Variation in the strength of lexical encoding across dialects

Cynthia G. Clopper

Ohio State University

Terrin N. Tamati

University Medical Center Groningen

Janet B. Pierrehumbert

University of Oxford e-Research Centre

Corresponding Author:
Cynthia G. Clopper
Department of Linguistics
Ohio State University
108A Ohio Stadium East
Columbus, OH 43210, USA

Email: clopper.1@osu.edu
Phone: 1-614-292-8235
Fax: 1-614-292-8833

Abstract

Lexical processing is slower and less accurate for unfamiliar dialects than familiar dialects. The goal of the current study was to test the hypothesis that dialect differences in lexical processing reflect differences in lexical encoding strength across dialects. Lexical encoding (i.e., updating the cognitive lexical representation to reflect the current token) was distinguished from lexical recognition (i.e., mapping the incoming acoustic signal to the target lexical category) in a series of lexical processing tasks with Midland and Northern American English. The experiments were conducted in the Midland region with Midland and Northern listeners. The results confirmed differential processing of the two dialects: the Midland dialect was processed more quickly than the Northern dialect. The results further revealed significantly larger repetition benefits (i.e., priming) and cross-dialect lexical interference effects for lexical forms in the Midland dialect than in the Northern dialect for both listener groups, particularly when the stimulus materials were presented in noise. These results suggest that lexical information is more strongly encoded for the contextually-local Midland dialect than for the non-local Northern dialect. We interpret these effects as reflecting cognitive processing costs associated with normalization for dialect variation, which lead to weaker lexical encoding under more difficult processing conditions.

Keywords: dialect variation; word recognition; recognition memory; lexical encoding

1. Introduction

1.1. Lexical recognition and encoding

Lexical processing involves mapping the incoming acoustic signal to the target lexical category (i.e., recognition) and updating the cognitive lexical representation to reflect the current token (i.e., encoding; Norris, McQueen, & Cutler, 2003). In an exemplar approach, lexical representations are groups of individual tokens that have been previously recognized and encoded (Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2001). For example, the representation of the word *bad* is the group of tokens that have been previously recognized and stored as exemplars of the category *bad*. The recognition process involves identifying the category to which the current token is most similar and the encoding process involves adding the current token to the existing representation for the target word (Goldinger, 1998; Johnson, 1997; Pierrehumbert, 2001). Lexical representations vary in their robustness as a function of the number and strength of the exemplars comprising them (Johnson, 1997; Pierrehumbert 2002, 2006). For example, frequent words have more robust representations than less frequent words because frequent words have more exemplars comprising their representations. Similarly, recently encountered words have more robust representations than less recently encountered words because recently encountered words have stronger (i.e., less-decayed) exemplars comprising their representations. Robust representations facilitate the word recognition process, leading to more accurate recognition for more frequent (Broadbent, 1967; Howes, 1957) and more recently encountered (Goldinger, 1996) words.

Lexical recognition and encoding are partially independent processes. Although a failure to recognize a target word also leads to a failure to encode the exemplar in memory, recognition does not necessarily lead to encoding. For example, a listener may fail to encode an exemplar

when the target word is correctly recognized, but a mispronunciation is detected (Garrett & Johnson, 2012). Lexical encoding can also vary in strength, so that even recent exemplars may contribute different weights to the overall representation. For example, “typical” variants are encoded more robustly than “atypical” variants (Nygaard, Burt, & Queen, 2000), and even atypical variants can be robustly encoded when they are “idealized” or socially salient (Sumner, Kim, King, & McGowan, 2014). Thus, variants associated with socially salient dialects are predicted to be more robustly encoded than variants associated with less salient dialects (Sumner et al., 2014). The goal of the current study was to explicitly test this prediction about differences in the strength of lexical encoding across dialects as a function of the listeners’ experience with regional dialect variation.

1.2. Lexical processing across dialects

Lexical processing is faster and more accurate for native and standard dialects than for unfamiliar, non-standard dialects (Clopper, Pierrehumbert, & Tamati, 2010; Floccia, Goslin, Girard, & Konopczynski, 2006; Impe, Geeraerts, & Speelman, 2008; Labov & Ash, 1997; Sumner & Samuel, 2009). This processing benefit for familiar dialects parallels the processing benefits observed for high-frequency words (Broadbent, 1967; Howes, 1957) and follows directly from the structure of exemplar-based lexical representations.¹ In particular, phonological variation is captured in this approach by linking individual exemplars to multiple different types of categories (Pierrehumbert, 2002). For example, an individual token of the word *bad* may be

¹ Following the previous research on which this study is based, we assume an exemplar-theoretic framework. However, the primary results may also be explained by an abstract model of lexical representation, in which the recognition and encoding processes can be distinguished. For example, in a model involving activation and competition among abstract lexical units, such as Cohort Theory (Marslen-Wilson & Welsh, 1978), TRACE (McClelland & Elman, 1986), or Shortlist (Norris, 1994), variation in encoding across dialects could be captured as variation in residual activation of lexical categories following successful recognition.

1
2
3
4 linked to the lexical category for *bad*, as well as to an individual talker category (e.g., the first
5 author's husband) and to a dialect category (e.g., Northern American English).
6
7

8
9 For an individual listener, different dialects can be familiar in different ways. The native
10 variety is robustly represented and therefore easy to process because it has been encountered
11 very frequently, including in the listener's own speech. The contextually-relevant local variety is
12 also robustly represented and easy to process because it has been encountered recently in
13 interactions with local talkers. The relevant standard varieties are robustly represented and easy
14 to process because they are socially salient (Sumner et al., 2014). Varieties that are less familiar
15 and not socially salient are less robustly represented due to limited exposure to and weaker
16 encoding of their variants and therefore exhibit a cost in lexical processing. We propose that this
17 processing cost reflects the normalization process, which in an exemplar approach involves
18 mapping an unfamiliar variant to the target lexical representation during recognition (Johnson,
19 1997). This processing cost then leads to less robust encoding of dialect-specific variants, even
20 when recognition is successful (see also Bent & Holt, 2013).
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 Sumner and Samuel (2009) obtained some preliminary evidence that differences in
39 encoding across dialects may be central to the observed processing costs associated with less-
40 robustly represented dialects. The dialects under investigation were the non-rhotic New York
41 City (NYC) variety and the rhotic General American variety. In a long-distance priming lexical
42 decision task (their Experiment 3), the stimulus words in the first block served as primes for the
43 targets in the second block. For the "overt NYC" listeners, who spoke the non-rhotic variety, a
44 repetition benefit was observed for all prime-target pairs, consistent with robust encoding of both
45 native and standard forms. For the "covert NYC" listeners, who spoke a rhotic variety but had
46 substantial experience with the non-rhotic variety, a repetition benefit was observed for both
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 rhotic and non-rhotic targets following only rhotic primes. This pattern is consistent with cross-
5
6 dialect processing costs associated with encoding non-salient variants: despite substantial
7
8 exposure, the non-rhotic primes were less robustly encoded than the native rhotic primes and
9
10 therefore did not facilitate later recognition of the same lexical forms. The “General American”
11
12 listeners, who had substantial exposure only to the rhotic variant, exhibited priming only for
13
14 rhotic prime-target pairs, consistent with their overall high error rates for the non-rhotic targets
15
16 (24% errors in some conditions), suggesting that they did not recognize (or, therefore, encode)
17
18 many of the non-rhotic targets as real words.
19
20
21
22

23
24 In the current study, we further explored the interaction between talker and listener
25
26 dialect in lexical encoding. Our study differs from Sumner and Samuel’s (2009) study in two
27
28 critical respects. First, we adopted a recognition memory paradigm with a word recognition task
29
30 to allow us to explicitly distinguish between the recognition and encoding processes (see section
31
32 1.3). Second, we examined cross-dialect word recognition for two varieties of American English
33
34 that exhibit subphonemic vowel variation, rather than categorically distinct variants, to ensure
35
36 that an equally-high proportion of the target words were correctly identified by all listeners,
37
38 allowing for an examination of encoding following successful recognition.
39
40
41
42

43
44 The dialects we examined were Midland and Northern American English. The Northern
45
46 dialect, spoken in the upper midwestern United States, is characterized by the Northern Cities
47
48 vowel shift (Labov, 1998; Labov, Ash, & Boberg, 2006), which is schematized in the right panel
49
50 of Figure 1. The Midland dialect, spoken in the lower midwestern United States, is characterized
51
52 by the fronting and lowering of /ɔ/ (approaching, but not necessarily merging with, /ɑ/) and back-
53
54 vowel fronting (Labov, 1998; Labov et al., 2006), as shown in the left panel of Figure 1.
55
56
57
58
59
60
61
62
63
64
65

Figure 1. Schematics of the vowel systems of the Midland (left) and Northern (right) dialects of American English. The phonetic symbols indicate the idealized variants of “standard” American English and the arrows indicate the direction and magnitude of the dialect-specific shifts in the F1 x F2 space.

