was NOT in the previous sentence?' Participants then chose between two words. The distractor word was closely related, either formally or semantically, to a word that had occurred in the sentence. Overall comprehension question accuracy rate was high, at 98.26%, suggesting that participants were attentive.

Participants completed a twenty-one-question postexperiment questionnaire including demographic information, their impressions of the experiment, and their musical consumption and tastes.

Analysis

I removed all outliers with raw reading times under 100 ms and over 2000 ms (following Fine, Jaeger, Farmer, & Qian 2013). Length- and position-adjusted residual reading times were then calculated for each word. First, I created a linear regression model for predicted reading time as a function of word length in characters and word position in sentence, with subject as a random effect. The residuals of this model (the difference between predicted and actual reading time for each observation) were then the dependent variable in the analysis (see e.g. Enochson & Culbertson 2015). Negative residual reading times indicate faster responses than predicted, and positive times indicate slower responses than predicted. Practice trials were excluded from the residual reading time calculation, but filler sentences were included. The final analysis includes the residual reading times of only the forty test target sentences.

Figure 1 shows the mean residual reading times across words 2–5 in the target sentences, divided into Context and NoContext groups of participants. Error bars represent the standard error of the mean. The critical word in the target sentences is word 3, which always corresponds to *don't/doesn't*. Note that word 3 generally has a lower residual reading time than other words in the sentence, due to the status of *don't/doesn't* as a function word. The visualization makes it appear that the effect of standardness is LARGER for the Context group than the NoContext group, contrary to Hypothesis 2.

Mixed-effects linear regression models were created for words 3–5: the test word plus two following words. For each word, I used an automated stepwise model comparison procedure using the R function {step} with {lme4}. This tests a maximal

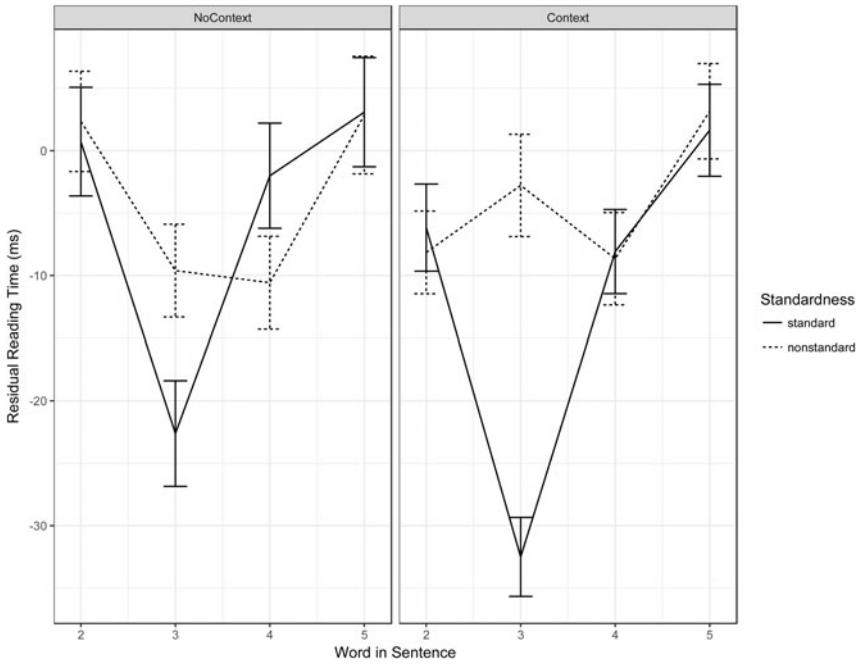

FIGURE 1. Residual reading times in target sentences by Standardness and Context condition (Experiment 1).

interaction structure, running a model-comparison procedure to determine which factors to retain for the best-fitting model. The dependent variable was residual reading time (in milliseconds). The fixed-effect predictor variables were trial number (because reading times naturally decrease across an experimental session), standardness, and context. Random slope terms were included for participants and items.

In the model, trial is a continuous variable, the reference factor level for standardness is Standard, and the reference factor level for context is NoContext. Therefore, estimates should be interpreted as estimating the effect of Nonstandard sentences relative to Standard, and Context participants relative to NoContext. I have used these labels in Table 2 to aid in interpretation.

Trial was significant; over the course of the experiment, reading times decreased. At *don't/doesn't* (word 3), reading times show a complicated set of significant predictors. Standardness does not reach significance as an independent main effect. Context is significant as a main effect: reading times in the Context condition were faster than those in the NoContext condition. Additionally, standardness is significant in two-way interactions with trial and context, and in a three-way interaction with trial and context.

TABLE 2. *Regression results for words 3–5 in target sentences (Experiment 1).*

WORD	FIXED EFFECTS	ESTIMATE	SE	*t*	*p*-VALUE
don't/doesn't					
(Word 3)	intercept	75.6225	11.0922	6.818	< .001
	trial	−1.2443	.1220	−10.201	< .001
	nonstandard	−17.6444	16.6043	−1.063	
	context	−43.1444	11.5602	−3.732	< .001
	trial:nonstandard	.4163	.1978	2.105	< .05
	trial:context	.4227	.1255	3.369	< .001
	nonstandard:context	47.2547	15.2079	3.107	< .01
	trial:nonstandard:context	−.3917	.1685	−2.324	< .05
Word 4	intercept	73.9837	6.174	11.983	< .001
	trial	−.9999	.0595	−16.799	< .001
	context	−17.0197	8.1171	−2.097	< .05
	trial:context	.1861	.0837	2.223	< .05
Word 5	intercept	63.8972	8.3402	7.661	< .001
	trial	−.7627	.0459	−16.618	< .001

To assist in interpreting these interactions, Figure 2 shows the linear model for residual reading times for *don't/doesn't* across trials in the experiment, divided by both standardness (the line type) and context (the two panels). The interaction between trial and context means that the difference between the two groups' reading times changes over the course of the experiment: the Context group is faster earlier in the experiment. The interaction between trial and standardness means that the standardness effect gets larger over the course of the experiment (this is opposite of what would be predicted, given that people tend to become more accommodating of variable forms with more exposure). The interaction between standardness and context shows that the effect of nonstandard *don't* is indeed larger for the Context group than the NoContext group, as apparent in Figure 1.

Note that in both characters and syllables, *doesn't* is a longer word. The residual reading time analysis adjusts for word length in number of characters, but not for syllables. Given these factors, in general *don't* should be faster than *doesn't*, not slower. Given that both are function words—and fulfill precisely the same function at that—this difference between them should be considered meaningful; if anything, it underestimates the relative quickness of the two-syllable standard variant (and, by implication, the difference between the two variants).

However, complicating things is the three-way interaction between these predictors. Whereas the Context group is consistently affected by the standardness manipulation over the course of the experiment, the NoContext group shows the effect of standardness in later trials of the experiment. Thus, the locus of the interactions seems to be in a difference between the two context groups, with the NoContext

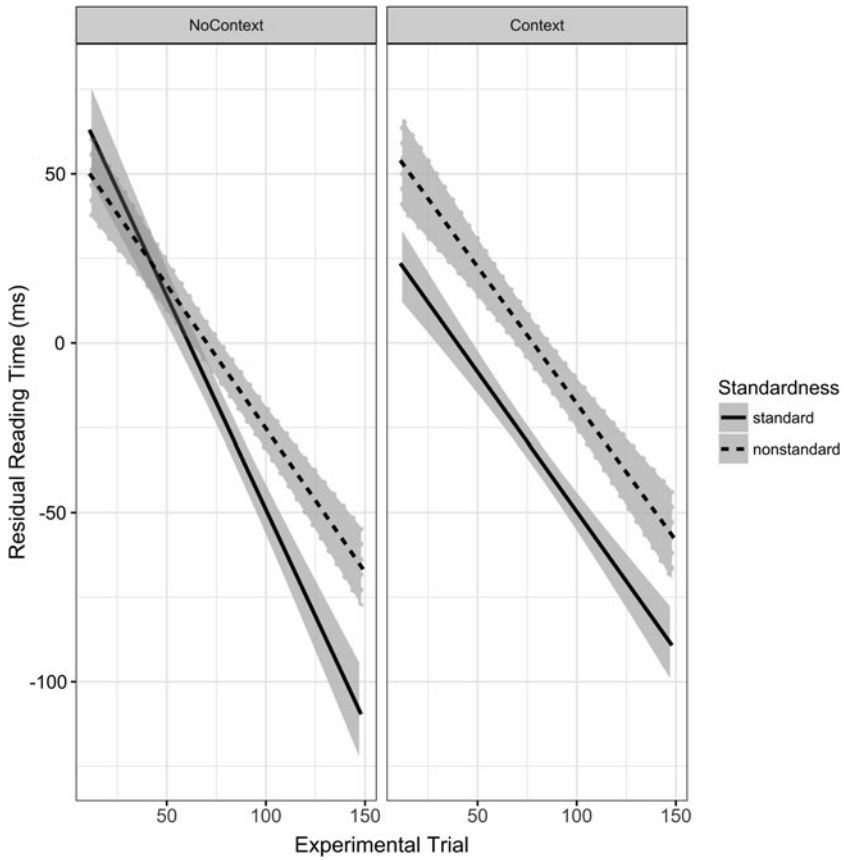

FIGURE 2. Residual reading-time regression lines as a function of Experimental Trial over the experiment, Standardness, and Context (Experiment 1)

group having longer reading times and no difference between the two grammatical conditions earlier in the experiment, but then converging with the pattern of the Context group by the end.

At word 4, context is a main effect, showing faster reading times for the Context group relative to the NoContext group. However, this too is mediated by an interaction with trial: the NoContext group is slower earlier in the experiment but not later. Neither words 4 nor 5 show effects of standardness, and context is not a meaningful predictor for word 5.

Discussion

Hypothesis 1 was confirmed, but with qualification. Nonstandard *don't* took longer to process than standard *doesn't*, but the effect was mediated by both context group

and trial. Hypothesis 2 was not confirmed. Context had an effect, but instead of a straightforward interaction between context and standardness, the two groups seemed to behave differently toward the experiment as a whole. While the Context group was affected by the standardness across the experiment and from the beginning, the NoContext group did not show a standardness effect until later.

My interpretation of these results is that rather than context information straightforwardly affecting linguistic expectations, it affected ORIENTATION TO THE PROCESSING TASK. Participants without the context information had noisier reading times at the beginning of the experiment—perhaps more difficulty accommodating the variety of linguistic forms encountered, between the test and filler sentences. Seen this way, genre information led the Context group to more consistent processing from the beginning. Additional trials made both groups faster, but for those with no context going into it, they also stabilized the grammatical effect.

The participants' postexperiment questionnaire responses lend credence to the interpretation that the two groups had different processing strategies, which did not map on to a simple difference in reading times as expected. Participants were asked after the experiment, 'Did you notice anything interesting about the sentences you just read? Please describe'. Representative comments from the NoContext group in (3a–c) reveal that participants were trying to form a logic for both the sentences themselves and their grammatical forms. As participants went through the sentences, they were trying to come up with a generic source for them: 'Where did these sentences come from?'

(3) a. There were grammatical mistakes in a lot of them that caught my eye. Also a lot of them sounded like quotes from something (i.e. song lyrics).
b. Sometimes the grammar wasn't 100% correct (ex. using 'them' instead of 'the' as a definite article) and some of the sentences just didn't make much sense (like the one about the clown honking someone's nose or something? that was weird).
c. They had no correlation, in my opinion, at first, though some of the sentences did go together. I also noticed the use of rep[e]tition of so[m]e words in the sentences. It also seemed as if a few went together, as they rhymed.