The experiment was conducted with Midland and Northern listeners in the Midland region to allow us to examine the potential interactions among different types of dialect familiarity. The Midland dialect is a subdialect of General American English (Labov, 1998) and has fewer dialect-specific vowel shifts that distinguish it from standard American English than the Northern dialect (see Figure 1). The shifts in the Midland dialect also generally preserve the structure of the vowel space, with the exception of the potential low-back merger, whereas the Northern Cities shift results in rotation and restructuring of the low and front parts of the vowel space, respectively. Thus, in the experimental context, the Midland dialect was both contextually relevant as the local dialect and socially salient as the objectively more standard variety for all listeners. For the Midland listeners, the Midland dialect is also the native dialect, whereas for the

Northern listeners, the Northern dialect is the native dialect. Thus, the Midland listeners were predicted to exhibit more robust encoding of the Midland dialect than the Northern dialect, whereas the Northern listeners were predicted to exhibit robust encoding of both dialects.

1.3. Assessing lexical encoding strength

Testing the prediction that dialects exhibit variation in the strength of their lexical encoding requires distinguishing between the recognition and encoding processes. The lexical recognition memory paradigm is ideal for making this distinction and was therefore applied in a series of two experiments in the current study. In Experiment 1, we conducted an explicit lexical recognition memory task (Craig & Kirsner, 1974; Goldinger, 1996; Palmeri, Goldinger, & Pisoni, 1993; Schacter & Church, 1992). In the study phase of this task, participants identified words in an open-set word recognition task by typing the words that they heard into a computer.² Although this task involves additional processing beyond recognition (i.e., mapping the recognized word to its orthographic representation), it requires the listeners to report the words that they heard, allowing us to confirm that the target words were recognized as intended by the talkers. This confirmation of successful recognition is critical, given that recognition errors may block encoding or lead to the encoding of targets as the wrong word.

In the test phase of the explicit recognition memory task, participants were presented with words from the study phase, as well as with words that they had not previously heard in the experiment, and were asked to identify each word as “old” if they had heard it in the study phase or as “new” if they had not heard it in the study phase. This task allows us to assess encoding during the study phase: words that are encoded in the study phase should be identifiable as “old”

² A typing task was used because typed responses can be objectively scored, unlike spoken responses which require impressionistic transcription. In the context of an experiment involving word recognition across dialects, an objective assessment of word recognition is critical for assessing recognition accuracy.

1
2
3
4 in the test phase because their representation includes a recent exemplar that is contextually
5
6 relevant to the experiment, whereas words that are not recognized and/or encoded in the study
7
8 phase should be misidentified as “new” because their representation does not include a recent,
9
10 contextually-relevant exemplar.
11
12

13
14 The nature of the recognition memory task therefore allows us to determine the locus of
15
16 the cross-dialect lexical processing costs observed in previous work (e.g., Clopper et al., 2010;
17
18 Floccia et al., 2006; Labov & Ash, 1997; Sumner & Samuel, 2009). In particular, if the cost is
19
20 associated with less accurate recognition of less familiar variants, we should observe poorer
21
22 performance overall in both the study and test phases for a less familiar dialect relative to a more
23
24 familiar dialect because both tasks involve mapping the acoustic signal to a lexical representation
25
26 (i.e., recognition). However, under this account, recognition memory should differ across
27
28 dialects only as a function of recognition accuracy. That is, recognition in the test phase should
29
30 be equivalent for all words that were correctly recognized in the study phase because the
31
32 mapping between the signal and the lexical form has already been established. Alternatively, if
33
34 the cost is associated with the strength of the encoding of the studied word, we should observe
35
36 reduced recognition memory when the studied word involves an unfamiliar variant relative to a
37
38 familiar variant because the lexical form of the unfamiliar variant will be only weakly encoded in
39
40 the study phase and therefore provide less facilitation during recognition of the same word in the
41
42 test phase.
43
44
45
46
47
48
49

50
51 Previous research using explicit recognition memory tasks has demonstrated that when
52
53 the same word is presented in the study and test phases, the similarity between the two tokens of
54
55 that word impacts the speed and accuracy of “old” identification judgments. For example, words
56
57 are easier to identify as “old” if they are presented in the same voice than in different voices
58
59
60
61
62
63
64
65

(Craig & Kirsner, 1974; Goldinger, 1996; Palmeri et al., 1993) and when different voices are used, voice similarity is correlated with recognition accuracy (Goldinger, 1996; cf. Goh, 2005; Palmeri et al., 1993). This pattern of results is consistent with the exemplar approach outlined in section 1.1. An exemplar that is encoded during the study phase is strongly weighted during the test phase because it was recently experienced, leading to easier identification of itself as “old.” Identifying a similar token as “old” is also supported by a recently encoded exemplar by facilitating the similarity-driven recognition process (Goldinger, 1996).

However, this effect of similarity on explicit recognition memory may be asymmetric. For example, Nygaard et al. (2000) observed significant repetition benefits for studied hyperarticulated forms that were repeated in both hyper- and hypoarticulated styles, but a repetition benefit for studied hypoarticulated forms was only observed when they were repeated in the hypoarticulated style. The forms were presented in all possible pairs, which means that the acoustic-phonetic similarity between the study and test forms was equivalent across conditions. Thus, this asymmetry in the similarity effect suggests that stronger encoding of hyperarticulated forms leads to repetition benefits for both identical and similar tokens, whereas weaker encoding of hypoarticulated forms limits repetition benefits to identical study and test tokens.

Given that our primary goal was to explore variation in encoding strength across dialects and that presenting identical tokens during the study and test phases has the potential to result in recency effects regardless of encoding strength, we manipulated token similarity for the repeated items across the study and test blocks of the experiment. Following Goldinger (1996), who compared repetitions by the same talker, by a different talker of the same gender, and by a different talker of the other gender, we examined recognition memory for words produced in the study and test phases by the same talker, by a different talker from the same dialect, and by a

1
2
3
4 different talker from the other dialect. We expected to replicate the finding that same talker
5 repetitions facilitate explicit recognition memory relative to different talker repetitions (Craik &
6
7 Kirsner, 1974; Goldinger, 1996; Palmeri et al., 1993). We also expected to observe a same-
8
9 dialect effect, in which repetitions by a different talker from the same dialect facilitate
10
11 identification of target words as “old” relative to repetitions by a different talker from a different
12
13 dialect (see also Sumner & Samuel, 2009). Finally, we expected larger effects of dialect
14
15 familiarity on lexical encoding to emerge as the similarity between the study and test tokens
16
17 decreased, consistent with the finding that stronger encoding leads to greater repetition benefits
18
19 for non-identical tokens (Nygaard et al., 2000). That is, the interaction between talker dialect and
20
21 listener dialect should be greatest when the study and test tokens are least similar (i.e., produced
22
23 by talkers from different dialects) and smallest when the study and test tokens are identical (i.e.,
24
25 produced by the same talker). Thus, both listener groups were expected to exhibit repetition
26
27 effects for both talker dialects in the same talker condition, in which the study and test tokens
28
29 were identical. However, the Midland listeners should exhibit stronger repetition effects in the
30
31 test phase for studied Midland targets than for studied Northern targets in the different talker and
32
33 different dialect conditions, in which the similarity between the study and test tokens is reduced.
34
35 The differences in repetition effects across talker dialects should be attenuated or even reversed
36
37 for the Northern listeners in these two conditions, given their familiarity with both dialects.
38
39
40
41
42
43
44
45
46
47

48 The subphonemic variation observed across the two talker dialects in the current study
49
50 allows us to examine a second dimension of encoding strength, beyond overall dialect
51
52 familiarity. Within each dialect, the vowel shifts shown in Figure 1 do not result in categorical
53
54 vowel mergers, with the exception of the potential merger of /ɔ, ɑ/ in the Midland dialect.
55
56 However, across dialects, this variation in the vowel system leads to perceptual phonological
57
58
59
60
61
62
63
64
65

1
2
3
4 confusions (Clopper et al., 2010; Jacewicz & Fox, 2012). For example, Northern variants of /ɪ, ʌ,
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

confusions (Clopper et al., 2010; Jacewicz & Fox, 2012). For example, Northern variants of /ɪ, ʌ, α/ may be ambiguous with /ɛ, ɔ, æ/, respectively, whereas Midland variants of /ɔ/ may be ambiguous with /ɑ/. In the study phase in Experiment 1, the words were presented without noise and we therefore expected that both listener groups would be able to overcome the potential phonological ambiguities and correctly recognize the target words (see also Figure 2). For the Midland listeners, however, the shifted Northern variants should be “atypical” and therefore only weakly encoded following recognition (Nygaard et al., 2000). We therefore included “catch” trials in the test phase of the experiment to explore the nature of the encoding of these potentially ambiguous variants. These catch trials involved new words that had not been presented in the study phase, but that formed minimal pairs with studied words. These minimally-paired study and test words differed only in their vowels and were selected to exhibit potential cross-dialect phonological confusions, based on previous descriptions of the dialects (Labov, 1998; Labov et al., 2006) and our own analysis of the stimulus talkers’ speech (see Figure 2; Clopper et al., 2010). Thus, we examined whether new test words that formed minimal pairs with potentially ambiguous study words were misidentified as “old” in the test phase. Just as stronger encoding during the study phase is expected to lead to stronger repetition benefits when the same word is presented in the test phase, stronger encoding is expected to lead to more interference when a minimally-paired word is presented in the test phase.

The design of the explicit recognition memory task in Experiment 1 thus allowed us to examine differential effects of dialect on lexical encoding as a function of listener dialect and experience, study-test token similarity, and the phonological ambiguity of the dialect-specific variants. However, although explicit recognition memory tasks have been shown to elicit robust token similarity effects (Craig & Kirsner, 1974; Goldinger, 1996; Palmeri et al., 1993), these

1
2
3
4 effects are stronger when the materials are presented in noise because overall processing is more
5
6 difficult, making it easier to observe improvement due to repetition (Goldinger, 1996). Similarly,
7
8 token identity (i.e., same vs. different talker) effects on long-distance repetition priming are
9
10 stronger when the study task is more difficult (McLennan & Luce, 2005). Further, recognition
11
12 memory is better overall when the same task is used in the study and test phases because the
13
14 same processes are applied at study and test (Goldinger, 1996), whereas tasks requiring speeded
15
16 responses (such as the explicit old/new recognition task in the test phase of our Experiment 1)
17
18
19 lead to weaker token identity effects (McLennan & Luce, 2005).
20
21
22

23
24 Given that our goal was to observe variation in encoding strength across dialects as a
25
26 function of listener experience, Experiment 2 was designed to enhance our ability to observe
27
28 token similarity effects and, by extension, variation in those effects due to encoding strength.
29
30 Specifically, in Experiment 2, we used an implicit lexical recognition memory task involving a
31
32 word recognition task in noise in both the study and test phases to further probe the effect of
33
34 dialect variation on lexical encoding strength. In this task, the improvement in word recognition
35
36 accuracy between the study and test phases was used to assess encoding: words that are strongly
37
38 encoded during the study phase should be easier to recognize when they are presented again in
39
40 the test phase than words that are more weakly encoded. Thus, unlike in Experiment 1, in which
41
42 participants' explicit old/new recognition judgments were used to assess encoding, in
43
44 Experiment 2, performance was compared across the study and test phases to provide an implicit
45
46 measure of recognition memory and lexical encoding (Church & Schacter, 1994; Goldinger,
47
48 1996; Schacter & Church, 1992). The addition of noise to the study phase increases the difficulty
49
50 of the task, which may lead to stronger repetition effects. The use of the same, non-speeded word
51
52 recognition task in noise in the test phase should also enhance the effect of token similarity and
53
54
55
56
57
58
59
60
61
62
63
64
65

other factors on the magnitude of the repetition effect. The remaining design factors were identical to Experiment 1 and our predictions were therefore parallel for the two experiments.

2. Experiment 1: Explicit lexical recognition memory

2.1. Listeners

Forty-two Ohio State University undergraduates (10 male, 32 female; 18-26 years old) participated as listeners in Experiment 1. They were all monolingual native speakers of American English with no reported hearing or speech disorders and were divided into two groups based on their residential history. The Midland (N=21) and Northern (N=21) listeners had lived exclusively in their respective dialect regions before age 18.

2.2. Stimulus materials

The stimulus materials were drawn from a pool of 168 familiar consonant-vowel-consonant English words (see Appendix). The words were produced by six female talkers (19-21 years old) from the Indiana Speech Project corpus (Clopper et al., 2002). Three talkers were lifetime residents of the Northern dialect region, with both parents also raised in the North. The remaining three talkers were lifetime residents of the Midland dialect region, with both parents also raised in the Midland. The mean first and second formant frequencies estimated at the temporal midpoint of the vowels in the stimulus words produced by the Midland and Northern talkers are shown in Figure 2. As expected, the Northern talkers show evidence of the Northern Cities shift, including lowering and backing of /ɪ, ɛ, ʌ/, raising and fronting of /æ/, and fronting and lowering of /ɑ/. Both talker dialects exhibit lowering and fronting of /ɔ/, as expected, as well as fronting of /u/, although this latter property is not typically described as a characteristic of the

Northern dialect (see Figure 1). The Midland talkers also show evidence of /ow/ fronting, but not /ʊ/ fronting. Thus, the stimulus materials exhibit the dialect-specific variants that are expected to produce cross-dialect phonological confusions and lead to differential lexical encoding across talker and listener dialects.

Figure 2. Mean F1 and F2 vowel spaces for the Midland and Northern talkers. Error bars are standard error of talker means.

In the study block, participants were presented with eleven different words produced by each of the six talkers, for a total of 66 study trials. In the test block, participants were presented

1
2
3
4 with the same 66 words from the study block as well as 66 additional words that were not
5
6 presented in the study block, for a total of 132 test trials. The 66 words from the study block
7
8 were “old” in the context of the experiment because they had previously been presented and the
9
10 66 additional words were “new” because they had not previously been presented. Although 66
11
12 words were presented in both the study and test blocks of the experiment, no words were
13
14 repeated within the study block or within the test block.
15
16
17

18
19 To explore the interaction between token similarity and dialect familiarity on lexical
20
21 recognition memory, one-third of the old words in the test block were produced by the same
22
23 talker as in the study block (Same Talker condition), one-third were produced by a different
24
25 talker from the same dialect region (Different Talker condition), and one-third were produced by
26
27 a different talker from the other dialect region (Different Dialect condition). Overall, each talker
28
29 produced approximately one-sixth of the old words in the test block. The old trials in the test
30
31 block were therefore characterized by a 3 (Trial Type: Same Talker, Different Talker, Different
32
33 Dialect) x 2 (Talker Dialect: Midland, Northern) design, with 11 trials per cell.
34
35
36
37

38
39 To explore the effects of cross-dialect phonological confusions on interference in lexical
40
41 recognition memory, approximately one-sixth of the new words in the test block formed minimal
42
43 pairs with words in the study block. These minimally-paired new words were predicted to be
44
45 confusable with their corresponding study words based on the vowel systems of the two talker
46
47 dialects (see Appendix).³ All of these minimally-paired new words were produced in the test
48
49 block by a talker from the other dialect than the corresponding word in the study block. Overall,
50
51 each talker produced approximately one-sixth of the new words in the test block. Thus, the new
52
53 trials in the test block were characterized by a 2 (Trial Type: Minimally-Paired, Other) x 2
54
55
56
57

58 ³ Given the large number of phonological CVC neighbors in English, many of the words in the stimulus
59 pool formed minimal pairs with other words in the pool. An analysis of the effect of phonological
60 neighborhood density on the results is presented in the Appendix.
61
62
63
64
65

(Talker Dialect: Midland, Northern) design. The trial type factor was unbalanced (approximately 11 Minimally-Paired and 55 Other new words per list) given constraints of the potential phonological confusion patterns across dialects and of the English lexicon.

To reduce the effects of dialect-specific tokens on the results, four separate stimulus lists were created to counter-balance talker dialect over items. To reduce talker-specific effects, talker-word pairings and talker pairings in the Different Talker and Different Dialect conditions were also varied across lists, but were not strictly counter-balanced to maintain a manageable number of lists. To reduce word-specific effects, old and new items were varied across lists, but were not strictly counter-balanced, to allow for the inclusion of the minimally-paired new words. Given that the stimulus materials were not fully counter-balanced across lists, the statistical analysis included both by-subject and by-item analyses. Each listener received one stimulus list, with 9-12 listeners per list. Table 1 provides examples from a single experimental list of the assignment of target words and talkers to the study and test phases of the experiment.

		Test Trial Type				
		Old			New	
Study Target	Test Target	Same Talker	Different Talker	Different Dialect	Minimally-Paired	Other
geese (MidlandA)	geese	MidlandA				
fish (NorthB)	fish	NorthB				
shake (MidlandB)	shake		MidlandA			
head (NorthC)	head		NorthB			
gap (NorthA)	gap			MidlandC		
cop (MidlandC)	cop			NorthA		
pig (NorthC)	peg				MidlandB	
bag (MidlandB)	beg				NorthC	
n/a	touch					MidlandC
n/a	hope					NorthA

Table 1. Examples from one stimulus list of the trial types in the recognition memory experiment.

The stimulus words were contained in individual digital sound files with a 22,050 Hz sampling rate and 16-bit resolution. The target words were presented without noise in both the study and test phases.

2.3. Procedure

Participants were seated at personal computers equipped with high-quality circumaural headphones, a keyboard, and a button box. In the study phase, participants were asked to listen to each word and respond by typing the word that they heard. No time limit was imposed for responding and the next trial was initiated when the participant pressed the Enter key.

Participants were encouraged to guess if they were unsure. They were not told that they would be tested on their memory for the study phase words. Prior to the analysis, responses were corrected for obvious typographical errors. Homophones were scored as correct (e.g., *doug* for *dug*).

In the test phase, participants were asked to listen to each word and press a button indicating “old” if they had heard the word in the study phase, regardless of the voice it was presented in, or a button indicating “new” if they had not heard the word in the study phase. They were asked to respond as quickly as possible without compromising accuracy. Button responses and response times were collected. Response times were measured from the onset of each stimulus item. Response times were analyzed for correct trials only. Prior to analysis, response times that exceeded 2.5 standard deviations beyond the participant mean were excluded (3.5% of correct responses).

2.4. Statistical analysis

A series of by-subject and by-item repeated measures ANOVAs was conducted to assess our predictions.⁴ First, we conducted an analysis of overall word recognition accuracy in the study block to determine the extent to which recognition accuracy was comparable across the two talker dialects for the two listener groups. Second, we conducted a series of analyses to explore the effects of dialect familiarity and token similarity on explicit lexical recognition memory accuracy and response times for the old words in the test block. Third, we conducted a series of analyses to explore the effects of cross-dialect phonological confusions on interference in lexical recognition memory for the new words in the test block.