The comments of the Context group in (4a–c) are strikingly different.

(4) a. They weren't all syntactically well-formed. Some of them didn't make sense in terms of syntax. And some of them I knew which songs they were lyrics to.
b. I thought I would recognize more of the lyrics but I really only recognized two of them.
c. The only thing that stood out to me was the word choice in the sentences. It is clear that there was some deviation from Standard American English. But that is to be expected in song lyrics.

These participants go into the experiment knowing to expect song lyrics. They seem to be reading with the question, 'Which song is this lyric from?'. These are two very different orientations to the task of reading the sentences. With context information, participants established a steady baseline and read for recognition of content. Without context information, participants read to understand the task itself.

If context affected task orientation, how does that orientation then relate to processing at the word or sentence level? The comments show that participants did make a metalinguistic connection between the grammar and the genre. Some NoContext participants tried to attribute the nonstandard forms to a particular genre (lyrics or poetry) or speaker (some mentioned imagining a Southern person), while the Context participants directly attributed the nonstandardisms to the genre of lyrics, as in (4c). Genre information gave the Context participants a rationalization for the presence of nonstandard features. Even if participants had not been actively predicting these features, their occurrence was rapidly integrated into a set of expectations for song lyrics that included nonstandard forms.

By contrast, since the NoContext participants did not have a reason to expect the nonstandard forms (in a laboratory setting in a university), it is possible that it took longer for them to adapt to the task at a basic level, with so many different kinds of grammatical forms coming as input in a fairly rapid span of time. This, I suggest, could have made the NoContext group's responses noisier at the beginning of the experiment. One thing to contribute to this noise was the presentation of multiple nonstandardisms across the different types of sentences. To investigate whether the nonstandard filler sentences could have contributed to the results in Experiment 1, I prepared Experiment 2 as a replication of the task with no nonstandard filler sentences.

EXPERIMENT 2

Methods and hypotheses

Experiment 2 was identical to Experiment 1 with the exception that all filler sentences occurred in their standard forms, so the only nonstandard sentences were in the target sentences containing NP_{SG} + *don't*. As noted above, this change was meant to address the possibility that the presence of several different kinds of nonstandardisms across sentence types created an unstable reading environment for the NoContext participants, who did not have context information available to scaffold or 'explain' the presence of nonstandard sentences.

Hypotheses 1 and 2 were identical to those of Experiment 1.

Participants

Eighty-seven participants completed the experiment for linguistics subject-pool or English-course extra credit. Eight participants were removed for being nonnative

speakers. The data of seventy-nine participants were analyzed: thirty-eight NoContext and forty-one Context. The self-reported demographics are given below.

- SEX/GENDER: thirty male, forty-nine female
- RACE/ETHNICITY: sixty-one White, not Hispanic; seven Black/African American; one Hispanic/Latino; three Multiracial/Mixed-race; five Asian American; two 'other'
- PARENTS' EDUCATION LEVEL: twenty no college degree; fifty-nine two-year college degree or higher

Materials, design, procedure

Materials were identical to those of Experiment 1 except that all fillers were made to be standard. The design and procedure were identical to that of Experiment 1.

Results

The procedure for analysis was the same as in Experiment 1.

Figure 3 shows residual reading times for words 2–5, divided by context group and standardness of target sentence. The NoContext group has a larger difference between the standard and nonstandard conditions than the Context group. The final regression model terms and results are listed in Table 3.

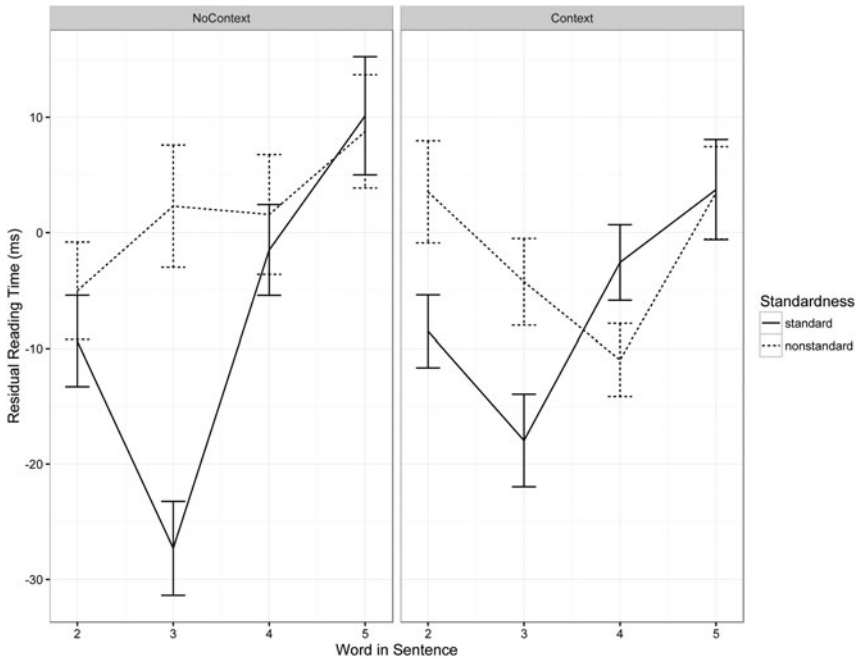

FIGURE 3. Residual reading times in target sentences by Standardness and Context condition (Experiment 2).