To further explore the effect of token similarity, we also conducted a by-item analysis with listener dialect and acoustic distance between the study and test tokens as factors. The acoustic distance measure was based on the acoustic analysis described in section 2.2 and summarized in Figure 2. Acoustic distance was defined as the Euclidean distance in the F1 x F2 vowel space in Bark between the study and test tokens of each target word. The analysis of recognition memory accuracy for the old words revealed a significant main effect of acoustic distance ($F_2(1, 118) = 6.97, p = .009$, partial $\eta^2 = .06$). As expected, the effect was negative: as acoustic distance between the study and test tokens increased, recognition accuracy decreased. However, this effect was driven primarily by the difference between the Same Talker trials (where acoustic distance = 0) and the other trial types. The effect of acoustic distance was not

⁴ We conducted repeated measures ANOVAs rather than mixed-effects regression models with crossed subject and item random effects for consistency with Experiment 2. As described in section 3.4, the dependent variable in Experiment 2 was a difference score reflecting improvement in word recognition from the study phase to the test phase. Logistic mixed-effect modeling with crossed subject and item random effects is inappropriate for this type of dependent variable because the trial-by-trial difference score (i.e., test accuracy – study accuracy by subject and item) would result in three possible values (-1, 0, 1) reflecting four outcomes (0-1, 1-1 or 0-0, 1-0, where incorrect is scored as 0 and correct is scored as 1). This variable is uninterpretable because 0 reflects two different outcomes (i.e., correct in both blocks and incorrect in both blocks).

significant in a parallel analysis of only the Different Talker and Different Dialect trials. The effect of acoustic distance was also not significant in the response time analysis. To explore the effect of token similarity on phonological interference, we conducted a parallel by-item analysis on the Minimally-Paired trials with listener dialect and Euclidean distance in the F1 x F2 vowel space in Bark between the study and test tokens of the minimally-paired words as factors. The effect of acoustic distance was not significant in the analysis of either recognition memory accuracy or response times for the minimally-paired words.⁵ The acoustic distance measure will therefore not be discussed further.

2.5. Results and discussion

2.5.1. Word recognition in the study phase

Overall word recognition accuracy in the study block was 87% for the Midland talkers and 89% for the Northern talkers, confirming that both talker dialects were highly intelligible when presented without noise. A pair of by-subject and by-item repeated measures ANOVAs on word recognition accuracy in the study block with talker dialect and listener dialect as factors revealed no significant main effects or interactions. Thus, both talker dialects were highly intelligible for both listener groups. This consistently high level of word recognition accuracy across conditions in the study phase allows us to interpret the test phase results in terms of encoding because any differences in the test phase cannot be attributed to differences in word recognition accuracy during the study phase.

⁵ The limited effect of the acoustic distance measure on recognition memory performance may reflect its failure to capture many potentially important dimensions of token similarity, including vowel trajectory and duration, consonantal phenomena, voice quality, and prosody.

2.5.2. Explicit recognition memory: Old words

A summary of the old/new recognition accuracy for old words (i.e., identifying the old words as “old” in the test block) is shown in Figure 3 for each of the three trial types (Same Talker, Different Talker, Different Dialect) for each of the two talker dialects (Midland, Northern) for each of the two listener dialects (Midland, Northern). A pair of by-subject and by-item repeated measures ANOVAs was conducted on old/new recognition accuracy for the old words with trial type, talker dialect, and listener dialect as factors. The main effect of trial type was significant in both analyses ($F_1(2, 80) = 7.41, p = .001$, partial $\eta^2 = .16$; $F_2(2, 114) = 7.25, p = .001$, partial $\eta^2 = .11$) and the talker dialect x trial type interaction was significant by subjects ($F_1(2, 80) = 3.33, p = .041$, partial $\eta^2 = .08$). None of the other main effects or interactions were significant.

Figure 3. Mean old/new recognition accuracy for the old words in the test block for each trial type (Same Talker, Different Talker, Different Dialect) for each talker dialect (Midland, Northern) for the Midland listeners (left) and the Northern listeners (right). Error bars are standard error of subject means.

Post-hoc t-tests confirmed that across talker dialects performance was more accurate for Same Talker trials than either Different Talker (Subjects: $t(41) = 3.86, p < .001$; Items: $t(103) = 2.07, p = .041$) or Different Dialect (Subjects: $t(41) = 2.51, p = .016$; Items: $t(96) = 2.05, p = .043$) trials, but performance on the Different Talker and Different Dialect trials did not differ. The significant talker dialect x trial type interaction reflects significant differences in accuracy between all three trial types for Midland talkers (all $t(41) > 2.03, p < .05$), but no significant effects of trial type on recognition accuracy for Northern talkers. As shown in Figure 3 for both listener dialects, for the Midland talkers, accuracy was highest for Same Talker trials, followed by Different Dialect trials, followed by Different Talker trials. Although no effects of trial type were observed for the Northern talkers, post-hoc one-sample t-tests confirmed that old/new recognition accuracy was significantly above chance in all three conditions for both talker dialects (all $t(41) > 3.02, p < .005$), suggesting strong encoding of all target words during the study phase.

A summary of the old/new recognition response times for correct old trials is shown in Figure 4 for each trial type for each talker dialect for each listener group. A pair of by-subject and by-item repeated measures ANOVAs was conducted on old/new recognition response times for the old words with trial type, talker dialect, and listener dialect as factors. The main effect of trial type was significant in both analyses ($F_1(2, 80) = 4.05, p = .021$, partial $\eta^2 = .09$; $F_2(2, 110) = 3.34, p = .039$, partial $\eta^2 = .06$). The main effect of talker dialect was significant by subjects ($F_1(1, 40) = 17.31, p < .001$, partial $\eta^2 = .30$) and marginal by items ($F_2(1, 98) = 3.42, p = .068$, partial $\eta^2 = .03$). None of the other main effects or interactions were significant.

As shown in Figure 4, responses were faster for Midland tokens than for Northern tokens across all three trial types for both listener groups. This finding replicates previous research demonstrating that less standard dialects (such as Northern American English) are processed more slowly than more standard dialects (such as Midland American English; e.g., Floccia et al., 2006). Post-hoc paired-sample by-subject t-tests confirmed that performance was slower for Different Talker trials than for either Same Talker ($t(41) = 2.90, p = .006$) or Different Dialect ($t(41) = 2.18, p = .035$) trials, but performance on the Same Talker and Different Dialect trials did not differ. Post-hoc independent-sample by-item t-tests revealed only a marginal difference between the Different Talker and Same Talker trials ($t(103) = 1.92, p = .057$). The other two by-item pairwise comparisons for trial type were not significant.

Figure 4. Mean old/new recognition response times for the old words in the test block for each trial type (Same Talker, Different Talker, Different Dialect) for each talker dialect (Midland, Northern) for the Midland listeners (left) and the Northern listeners (right). Error bars are standard error of subject means.

1
2
3
4
5
6
7 The main effects of trial type on recognition accuracy and response times for the old
8
9 words in the test phase reveal a consistent pattern of higher accuracy and faster response times
10
11 for Same Talker trials than Different Talker trials, replicating previous studies (Craig & Kirsner,
12
13 1974; Goldinger, 1996; Palmeri et al., 1993). However, the results for the Different Dialect trials
14
15 were unexpected. In the accuracy analysis, Different Dialect trials were more accurate than
16
17 Different Talker trials for Midland talkers and in the response time analysis, Different Dialect
18
19 trials were faster than Different Talker trials across talker groups. Together, these results suggest
20
21 that identifying a target word as “old” is easier when the study and test tokens are produced by
22
23 different talkers from different dialects than when they are produced by different talkers from the
24
25 same dialect. Further analysis of talker and/or token similarity, both within and across dialects,
26
27 may provide a clearer understanding of the role of token similarity in this task (Goh, 2005;
28
29 Goldinger, 1996; Johnson, 2006).
30
31
32
33
34
35
36
37

38 2.5.3. *Explicit recognition memory: New words*

39

40
41 The overall old/new recognition accuracy and response times for the Minimally-Paired
42
43 and Other new word targets in the test phase are shown for each talker dialect for each listener
44
45 group in Table 2. A pair of by-subject and by-item repeated measures ANOVAs on old/new
46
47 recognition accuracy for the new words with trial type (Minimally-Paired, Other), talker dialect,
48
49 and listener dialect as factors was conducted. The main effect of trial type was significant in both
50
51 analyses ($F_1(1, 40) = 27.04, p < .001$, partial $\eta^2 = .40$; $F_2(1, 124) = 22.13, p < .001$, partial $\eta^2 =$
52
53 $.15$). The main effect of listener dialect was significant by items ($F_2(1, 125) = 30.17, p < .001$,
54
55 partial $\eta^2 = .19$) and marginal by subjects ($F_1(1, 40) = 3.25, p = .079$, partial $\eta^2 = .08$). None of
56
57
58
59
60
61
62
63
64
65

the other main effects or interactions were significant. Performance was less accurate for the Minimally-Paired than the Other new words, suggesting interference for minimally-paired test targets from their corresponding study targets. Performance was also more accurate for the Northern listeners than the Midland listeners, suggesting that exposure to multiple dialects facilitates performance in the explicit recognition memory task.

Midland Listeners	Minimally-Paired New Words		Other New Words	
Talker Dialect	Accuracy (%)	RT (ms)	Accuracy (%)	RT (ms)
Midland	58 (24)	1170 (265)	76 (25)	1142 (135)
North	57 (27)	1125 (301)	78 (25)	1229 (209)
Northern Listeners	Minimally-Paired New Words		Other New Words	
Talker Dialect	Accuracy (%)	RT (ms)	Accuracy (%)	RT (ms)
Midland	64 (27)	1398 (618)	86 (20)	1190 (393)
North	70 (23)	1216 (269)	86 (18)	1131 (140)

Table 2. Mean old/new recognition accuracy and response times in the test block for Minimally-Paired and Other new words produced by Midland and Northern talkers for Midland (top) and Northern (bottom) listeners. Standard deviations are shown in parentheses.

A pair of by-subject and by-item repeated measures ANOVAs on correct old/new recognition response times for the new words with trial type (Minimally-Paired, Other), talker dialect, and listener dialect as factors was also conducted.⁶ The trial type x listener dialect interaction was significant in both analyses ($F_1(1, 36) = 4.16, p = .049$, partial $\eta^2 = .10$; $F_2(1, 123) = 8.45, p = .004$, partial $\eta^2 = .06$). The talker dialect x trial type interaction was significant by items ($F_2(1, 99) = 4.54, p = .036$, partial $\eta^2 = .04$) and marginal by subjects ($F_1(1, 36) = 3.83, p = .058$, partial $\eta^2 = .10$). The main effects of trial type ($F_2(1, 120) = 10.21, p = .002$, partial η^2

⁶ Prior to the analysis, data from four participants who had zero observations in one or more cells of the design were excluded.

= .08) and talker dialect ($F_2(1, 99) = 4.60, p = .034$, partial $\eta^2 = .04$) were also significant by items. None of the other main effects or interactions were significant. Responses were slower for Northern talkers than Midland talkers, consistent with the results for the old words. In addition, responses were slower for the Minimally-Paired than the Other new words, consistent with the main effect of trial type observed in the accuracy analysis and providing further evidence for strong encoding of both talker dialects by both listener groups during the study phase.

The talker dialect x trial type interaction reflects a significant by-item effect of trial type for the Midland talkers ($t(112) = 2.57, p = .012$), but not for the Northern talkers. Responses for the Midland talkers were slower for the Minimally-Paired than the Other new words, consistent with the overall main effect. The trial type x listener dialect interaction reflects a significant effect of trial type for the Northern listeners (Subjects: $t(19) = 1.92, p = .070$; Items: $t(127) = 2.75, p = .007$), but not for the Midland listeners. For the Northern listeners, responses were slower for Minimally-Paired than for Other new words, consistent with the overall main effect. Although the three-way interaction was not significant, an inspection of Table 2 suggests that both of the significant two-way interactions may be driven primarily by the slow response times by the Northern listeners to the Midland Minimally-Paired targets. This pattern is unexpected from the perspective of encoding because the Northern listeners should robustly encode both talker dialects and therefore exhibit comparable degrees of interference from the Minimally-Paired targets in the test phase. Further, although the interactions appear to suggest stronger encoding of the native Northern dialect than the non-native Midland dialect for the Northern listeners, a parallel effect was not observed for the Midland listeners. Thus, these interactions may reflect processing strategy differences across listener groups rather than differential encoding of the two talker dialects. In particular, the Northern listeners may take longer than the

1
2
3
4 Midland listeners to make judgments in the old/new recognition task for the Midland Minimally-
5
6 Paired targets because multiple lexical candidates are activated over a longer period of time
7
8
9 (Clopper & Walker, in press).
10

11 Taken together, the old word analysis revealed above-chance old/new recognition
12
13 accuracy in all conditions and the new word analysis revealed significant interference in both
14
15 accuracy and response times for Minimally-Paired targets across talker and listener dialects.
16
17 Both of these results suggest strong encoding of both talker dialects during the study phase for
18
19 both listener groups. The lack of strong evidence for differential encoding across dialects may
20
21 reflect properties of the task. As discussed in section 1.3, token similarity effects on recognition
22
23 memory and long-distance priming are larger when the task is more difficult, when the same task
24
25 is used in the study and test phases, and when a non-speeded task is used (Goldinger, 1996;
26
27 McLennan & Luce, 2005). In Experiment 2, we examined implicit lexical recognition memory
28
29 using a more difficult, non-speeded word recognition task in noise in both the study and test
30
31 blocks to further probe the effects of talker and listener dialect on lexical encoding. We
32
33 compared performance across the study and test blocks to obtain an implicit measure of
34
35 recognition memory for the studied forms.
36
37
38
39
40
41
42
43
44

45 **3. Experiment 2: Implicit lexical recognition memory**

46 *3.1. Listeners*

47
48 Forty-two Ohio State University undergraduates (16 male, 26 female; 17-23 years old)
49
50 participated as listeners in Experiment 2. They were all monolingual native speakers of
51
52 American English with no reported hearing or speech disorders and were divided into two groups
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 based on their residential history. The Midland (N=22) and Northern (N=20) listeners had lived
5
6 exclusively in their respective dialect regions before age 18.
7
8
9

10 11 *3.2. Stimulus materials* 12 13

14 The same stimulus materials and lists were used in Experiment 2 as in Experiment 1,
15
16 except that the stimulus words were mixed with speech-shaped noise at a signal-to-noise ratio of
17
18 +2 dB. Each sound file consisted of a 400 ms silent lead, followed by 500 ms of noise only, the
19
20 target word mixed with noise, and a 500 ms noise tail. Each listener received one stimulus list,
21
22 with 9-12 listeners per list.
23
24
25
26
27

28 29 *3.3. Procedure* 30

31 Participants were seated at personal computers equipped with high-quality circumaural
32
33 headphones and a keyboard. On each trial in both the study and test blocks, participants were
34
35 presented with a single word mixed with noise over the headphones and were asked to respond
36
37 by typing the word that they heard. No time limit was imposed for responding and the next trial
38
39 was initiated when the participant pressed the Enter key. Participants were encouraged to guess if
40
41 they were unsure. The trials within each block were randomized separately for each participant
42
43 and repetitions of words across the study and test blocks were separated by a mean of 98 trials
44
45 (SD = 42 trials). The study and test phases were separated by a short break. Prior to the analysis,
46
47 responses were corrected for obvious typographical errors. Homophones were scored as correct.
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

3.4 Statistical analysis

As in Experiment 1, a series of statistical analyses was conducted to assess our predictions. First, we conducted an analysis of overall word recognition accuracy in the study block to determine the extent to which recognition accuracy was comparable across the two talker dialects for the two listener groups. Second, we conducted a series of analyses to explore the effects of dialect familiarity and token similarity on the repetition benefit for the old words in the test block. The dependent variable in these analyses was a difference score calculated from the mean accuracy of the old words in each of the two blocks (test – study) of the experiment. Positive difference scores reflect a repetition benefit, whereas difference scores close to 0 reflect the lack of an effect of repetition on recognition accuracy. Third, we conducted a series of analyses to explore the effects of cross-dialect phonological confusions on interference in lexical recognition memory for the new words in the test block with word recognition accuracy in the test block as the dependent variable.

We also conducted a by-item analysis examining the effects of listener dialect and acoustic distance between the study and test tokens, defined as in Experiment 1, on the old trials and on the Minimally-Paired trials. The effect of acoustic distance was not significant in either analysis and this measure will not be discussed further.

3.4.1. Analysis of the word recognition error patterns

Unlike in Experiment 1, the responses in the test block in Experiment 2 were word recognition judgments. We were therefore able to examine what the listeners reported hearing in the study and test phases for more detailed evidence of differential lexical encoding across dialects in the study phase. The vowel in each typed word response was transcribed. Nonword

1
2
3
4 responses, multisyllabic responses, and homographs for which a unique vowel transcription is
5
6 not available (e.g., *dove*, *read*) were transcribed as “Other.”
7
8

9 For the old words in the test block, we examined the repetition benefit separately by
10
11 target vowel. We then created separate stimulus-response vowel confusion matrices for the study
12
13 phase and for the old words in the test phase to determine the possible locus of the repetition
14
15 benefits. Given that dialect familiarity effects were expected to emerge more strongly for non-
16
17 identical study-test pairs, we focused on the Different Talker and Different Dialect trials in this
18
19 analysis. Given that each listener responded to only one target word with each vowel in each
20
21 condition for each talker dialect, these analyses are presented qualitatively in section 3.5.2.
22
23

24
25 We conducted a similar qualitative analysis to examine the cross-dialect interference
26
27 effect on minimally-paired new words in the test block. As in the previous analysis, we
28
29 examined recognition accuracy separately by target vowel. We then created stimulus-response
30
31 vowel confusion matrices for the Minimally-Paired new words and the Other new words in the
32
33 test phase to provide a more detailed assessment of the interference arising from an encoded
34
35 study word on the recognition of a minimally-paired test word.
36
37
38
39
40
41
42

43 *3.5. Results and discussion*

44 45 *3.5.1. Word recognition in the study phase*

46
47 Overall word recognition accuracy in the study phase was 60% for the Midland talkers
48
49 and 68% for the Northern talkers. A pair of by-subject and by-item repeated measures ANOVAs
50
51 on word recognition accuracy in the study phase with talker dialect and listener dialect as factors
52
53 was conducted. The main effect of talker dialect was significant by subjects and marginal by
54
55 items ($F_1(1, 40) = 8.45, p = .006$, partial $\eta^2 = .17$; $F_2(1, 118) = 3.54, p = .062$, partial $\eta^2 = .03$).
56
57
58
59
60
61
62
63
64
65

The main effect of listener dialect was significant by items ($F_2(1, 118) = 5.14, p = .025$, partial $\eta^2 = .04$) and the talker dialect x listener dialect interaction was significant by subjects ($F_1(1, 40) = 4.28, p = .045$, partial $\eta^2 = .10$). None of the other main effects or interactions were significant. Accuracy was higher for the Northern talkers than the Midland talkers overall and for the Midland listeners than the Northern listeners overall. Post-hoc paired-sample t-tests revealed a significant effect of talker dialect for the Midland listeners ($M_{\text{Midland}} = 59\%$, $M_{\text{North}} = 70\%$, $t(21) = 3.60, p = .002$), but no effect of talker dialect for the Northern listeners ($M_{\text{Midland}} = 62\%$, $M_{\text{North}} = 66\%$).