TABLE 3. *Regression results for words 3–5 in target sentences (Experiment 2).*

WORD	FIXED EFFECTS	ESTIMATE	SE	T	*p*-VALUE
don't/doesn't					
(Word 3)	intercept	43.420	7.2814	5.963	< .001
	trial	−.8931	.0479	−18.653	< .001
	context	9.165	6.8995	1.328	
	nonstandard	31.5964	5.6987	5.544	< .001
	context:nonstandard	−15.7416	7.904	−1.992	< .05
Word 4	intercept	78.7978	7.1988	10.946	< .001
	trial	−.9821	.0619	−15.869	< .001
	context	−22.5473	8.5791	−2.628	< .01
	trial:context	.1965	.0853	2.303	< .05
Word 5	intercept	79.8332	8.4324	8.4	< .001
	trial	−.8029	.0503	−15.98	< .001

This experiment's results support both hypotheses. Nonstandard *don't* took longer to read than nonstandard *doesn't.* There was also an interaction of context and standardness: Context participants were less strongly affected by the standardness manipulation than were NoContext participants.

At word 4, the main effect of trial and context, and the interaction effect of trial and context, were found to be similar as in Experiment 1. And, as with Experiment 1, word 5 showed no effect of either context or standardness.

Discussion

Both participant groups were affected by the presence of nonstandard *don't*, but the effect was smaller for the group of participants who had genre information given to them. It seems that the expectation for song lyrics activated, in part, a loosening of grammatical standards. (NB: This assumes that these are speakers whose baseline expectations would be 'standard' grammatical forms. Given a number of factors, including the fact that the experiment occurred in a laboratory reading setting on a university campus and the demographics of the sample, it is a reasonable assumption that the baseline expectation for a written text would be standard. See Squires 2014c for some discussion.)

To test for even stronger degrees of expectation shift, I ran an additional version of the experiment with more explicit context information. Can participants' expectations away from standard grammatical forms be made strong enough to all but mitigate the processing effect of nonstandard *don't*?

EXPERIMENT 3

Methods and hypotheses

In Experiment 3, participants were given even more explicit instructions that linked linguistic nonstandardness with song lyrics as a genre. In order to maintain

nonstandardness throughout the experiment and not only in the relatively few target trials, nonstandard fillers were again used. The new participants' responses were then compared to those of the Context participants from Experiment 1, who had been exposed to a weaker context manipulation. For this analysis, the between-subjects variable is ContextStrength: Experiment 1 participants are the 'Weak' context group and Experiment 3 participants are the 'Strong' context group. The hypotheses for Experiment 3 are given below.

H1. STANDARDNESS MAIN EFFECT: Reading times for nonstandard *don't* will be slower than reading times for standard *doesn't*.

H2. STANDARDNESS X CONTEXTSTRENGTH INTERACTION EFFECT: The effect of standardness will be smaller for Strong context participants than for Weak context participants.

Participants

The new experiment version was run on thirty-eight participants. The data of seven were removed due to experiment failure or nonnative speaker status, leaving thirty-one participants in Experiment 3.

The data of the forty-eight Context participants from Experiment 1 were combined with the new data into one data set, for a total of seventy-nine participants. The self-reported demographics of the combined sample are given below.

- SEX/GENDER: twenty-eight male, fifty-one female
- RACE/ETHNICITY: fifty-four White, not Hispanic; eight Black/African American; two Hispanic/Latino; four Multiracial/Mixed-race; eleven Asian American
- PARENTS' EDUCATION LEVEL: twenty-five no college degree; fifty-four two-year degree or higher

Materials, design, procedure

The experiment design and procedure were identical to Experiments 1 and 2, except that the instructions explicitly linked grammar with the genre information. The first slide of instructions (modifications from Experiments 1 and 2 are underlined) are given in (5) below.

(5) In this experiment, you will be silently reading a series of sentences.
The sentences come from song lyrics from a number of genres.
These genres include pop, country, rap, rock, folk, hip-hop, alternative, and R and B.
<u>The sentences may contain language that is not considered 'correct' or 'standard'.</u>
(after practice trials)

Good job!
Now you are ready to begin the experiment.
Remember, these sentences are from song lyrics, and some of them may contain 'incorrect' or 'nonstandard' language.

Results and discussion

In order to treat the observations as a single data set, the raw reading times for all participants were combined, and new residual reading times for this combined data set were created. The residual reading-time calculation is sensitive to the other observations in the dataset, hence combining the raw data was seen as more valid (i.e. it would not be appropriate to create residual reading times for each group when the group difference is the effect of interest). Figure 4 shows residual reading times across words in the target sentences for the two context groups (Weak, Experiment 1, N = 48; Strong, Experiment 3, N = 31).

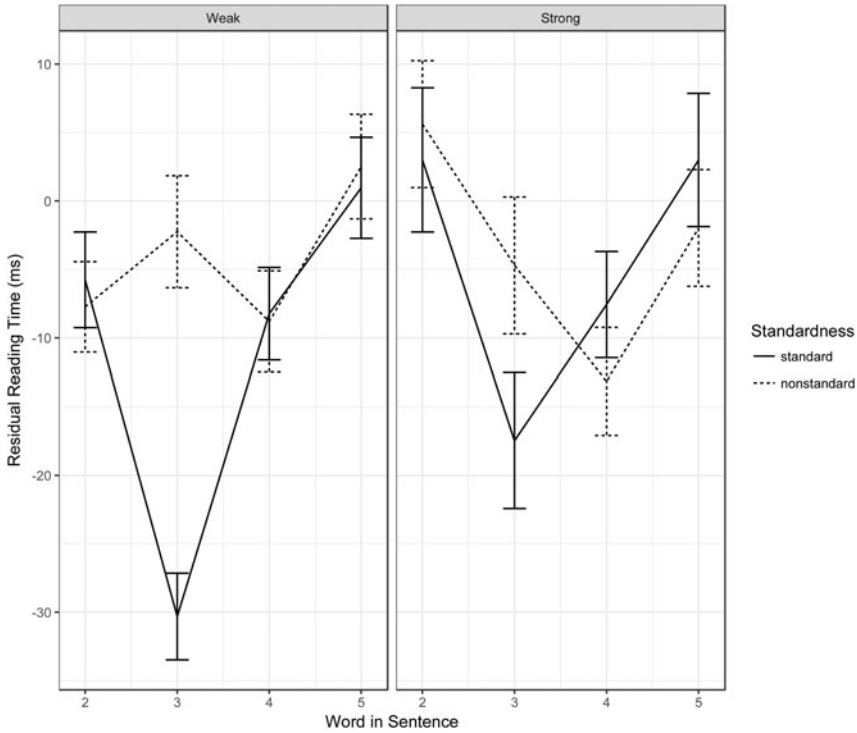

FIGURE 4. Residual reading times in target sentences by Standardness and ContextStrength condition (Experiment 3).

TABLE 4. *Regression results for words 3–5 in target sentences (Experiment 3).*

WORD	FIXED EFFECTS	ESTIMATE	SE	T	*P*-VALUE
don't/doesn't	intercept	39.31612	5.84962	6.721	< .001
	trialposition	−.81691	.04729	−17.274	< .001
	nonstandard	23.94788	3.90759	6.129	< .001
Word 4	intercept	53.18543	4.60170	11.56	< .001
	trial	−.77773	.04216	−18.45	< .001
Word 5	intercept	56.73737	7.74103	7.329	< .001
	trial	−.69130	.04515	−15.31	< .001

As evidenced in the figure, the means do show a smaller difference between the standard and nonstandard conditions for the Strong participants than for the Weak participants, in line with the hypothesis. However, the interaction does not reach the level of significance (or more precisely, adding the interaction effect to the statistical model does not produce a better-fitting model; see Table 4). Nonetheless, this exploratory experimental condition suggests that it would be fruitful to continue to probe differing levels of context information, as I discuss further below.

GENERAL DISCUSSION AND CONCLUSION

These experiments tested the broad hypotheses that generic links between form and context are part of speakers' sociolinguistic knowledge, and that those links can be activated by context information, thereby shifting expectations for grammatical form. Experiment 1 found a complicated interaction between standardness, context, and time across the experiment. Experiment 2 found an effect of standardness mediated by context: participants given information about song lyrics had a smaller standardness effect than those without context information. Experiment 3 found a standardness effect and a small but not statistically significant difference between participants given 'weak' versus 'strong' context information. From these findings, three immediate issues invite further discussion: (i) the locus of the effect of genre information, (ii) the presence of one versus many nonstandardisms across the sentences, and (iii) future possibilities for exploring different forms of 'context information' and its relevance to processing.

First, to revisit the motivating question for this study, the results suggest that speech genre can serve as an expectation-shifting sociolinguistic cue during sentence processing. Participants with context information were certainly oriented differently to the task of reading the sentences compared to those without: THEY WERE EXPECTING DIFFERENT KINDS OF SENTENCES. But the expectation for 'difference' seemed to be a general one, including ideas about linguistic form but not being specific to them (as with Seifeldin et al. 2015 on the processing of copula absence).

Both reading times and the postexperiment comments provide good evidence that morphosyntactic nonstandardness IS linked in speakers' knowledge to song lyrics as a genre. Yet there were also other features that either triggered the interpretation of the sentences as being from an 'atypical' genre or from song lyrics specifically—namely, the themes and content of the sentences, as well as a sense that the sentences were somehow 'poetic'. Thus, the direct difference between the NoContext and Context groups is one of task (reading) orientation, which secondarily affected grammatical expectation (for some participants). One kind of experiment that may be able to follow up on these findings would be self-paced LISTENING of similar stimuli. As a more passive task, listening would take out the process of text comprehension en route to language comprehension (though of course still requiring, for example, speech recognition and lexical access, and, as discussed before, with the added social complications brought in by voices).

Second, Experiments 1 and 2 differed in the number and type of nonstandardisms in the sentences. While Experiment 1's NoContext participants may have had noisy reading times due to other factors, the results suggest at least that the number and type of nonstandardisms is pertinent to processing. I've looked elsewhere at the effect of the number of instances of the SAME feature across an experiment (Squires 2016), finding that participants' level of awareness about the feature's presence seems stable whether they are exposed to only a handful of instances or many (see also Labov et al. 2011). I have not, however, explored the effect of different features combined. This seems an especially apt area of exploration in the drive to connect laboratory processing with 'real-world' language use: in real situations of interaction between different-dialect speakers, there would typically be not just one feature of difference but a bundle.

Third, while I tried to elicit a stronger expectation shift by offering a more overt context cue in Experiment 3, the result was not strong enough to conclude its effectiveness. My previous work has suggested that shifts in perception/processing towards favoring nonstandard variants are difficult to elicit and generally small (Squires 2013, 2014a); this was also the case here. It is true that other studies have found rapid and almost complete accommodation to unfamiliar dialect variants or ungrammatical sentences, as referenced in the literature review. This has not been the outcome of my own work, and future studies can more carefully tease apart variants that do some language-ideological work versus those that do not (whether they are familiar or totally unknown). There might simply be a ceiling effect for the level of 'acceptance' or accommodation possible for other-dialect features, especially features that may, absent context, be read as stigmatized to the participant population. It is also possible that telling participants to expect nonstandard forms (metalinguistically priming them) caused them to fixate MORE on the nonstandard forms when they did encounter them, thus washing out a mitigation effect. Perhaps even more fine-grained measures, such as eye-tracking or ERPs, could tease apart surprisal from fixation.