The lack of a talker dialect effect on word recognition accuracy for the Northern listeners is expected, given their robust representations of both dialects. However, the Midland listeners' more accurate word recognition for the Northern talkers than the Midland talkers is surprising. A further inspection of vowel recognition accuracy reveals higher vowel accuracy for the Midland talkers than the Northern talkers for both listener groups (Midland listeners: $M_{\text{Midland}} = 82\%$, $M_{\text{North}} = 80\%$; Northern listeners: $M_{\text{Midland}} = 85\%$, $M_{\text{North}} = 78\%$), suggesting that the Midland listeners may have found this set of Northern talkers' consonants more recognizable in noise (see also Clopper et al., 2010).

Given the hypothesis that recognition failure leads to encoding failure (Garrett & Johnson, 2012), this difference in recognition accuracy during the study phase has the potential to impact our interpretation of the test phase results with respect to encoding. In particular, the Midland tokens may have been less strongly encoded than the Northern tokens in the study phase by the Midland listeners because they were recognized less accurately. However, the results of the test phase (see sections 3.5.2 and 3.5.3) provide evidence that this difference in recognition accuracy did not eliminate the predicted talker dialect differences in encoding.

3.5.2. *Implicit recognition memory: Old words*

A summary of the recognition accuracy difference scores between the study and test blocks is shown in Figure 5 for each of the three trial types (Same Talker, Different Talker, Different Dialect) for each of the two talker dialects for each of the two listener dialects. A pair of by-subject and by-item repeated measures ANOVAs on the repetition effect with trial type, talker dialect, and listener dialect as factors was conducted. The talker dialect x trial type interaction ($F_1(2, 80) = 10.85, p < .001, \text{partial } \eta^2 = .21$) and the talker dialect x trial type x listener dialect interaction ($F_1(2, 80) = 6.23, p = .003, \text{partial } \eta^2 = .13$) were significant by subjects. The main effect of listener dialect was significant by items ($F_2(1, 114) = 4.71, p = .032, \text{partial } \eta^2 = .04$). None of the other main effects or interactions were significant.

As shown in Figure 5, the talker dialect x trial type interaction reflects repetition benefits from study to test for the Midland Same Talker, Midland Different Talker, and Northern Different Dialect trials, but repetition effects close to 0% for the Northern Same Talker, Northern Different Talker, and Midland Different Dialect trials. To determine which of these repetition effects were significant, posthoc one-sample t-tests comparing the mean difference score for each talker dialect in each trial type to 0 were conducted. These posthoc tests revealed that the repetition effect was significantly greater than 0 for Midland Same Talker trials ($t(41) = 3.18, p = .002$), Midland Different Talker trials ($t(41) = 2.72, p = .010$), and Northern Different Dialect trials ($t(41) = 5.68, p < .001$). The repetition effect was not significant in any of the other conditions. Thus, the repetition benefit was limited to target words produced in the study phase by a Midland talker and a repetition benefit was not observed for target words produced in the

study phase by a Northern talker. This pattern of results suggests stronger encoding of studied Midland tokens than Northern tokens.

Figure 5. Mean proportion correct difference scores (mean accuracy during the test block – mean accuracy during the study block) for the old words for each trial type (Same Talker, Different Talker, Different Dialect) for each talker dialect (Midland, Northern) for the Midland listeners (left) and the Northern listeners (right). Error bars are standard error of subject means.

Figure 5 further shows that the Northern listeners exhibited larger repetition benefits in these three conditions than the Midland listeners, leading to the significant effect of listener dialect by items and the significant three-way interaction by subjects. Post-hoc independent-sample by-subject t-tests confirmed a significant difference in the magnitude of the repetition effect across listener dialects for the Northern Different Dialect trials ($t(40) = 2.36, p = .023$) and a marginal difference in the magnitude of the repetition effect across listener dialects for the Midland Different Talker trials ($t(40) = 1.83, p = .074$). None of the other pairwise comparisons

1
2
3
4 by listener dialect were significant. These findings are consistent with the listener group
5
6 differences in processing strategy proposed in section 2.5.3 (see also Clopper & Walker, in
7
8 press). In particular, Northern listeners may activate a larger set of exemplars when assessing the
9
10 similarity of the current token to existing representations, leading to larger cross-dialect
11
12 repetition benefits because the activated exemplars are more likely to include a wider range of
13
14 dialect-specific forms.
15
16
17

18
19 To more fully understand the effects of dialect variation on lexical encoding, we
20
21 examined the repetition benefits separately by vowel category. A summary of the repetition
22
23 benefits observed for each vowel category for the Different Talker and Different Dialect trials for
24
25 each talker dialect for each listener dialect is shown in Table 3. The table clearly shows a mixed
26
27 pattern of positive (i.e., repetition benefit/priming) and negative (i.e., repetition cost/interference)
28
29 values with a range of magnitudes. Given the sparsity of the data, we focus on absolute
30
31 difference scores greater than or equal to 0.20 and these cells are shaded in the table. The overall
32
33 pattern of repetition benefits and costs is consistent with the data shown in Figure 5. For the
34
35 Midland Different Talker trials, repetition benefits were observed for /æ, ɔ/ for the Midland
36
37 listeners and for /i, ɪ, ɔ, ow, u/ for the Northern listeners. The repetition benefits were offset only
38
39 by a repetition cost for /a/ for the Midland listeners, leading to an overall repetition benefit. For
40
41 the Northern Different Talker trials, repetition benefits were observed for /ɛ, ʊ/ for the Midland
42
43 listeners and for /ɛ, a, ʊ/ for the Northern listeners, but these benefits were offset by repetition
44
45 costs for /ej, ow, u/ for the Midland listeners and for /ej, ʌ, ow, u/ for the Northern listeners,
46
47 leading to an overall null repetition effect. For the Midland Different Dialect trials, repetition
48
49 benefits were observed for /ɛ, ɔ/ for the Midland listeners and for /ɛ, ɔ, ow/ for the Northern
50
51 listeners, but these benefits were offset by repetition costs for /i, u/ for the Midland listeners and
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 for /ɪ, ej, ʌ, u/ for the Northern listeners, leading to an overall null repetition effect. Finally, for
5
6 the Northern Different Dialect trials, repetition benefits were observed for /i, ɪ, ej, ʌ, ʊ/ for the
7
8 Midland listeners and for /ɪ, ej, ʌ, u/ for the Northern listeners, and these benefits were offset by
9
10 repetition costs only for /ɛ, ʌ/ for the Midland listeners and for /ʌ/ for the Northern listeners,
11
12 leading to an overall repetition benefit. Thus, more relative benefits were observed overall for the
13
14 Midland Different Talker and Northern Different Dialect conditions than for the other two
15
16 conditions, consistent with the pattern of significant repetition benefits observed in Figure 5.
17
18 More relative benefits were also observed for the Northern listeners than the Midland listeners,
19
20 consistent with the larger repetition benefits observed for the Northern listeners than the Midland
21
22 listeners overall.
23
24
25
26
27

28
29 Many of the vowels exhibiting repetition benefits are involved in an ongoing shift in one
30
31 or both of the two talker dialects (see Figure 1), which allows us to consider the effect of
32
33 phonological ambiguity on the repetition benefit. We therefore examined stimulus-response
34
35 vowel confusion matrices from the study and test phases to further explore these vowel-specific
36
37 benefits. For the Midland Different Talker trials, the repetition benefit for /æ/ was accompanied
38
39 by fewer confusions with /ɛ/, the repetition benefit for /ɔ/ was accompanied by fewer confusions
40
41 with /ʌ/, and the repetition benefit for /ow/ was accompanied by fewer confusions with /ʊ/. For
42
43 the Northern Different Talker trials, the repetition benefit for /ɛ/ was accompanied by fewer
44
45 confusions with /æ/, the repetition benefit for /ʌ/ was accompanied by fewer confusions with /ɔ/,
46
47 and the repetition benefit for /ʊ/ was accompanied by fewer confusions with /ʌ, ow/. Thus, for
48
49 the Different Talker trials, the repetition benefits primarily reflected fewer of the predicted
50
51 dialect-specific phonological confusions for the target dialect. Further, robust repetition benefits
52
53 for some vowel categories for the Northern Different Talker trials suggests that some Northern
54
55
56
57
58
59
60
61
62
63
64
65

study tokens were encoded strongly enough to produce repetition benefits for talkers from the same dialect in the test phase.

	Midland Talkers		Northern Talkers	
Midland Listeners	DT	DD	DT	DD
i	-0.05	-0.32	0.05	0.27
ɪ	0.18	-0.09	-0.09	0.20
ej	-0.18	-0.05	-0.45	0.32
ɛ	-0.14	0.59	0.41	-0.41
æ	0.32	-0.09	0.09	0.18
ɑ	-0.23	0.17	0.09	-0.38
ɔ	0.41	0.42	0.10	-0.18
ʌ	0.18	-0.09	-0.05	0.23
ow	-0.16	-0.18	-0.23	-0.09
ʊ	-0.09	-0.07	0.73	0.28
u	0.09	-0.23	-0.36	0.14
Northern Listeners				
i	0.30	-0.05	0.05	0.05
ɪ	0.20	-0.25	-0.05	0.21
ej	-0.05	-0.30	-0.25	0.20
ɛ	0.00	0.30	0.25	-0.10
æ	-0.15	-0.10	0.00	0.05
ɑ	0.05	0.07	0.20	-0.20
ɔ	0.35	0.40	-0.19	0.06
ʌ	0.15	-0.35	-0.30	0.45
ow	0.31	0.23	-0.24	0.15
ʊ	-0.10	-0.10	0.50	0.12
u	0.30	-0.35	-0.30	0.45

Table 3. Mean proportion correct difference scores (mean accuracy during the test block – mean accuracy during the study block) for the old words for each vowel category for the Different Talker (DT) and Different Dialect (DD) trials for the Midland and Northern talkers for the Midland (top) and Northern (bottom) listeners. Shaded cells have an absolute difference score ≥ 0.20 .

For the Different Dialect trials a different pattern emerges. For the Midland Different Dialect trials, the repetition benefits for /ε, ɔ, ow/ were not accompanied by clear improvements in vowel recognition accuracy. However, for the Northern Different Dialect trials, the repetition benefits for /ʌ, ʊ/ were accompanied by fewer confusions with /ʊ, ʌ/, respectively, and the repetition benefit for /ɪ/ was accompanied by fewer confusions with /æ/. Thus, cross-dialect reductions in dialect-specific phonological confusions are observed only for the Northern Different Dialect trials, suggesting stronger encoding of Midland tokens in the study phase, leading to robust repetition effects in the Northern Different Dialect trials in the test phase.

3.5.3. *Implicit recognition memory: New words*

The overall recognition accuracy for the Minimally-Paired and Other new word targets in the test phase is shown for each talker dialect for each listener dialect in Table 4. A pair of by-subject and by-item repeated measures ANOVAs with trial type (Minimally-Paired, Other), talker dialect, and listener dialect as factors was conducted. The main effect of trial type was significant in both analyses ($F_1(1, 40) = 32.76, p < .001$, partial $\eta^2 = .45$; $F_2(1, 115) = 5.25, p = .024$, partial $\eta^2 = .04$). The main effect of talker dialect ($F_1(1, 40) = 32.99, p < .001$, partial $\eta^2 = .45$) and the talker dialect x trial type interaction were also significant by subjects ($F_1(1, 40) = 31.01, p < .001$, partial $\eta^2 = .44$). The other main effects and interactions were not significant.

Midland Listeners		
Talker Dialect	Minimally-Paired New Words	Other New Words
Midland	66% (22%)	69% (10%)
North	43% (22%)	64% (11%)
Northern Listeners		
Talker Dialect	Minimally-Paired New Words	Other New Words
Midland	60% (18%)	67% (8%)
North	37% (20%)	72% (11%)

Table 4. Mean word recognition accuracy in the test block for Minimally-Paired and Other new words produced by Midland and Northern talkers for Midland (top) and Northern (bottom) listeners. Standard deviations are shown in parentheses.

As shown in Table 4, accuracy was lower for the Minimally-Paired new words than for the Other new words and for the Northern talkers than for the Midland talkers. However, both of these main effects were mediated by the talker dialect x trial type interaction. In particular, recognition of Minimally-Paired new words produced by the Northern talkers in the test phase was significantly less accurate than either recognition of Minimally-Paired new words produced by the Midland talkers ($t(41) = 6.19, p < .001$) or recognition of Other new words produced by either group of talkers ($t(41) = 7.85, p < .001$ for Midland Other new words; $t(41) 7.05, p < .001$ for Northern Other new words). None of the other pairwise comparisons were significant. Given that the Minimally-Paired new words in the test phase were produced by a talker from the other dialect than the corresponding word in the study block, the Minimally-Paired targets are comparable to the Different Dialect trials for the old words. The poor overall accuracy for the Northern Minimally-Paired new words (Table 4) is particularly striking in comparison to the large repetition benefit for the Northern Different Dialect old words (Figure 5). However, both results reflect the same strong encoding of Midland words during the study phase: strong

1
2
3
4 encoding leads to more accurate recognition of the same word, but greater interference for a
5
6 minimally-paired word.
7
8

9 A summary of the word recognition accuracy for each vowel category for the Minimally-
10 Paired and Other new trials for each talker dialect for each listener dialect is shown in Table 5. A
11
12 difference score (Other – Minimally-Paired) is also provided for each talker dialect for each
13
14 listener dialect. Positive difference scores reflect interference for the Minimally-Paired trials
15
16 (i.e., higher accuracy for the Other new words than the Minimally-Paired new words), whereas
17
18 negative difference scores reflect a processing benefit for the Minimally-Paired trials (i.e., higher
19
20 accuracy for the Minimally-Paired new words than the Other new words). Given the sparsity of
21
22 the data for the Minimally-Paired new words, we focus on absolute difference scores greater than
23
24 or equal to 0.20 and these cells are shaded in Table 5. The overall patterns of interference are
25
26 consistent with the data shown in Table 4. For the Midland talkers, interference was observed for
27
28 the /ʌ, ʊ/ Minimally-Paired new words for the Midland listeners and for the /ʌ/ Minimally-Paired
29
30 new words for the Northern listeners. This interference was partially offset by a benefit for /ɔ/ for
31
32 the Minimally-Paired new words for both listener groups. For the Northern talkers, interference
33
34 was observed for /æ, ɑ, ʊ/ for the Midland listeners and for /æ, ɛ, ʌ, ʊ/ for the Northern listeners.
35
36 This interference was partially offset by a benefit for /ʌ/ for the Midland listeners and a benefit
37
38 for /ɑ/ for the Northern listeners. Thus, interference was more widespread than benefits for the
39
40 Minimally-Paired new words for both talker dialects, consistent with the overall interference
41
42 effects shown in Table 4. Similarly, interference was more widespread for the Northern talkers
43
44 than the Midland talkers, consistent with the talker dialect effect on interference.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

	Midland Talkers			Northern Talkers		
Midland Listeners	MP	Other	Other - MP	MP	Other	Other - MP
æ	0.70	0.80	0.10	0.25	0.48	0.23
ɛ	0.82	0.65	-0.17	0.39	0.48	0.09
ɑ	0.17	0.32	0.15	0.25	0.60	0.35
ɔ	0.77	0.49	-0.28	0.63	0.68	0.06
ʌ	0.00	0.90	0.90	1.00	0.66	-0.34
ʊ	0.50	0.89	0.39	0.38	0.67	0.29
Northern Listeners						
æ	0.65	0.57	-0.08	0.37	0.64	0.27
ɛ	0.68	0.75	0.06	0.23	0.69	0.45
ɑ	0.40	0.48	0.08	0.60	0.33	-0.27
ɔ	0.90	0.49	-0.41	0.60	0.75	0.15
ʌ	0.00	0.81	0.81	0.44	0.73	0.28
ʊ	0.67	0.63	-0.04	0.40	0.86	0.46

Table 5. Mean word recognition accuracy for the Minimally-Paired (MP) and Other new words

in the test phase and mean word recognition accuracy difference scores (Other new words – Minimally-Paired new words) for each vowel category for the Midland and Northern talkers for the Midland (top) and Northern (bottom) listeners. Shaded cells have an absolute difference score ≥ 0.20 .

4. General discussion

4.1. Variation in the strength of lexical encoding across dialects

Taken together, the results of the two recognition memory experiments provide novel empirical evidence for differential lexical encoding across dialects, leading to processing costs for the more weakly encoded dialect. Specifically, the results suggest that the processing costs associated with the Northern dialect reflect weaker lexical encoding of Northern forms than Midland forms for both listener groups. In Experiment 1, explicit old/new recognition responses to Northern targets were slower, although not less accurate, than responses to Midland targets, consistent with previous research demonstrating cross-dialect processing costs in lexical access

(e.g., Floccia et al., 2006; Sumner & Samuel, 2009). In Experiment 2, stronger implicit recognition memory benefits and more minimal pair interference were observed for studied Midland tokens than for studied Northern tokens, suggesting that the lexical form of the Northern study tokens was not as strongly encoded in memory and therefore produced less facilitation in recognition of the same word and less interference in recognition of a phonologically similar word. The combined evidence for slower processing of Northern tokens and weaker encoding of lexical information for Northern tokens suggests that the lexical processing costs associated with some dialects reflect, at least in part, weaker overall lexical encoding and representation of those dialects. That is, the lexical information associated with exemplars from those dialects may be more weakly encoded so that recognition of similar exemplars in the future is less impacted by previous experience.