While experimental in nature, this study is about the links between social and linguistic material that are a part of speakers' knowledge of language. I have thus

endeavored to make explicit connections between laboratory experimental methods, focused on fine-grained measurable responses, and analytical concepts typically treated more anthropologically—what I will broadly call *variation concepts*. These point to facts about language variation, speakers' metalinguistic beliefs about language/variation, or both. As I have been incorporating it here, *genre* is a variation concept, naming types of speech recognizable by members of a speech community. *Enregisterment* is another (Agha 2007), naming the process by which linguistic features become construed as belonging to a language variety. A related term is *language ideology*, the beliefs and ideas speakers have and use to understand and rationalize linguistic behavior (Silverstein 1979; Woolard 1998). Yet another is the 'third-wave' conception of *style* (e.g. Moore & Podesva 2009; Eckert 2012), which points to clusters of both linguistic and nonlinguistic elements that carry social meaning. Variation concepts name relationships we observe based on linguistic and metalinguistic behavior. If these relationships are part of the mental models speakers have of language in their world, they should also affect 'on the ground' language comprehension. It has now been robustly shown that social properties of individual speakers matter to speech perception (e.g. Hay & Drager 2010; Drager 2011; McGowan 2015), but many other elements of context remain to be investigated (though see Hay et al. 2017), and much more work is needed on sentence processing/comprehension.

By using variation concepts as starting points for experimental investigations, we can understand more about the cognitive mechanisms by which sociolinguistic categories form and the social effects they carry in interaction (see related discussion in Campbell-Kibler 2016). For example, one aim of this study was to expand upon which types of social factors affect linguistic expectation, by focusing on genre (context) rather than persona (speaker). Yet participants' qualitative questionnaire responses make clear that despite the lack of cuing to do so during the experiment, they readily imagined a speaker for the sentences. This imaginary speaker was frequently described by participants as, unsurprisingly given the stereotypical indexicalities of nonstandard features, 'less educated', 'Southern', 'rural', or 'African American'. While genre can exist as a category independent of individuals, speakers may nonetheless fundamentally organize knowledge about variation in accordance with types of people (Campbell-Kibler 2009). This is not at all surprising—it aligns with approaches to sociolinguistic production and perception that foreground notions of *style* and *persona* (e.g. Campbell-Kibler 2007; Podesva 2007; Eckert 2012; D'Onofrio 2015).

Moreover, if we want to understand the formation, maintenance, and change of language ideologies, we should be interested in two key domains of speakers' exposure: experiences with language/variation and metalinguistic discourses about language/variation. I believe that exploring in-the-moment processing of language variation can illuminate both domains—and their relationship. At the experiential level, how does a speaker's processing apparatus handle variant A versus variant B? And does it handle variant A differently when encountered in condition X

versus condition Y? And to what extent do the existing metalinguistic beliefs the speaker has about language influence their handling?

Understanding how dialect differences are processed is an important goal for linguistics for theoretical reasons, but I would suggest that it can also inform more 'practical' problems in the study of language in society. In their critically important discussion of jurors' mishearing of witness Rachel Jeantel during the murder trial of George Zimmerman, Rickford & King (2016) note the need for more research into how non-AAVE speakers comprehend AAVE. As they put it, more generally, 'the receptive side of competence in variation studies has been underinvestigated' (2016:973). Language processing is part of this 'receptive competence'—it is the very foundation of it. We have ample research documenting people's explicit judgments of others' language, as well as self-assessments of whether they understood another person or not. We also have ample research showing that those judgments and assessment are strongly influenced by social beliefs and biases (e.g. Rubin 1992). But explicit evaluation of language is one thing; what actually happens as someone processes language is another. More information about the latter should inform our understanding of the former, and better position us to ask how linguistic differences come to be understood and evaluated as socially meaningful differences.

Appendix: Test sentences (nonstandard version)

The boss don't care about the dress code
A heart don't forget something like that
This hair don't mean a thing
This game don't seem like much fun
A man don't have to die
This industry don't come with benefits
The dark don't hide it
This life don't last forever
The moon don't shine every night
This flower don't belong to me
This part don't cost any money
This clown don't smile and honk your nose
That dog don't hunt in the woods
This kid don't usually let them stay
This guy don't want to battle
This money don't make a difference
This ship don't sail far
The king don't have to wear his crown
This road don't stop shifting
The lady don't mind a bit
Your ghost don't scare me now
The government don't want to see that
This mirror don't look the same

The highway don't know you're alive
The camera don't lie to you
The pump don't work right
The music don't feel like it did
That car don't come out until next year
The motherland don't love you
This ink don't come off even if rings do
This drink don't spill over
This train don't stop there anymore
This diet don't seem to work
The coroner don't need sheets
This rig don't dig any overload
That rocket don't stop until I'm done
This house don't feel like home when I'm alone
That hype don't feel the same next year
This cat don't chase any mice
That crown don't make you a prince

NOTES

*Thanks to the editor and anonymous reviewer, who provided critical and helpful readings of this article. I would also like to acknowledge Robin Queen and Julie Boland for longstanding conversations that have shaped this research; my research assistants, Bethany Toma and Jamie Smith; and my departmental colleagues, Tracee Mohler and Wayne Lovely, who helped me in various ways to get this work done. I am also grateful to audiences at SVALP, Stanford University, and the Graduate Center at CUNY for their valuable comments and questions on this work as it evolved. All shortcomings are, as always, my own responsibility.

[1]http://paradigmexperiments.com

[2]The test sentences are listed in the appendix.

REFERENCES

Agha, Asif (2007). *Language and social relations*. Cambridge: Cambridge University Press.

Bakhtin, Mikhail M. (1986). *Speech genres and other late essays*. Ed. by Caryl Emerson & Michael Holquist. Trans. by Vern W. McGee. Austin: University of Texas Press.

Beal, Joan C. (2009). 'You're not from New York City, you're from Rotherham': Dialect and identity in British indie music. *Journal of English Linguistics* 37(3):223–40.

Bértoli-Dutra, Patricia (2014). Multi-dimensional analysis of pop songs. In Tony Berber Sardinha & Marcia Veirano Pinto (eds.), *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber*, 149–76. Amsterdam: John Benjamins.

Biber, Douglas (1993). The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities* 26:331–45.

———, & Susan Conrad (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Bieber, Justin; Johntá Austin; & Bryan-Michael Cox (2010). Never let you go [Recorded by Justin Bieber]. On *My world 2.0* [Digital download]. Atlanta, GA: Island.

———, & Sam Hook (2014). We were born for this [Recorded by Justin Bieber]. [Soundcloud download]. Online: https://soundcloud.com/justinbieber/we-were-born-for-this.

———; Julia Michaels; Justin Tranter; Sonny Moore; & Michael Tucker (2015). Sorry [Recorded by Justin Bieber]. On *Purpose* [Digital download]. Los Angeles, CA: Def Jam.

———; Jason 'Poo Bear' Boyd; & Mason Levy (2015). What do you mean? [Recorded by Justin Bieber]. On *Purpose* [Digital download]. Los Angeles, CA: Def Jam.
Boland, Julie; Guadalupe de los Santos; Julie Carranza; & Michael Kaschak (2015). Self-paced reading time as a measure of learning novel constructions. 28th Annual CUNY Conference on Sentence Processing. University of Southern California, March 19–21.
Briggs, Charles, & Richard Bauman (1992). Genre, intertextuality and social power. *Journal of Linguistic Anthropology* 2:131–72.
Campbell-Kibler, Kathryn (2007). Accent, (ING) and the social logic of listener perceptions. *American Speech* 82(1):32–64.
——— (2009). The nature of sociolinguistic perception. *Language Variation and Change* 21(1):135–56.
——— (2016). Towards a cognitively realistic model of meaningful sociolinguistic variation. In Anna M. Babel (ed.), *Awareness and control in sociolinguistic research*, 123–51. Cambridge: Cambridge University Press.
Childs, Rebecca (2016). Who I am and who I want to be: Variation and representation in a messaging platform. In Lauren Squires (ed.), *English in computer-mediated communication: Variation, representation, and perception*, 261–80. New York: Mouton de Gruyter.
Coupland, Nikolas (2011). Voice, place and genre in popular song performance. *Journal of Sociolinguistics* 15:573–602.
Curzan, Anne (2014). *Fixing English: Prescriptivism and language history*. Cambridge: Cambridge University Press.
D'Onofrio, Annette (2015). Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics* 19(2):241–56.
Drager, Katie (2011). Speaker age and vowel perception. *Language and Speech* 54(1):99–121.
Duncan, Daniel (2017). Australian singer, American features: Performing authenticity in country music. *Language & Communication* 52:31–44.
Eberhardt, Maeve, & Kara Freeman (2015). 'First things first, I'm the realest': Linguistic appropriation, white privilege, and the hip-hop persona of Iggy Azalea. *Journal of Sociolinguistics* 19(3):303–27.
Eckert, Penelope (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41:87–100.
Enochson, Kelly, & Jennifer Culbertson (2015). Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PloS one* 10(3):e0116946.
Ferguson, Charles A. (1994). Dialect, register, and genre: Working assumptions about conventionalization. In Douglas Biber & Edward Finegan (eds.), *Sociolinguistic perspectives on register*, 15–30. Oxford: Oxford University Press.
Fine, Alex B.; T. Florian Jaeger; Thomas A. Farmer; & Ting Qian (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one* 8(10):e77661.
Gibson, Andy (2010). *Production and perception of vowels in New Zealand popular music*. Auckland: Auckland University of Technology MPhil thesis.
Giltrow, Janet, & Dieter Stein (eds.) (2009). *Genres in the internet: Issues in the theory of genre*. Amsterdam: John Benjamins.
Hanks, William F. (1987). Discourse genres in a theory of practice. *American Ethnologist* 14:668–92.
Hanulíková, 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(4):878–87.
———, & Manuel Carreiras (2015). Electrophysiology of subject-verb agreement mediated by speakers' gender. *Frontiers in Psychology* 6:1396. doi: 10.3389/fpsyg.2015.01396.
Hay, Jennifer, & Katie Drager (2010). Stuffed toys and speech perception. *Linguistics* 48(4):865–92.
———; Ryan Podlubny; Katie Drager; & Megan McAuliffe (2017). Car-talk: Location-specific speech production and perception. *Journal of Phonetics* 65:94–109.
Hesson, Ashley, & Madeline Shellgren (2015). Discourse marker *like* in real time: Characterizing the time-course of sociolinguistic impression formation. *American Speech* 90(2):154–86.

Hymes, Dell (1974). *Foundations of sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.
Jaeger, T. Florian, & Kodi Weatherholtz (2016). What the heck is salience? How predictive language processing contributes to sociolinguistic perception. *Frontiers in Psychology* 7:1115. doi: 10.3389/fpsyg.2016.01115.
Kamide, Yuki (2012). Learning individual talkers' structural preferences. *Cognition* 124(1):66–71.
Kaschak, Michael P. (2006). What this construction needs is generalized. *Memory & Cognition* 34 (2):368–79.
———, & Arthur M. Glenberg (2004). This construction needs learned. *Journal of Experimental Psychology: General* 133(3):450–67.
Kreyer, Rolf, & Joybrato Mukherjee (2007). The style of pop song lyrics: A corpus-linguistic pilot study. *Anglia* 125(1):31–58.
Kuperberg, Gina R., & T. Florian Jaeger (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31(1):32–59.
Labov, William (1973). Where do grammars stop? In Roger W. Shuy (ed.), *Report of the Twenty-Third Annual Round Table Meeting on Linguistics and Language Studies*, 43–88. Washington, DC: Georgetown University Press.
———; Sharon Ash; Maya Ravindranath; Tracey Weldon; Maciej Baranowski; & Naomi Nagy (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15(4):431–63.
Levon, Erez, & Isabelle Buchstaller (2015). Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. *Language Variation and Change* 27(3):319–48.
Logan, Beth; Andrew Kositsky; & Pedro Moreno (2004). Semantic analysis of song lyrics. In *Proceedings of the* 2004 *IEEE International Conference on Multimedia and Expo* (ICME'04), vol. 2, 827–30. doi: 10.1109/ICME.2004.1394328.
Luka, Barbara J., & Lawrence W. Barsalou (2005). Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52(3):436–59.
McGowan, Kevin B. (2015). Social expectation improves speech perception in noise. *Language and Speech* 58(4):502–21.
Milroy, James (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics* 5(4):530–55.
Milroy, Lesley, & James Milroy (1999). *Authority in language: Investigating Standard English*. 3rd edn. London: Routledge.
Moore, Emma, & Robert J. Podesva (2009). Style, indexicality, and the social meaning of tag questions. *Language in Society* 38(4):447–85.
Murphey, Tim (1990). *Music and song in language learning: An analysis of pop song lyrics and the use of music and song in teaching English to speakers of other languages*. Bern: Lang.
——— (1992). The discourse of pop songs. *TESOL Quarterly* 26:770–74.
Podesva, Robert J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11(4):478–504.
Rickford, John R., & Sharese King (2016). Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language* 92(4):948–88.
Rubin, Donald L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33(4):511–31.
Sanchez, Kauyumari; Jennifer Hay; & Elissa Nilson (2015). Contextual activation of Australia can affect New Zealanders' vowel productions. *Journal of Phonetics* 48:76–95.
Seifeldin, Maryam; Max Cantor; Julie Boland; & Jonathan Brennan (2015). They dropping copulas: Salient cues in the integration of speaker identity and syntax. 28th Annual CUNY Conference on Sentence Processing. University of Southern California, March 19–21.

Sheeran, Ed; Benny Blanco; & Justin Bieber (2015). Love yourself [Recorded by Justin Bieber]. On *Purpose* [Digital download]. Los Angeles, CA: Def Jam.
Silverstein, Michael (1979). Language structure and linguistic ideology. In Paul R. Clyne, William F. Hanks, & Carol L. Hofbauer (eds.), *The elements: A parasession on linguistic units and levels*, 193–247. Chicago: Chicago Linguistic Society.
Simpson, Paul (1999). Language, culture and identity: With (another) look at accents in pop and rock singing. *Multilingua: Journal of Cross-Cultural and Interlanguage Communication* 18(4):343–68.
Squires, Lauren (2010). Enregistering internet language. *Language in Society* 39(4):457–92.
——— (2013). It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of Sociolinguistics* 17(2):200–237.
——— (2014a). Knowledge, processing, evaluation: Testing the sociolinguistic perception of English subject-verb agreement variation. *Journal of English Linguistics* 42(2):144–72.
——— (2014b). Talker specificity and the perception of grammatical variation. *Language, Cognition and Neuroscience* 29(7):856–76.
——— (2014c). Social differences in the processing of grammatical variation. *Penn Working Papers in Linguistics* 20(2):Article 20. Online: http://repository.upenn.edu/pwpl/vol20/iss2/20/.
——— (2016). Processing grammatical differences: Perceiving v. noticing. In Anna M. Babel (ed.), *Awareness and control in sociolinguistic research*, 80–103. Cambridge: Cambridge University Press.
Steinhauer, Karsten, & John F. Connolly (2008). Event-related potentials in the study of language. In Brigitte Stemmer & Harry A. Whitaker (eds.), *Handbook of the neuroscience of language*, 91–104. Amsterdam: Elsevier.
Terhune, Todd (1997). Pop songs: Myths and realities. *The English Connection* 1(1):8–12.
Tesink, Cathelijne M. J. Y.; Karl Magnus Petersson; Jos J. A. van Berkum; Daniëlle van den Brink; Jan K. Buitelaar; & Peter Hagoort (2009). Unification of speaker and meaning in language comprehension: An fMRI study. *Journal of Cognitive Neuroscience* 21(11):2085–99.
Trudgill, Peter (1983). Acts of conflicting identity: The sociolinguistics of British pop-song pronunciation. In Peter Trudgill, *On dialect: Social and geographical perspectives*, 141–60. Oxford: Basil Blackwell and New York: New York University Press.
Van Berkum, Jos J. A.; Daniëlle van den Brink; Cathelijne M. J. Y. Tesink; Miriam Kos; & Peter Hagoort (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience* 20(4):580–91.
Weatherholtz, Kodi; Kathryn Campbell-Kibler; & T. Florian Jaeger (2014). Socially-mediated syntactic alignment. *Language Variation and Change* 26(3):387–420.
Werner, Valentin (2012). Love is all around: A corpus-based study of pop lyrics. *Corpora* 7(1):19–50.
Wolfram, Walt (1982). Language knowledge and other dialects. *American Speech* 57(1):3–18.
Woolard, Kathryn (1998). Introduction: Language ideology as a field of inquiry. In Bambi B. Schieffelin, Kathryn A. Woolard, & Paul V. Kroskrity (eds.), *Language ideologies: Practice and theory*, 3–47. New York: Oxford University Press.
Yildirim, Ilker; Judith Degen; Michael K. Tanenhaus; & T. Florian Jaeger (2016). Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language* 87:128–43.

(Received 23 October 2017; revision received 08 June 2018; accepted 28 June 2018; final revision received 10 July 2018)

Address for correspondence:
Lauren Squires
The Ohio State University
421 Denney Hall
164 Annie & John Glenn Ave.
Columbus, OH 43210
squires.41@osu.edu