The observed pattern of repetition benefits across experiments clearly highlights the distinction between recognition and encoding. In Experiment 1, the words were presented without noise and both Northern and Midland targets were successfully encoded in memory, as evidenced by the lack of a talker dialect effect on accuracy in the test phase. In Experiment 2, the words were presented with noise and the Midland targets were more successfully encoded in memory, despite lower overall recognition accuracy in the study phase. Thus, the talker dialect effects on performance observed in the test phase reflect the extent to which the lexical forms were encoded in memory after recognition in the study phase, rather than the extent to which they were accurately recognized.

Our results further suggest a mechanism for the weaker encoding of particular dialects: for variation associated with dialects, the cost of normalization for this variation reduces the probability that the stimulus will be encoded, even if the lexical form is accurately identified.

1
2
3
4 When the stimulus materials were presented without noise in Experiment 1, recognition was
5
6 highly successful, but was slower for the Northern tokens than the Midland tokens, suggesting
7
8 that normalization for Northern variants requires more time than normalization for Midland
9
10 variants. The addition of noise to the signal in Experiment 2 reduced the available information in
11
12 the signal, making the normalization process more difficult, and leading to lower overall
13
14 recognition accuracy for both talker dialects. This additional cost associated with dialect
15
16 normalization in noise also led to the emergence of dialect differences in encoding (see also
17
18 Lewis-Peacock & Norman, 2014; Newman & Norman, 2010). The Midland variants were more
19
20 strongly encoded than the Northern variants, consistent with a greater normalization processing
21
22 cost for the Northern dialect than the Midland dialect. Thus, the results of our recognition
23
24 memory tasks suggest that even when the normalization process is successful and accurate
25
26 lexical access is achieved, the lexical form may not be encoded well enough to affect future
27
28 processing. These results also reveal two important factors impacting normalization processes
29
30 and, by extension, the strength of encoding across dialects: the presence of noise and talker
31
32 dialect.
33
34
35
36
37
38
39
40
41
42

43 *4.2. Effects of noise on lexical encoding across dialects*

44

45 The different patterns of recognition memory observed across dialects in the two
46
47 experiments reveal that the strength of the encoding of the lexical form across dialects is
48
49 critically dependent on task effects, including whether or not the targets are presented in noise.
50
51 Although noise-masking and dialect variation have differential consequences for speech
52
53 processing (Mattys, Davis, Bradlow, & Scott, 2012), noise-masking can highlight the costs
54
55 associated with normalization and encoding of lexical information in the context of talker or
56
57
58
59
60
61
62
63
64
65

1
2
3
4 dialect variation. For example, Church and Schacter (1994) observed significant repetition
5
6 benefits with low-pass filtered stimulus materials in their implicit recognition memory task, but
7
8 no significant repetition benefits with stimulus materials presented in the quiet in their explicit
9
10 recognition memory task, suggesting differential processing of talker-specific information as a
11
12 function of noise and/or task demands. More recently, Gilbert, Chandrasekaran, and Smiljanic
13
14 (2014) found stronger explicit recognition memory in noise when the speech to be remembered
15
16 was hyperarticulated. They interpreted this finding as reflecting stronger linguistic encoding
17
18 when less processing was required to accurately identify the targets, consistent with our proposal
19
20 that lexical information is encoded more robustly when the cost of normalization is smaller.
21
22 Furthermore, Gilbert et al. (2014) observed that overall intelligibility was not fully predictive of
23
24 encoding strength (see also Mattys & Liss, 2008), consistent with our observation from
25
26 Experiment 1 that processing dialect-specific information is costly even when word recognition
27
28 is ultimately successful. Our results provide new evidence that the processing costs associated
29
30 with normalization for dialect variation in noise play a critical role in lexical encoding strength:
31
32 the presence of noise in Experiment 2 imposed a greater processing cost on the Northern forms,
33
34 resulting in weaker encoding of the Northern forms in memory, relative to the Midland forms.
35
36
37
38
39
40
41
42
43
44

45 46 *4.3. Effects of talker dialect on lexical encoding* 47

48 The differential pattern of lexical encoding across the Midland and Northern talkers was
49
50 observed in this study for both Midland and Northern listeners, who were all living in the
51
52 Midland region at the time of the experiments. Although we expected the Midland listeners to
53
54 exhibit weaker encoding for the non-native Northern dialect than for their native Midland dialect,
55
56 we expected the Northern listeners to robustly encode both their native Northern dialect and the
57
58
59
60
61
62
63
64
65

1
2
3
4 more standard and contextually-local Midland dialect. The Northern listeners exhibited some
5
6 evidence of stronger cross-dialect interference for minimally-paired study and test words in
7
8 Experiment 1 and stronger cross-dialect repetition benefits for the old test words in Experiment
9
10 2, but no evidence of stronger encoding of Northern study words than the Midland listeners in
11
12 either experiment. From the perspective of an exemplar model in which frequent or familiar
13
14 forms are robustly represented, this result is striking because the native dialect should be robustly
15
16 encoded for our Northern listeners, who lived in their native Northern region until at least age 18.
17
18
19
20

21 One possible explanation for the observed pattern of results is that locally-relevant
22
23 dialects are recognized and encoded better than non-locally-relevant dialects, regardless of the
24
25 native dialect. The current study was conducted in the Midland dialect region, which means that
26
27 the contextually-local dialect was also the more standard dialect. Thus, the limited listener
28
29 dialect effect may reflect the location of the experiment in the Midland region, if the local dialect
30
31 context can outweigh the native dialect during processing (i.e., a recency effect outweighing a
32
33 frequency effect). However, the local context of Sumner and Samuel's (2009) experiment did not
34
35 facilitate priming of the local New York City variant for non-local listeners and the local context
36
37 of Clopper and Bradlow's (2008) study did not facilitate local Northern dialect intelligibility for
38
39 native Northern listeners. Thus, the local (i.e., regional) context may have the strongest impact
40
41 on processing when the local dialect is also relatively standard or when it is not superseded by a
42
43 smaller, local university context (see also Hay & Drager, 2010; Walker, 2014).
44
45
46
47
48
49

50 A processing benefit for standard dialect forms is also consistent with previous research
51
52 demonstrating greater form priming for canonical consonant variants (e.g., /nt, t/) than for
53
54 reduced consonant variants (e.g., /ɾ, ɾ, ʔ/; Ranbom & Connine, 2007; Ranbom, Connine, &
55
56 Yudman, 2009; Sumner & Samuel, 2005; cf., McLennan, Luce, & Charles-Luce, 2003). In
57
58
59
60
61
62
63
64
65

1
2
3
4 particular, the consonant variation results are interpreted as demonstrating the dominance of
5
6 canonical representations over reduced representations, despite the relatively high frequency of
7
8 reduced forms in natural speech. This interpretation directly parallels our finding that the more
9
10 standard Midland tokens were encoded more strongly than the less standard Northern tokens for
11
12 both listener groups. Sumner et al. (2014, p. 8) drew similar parallels between reduction
13
14 phenomena and dialect variation and made a distinction between socially-idealized, or standard,
15
16 forms and non-idealized, or non-standard, forms. They argued that infrequent standard forms
17
18 will be more robustly encoded than more frequent “default” or “subjectively standard” forms,
19
20 even when both forms are correctly recognized, and predicted that these encoding differences
21
22 will emerge in recognition memory tasks (p. 9). Thus, the limited listener dialect effect in our
23
24 study may reflect the inclusion of the relatively standard Midland dialect, if a standard dialect
25
26 can outweigh a non-standard native dialect during processing (i.e., a social salience effect
27
28 outweighing a frequency effect).⁷
29
30
31
32
33
34
35

36 Taken together, the observed talker dialect encoding differences and the limited evidence
37
38 for a listener dialect effect suggest that the strength of encoding and representation of dialect
39
40 variation does not directly reflect the listener’s previous dialect exposure. Further, although
41
42 ideologies about standard varieties may play a role, as proposed by Sumner et al. (2014),
43
44 additional factors, such as the local context, may also critically impact the strength of lexical
45
46 encoding across dialects.
47
48
49
50
51
52
53

54
55 ⁷ Although the Northern dialect is objectively linguistically less standard than the Midland dialect (Labov,
56 1998), the Northern dialect is not socially enregistered as non-standard (Campbell-Kibler, 2012). Thus,
57 the difference in encoding between the Midland and Northern tokens in our study cannot be attributed to
58 differences in subjective standardness because both varieties are ideologically standard for our
59 participants, and may instead reflect the mismatch between ideologies about standardness and the
60 phonetic reality of dialect variation in the American Midwest (see also Niedzielski, 1999).
61
62
63
64
65

4.4. *Conclusions*

We argue that the processing costs observed in cross-dialect lexical processing tasks reflect weaker representations of some dialects than others, resulting in part from differential encoding of dialect-specific variants, particularly when those variants are presented in noise. In the context of an implicit recognition memory task in noise, the weaker encoding of some dialect-specific variants results in reduced repetition benefits from study to test for those forms and reduced lexical interference for minimally-paired study and test forms. We propose that the mechanism underlying this effect reflects costs associated with normalization for dialect variation. In particular, lexical information is less strongly encoded when the costs associated with normalization for dialect variation are greater. The costs associated with normalization for dialect variation are impacted by properties of the experimental context, including the presence or absence of noise and the contextually-relevant local and/or standard dialects. For Northern American English listeners, these contextual factors can outweigh native-dialect biases and lead to weaker encoding of the native Northern dialect than of the non-native, but standard and contextually-relevant Midland dialect.

Acknowledgments

This work was partially supported by the NSF [BCS-1056409 to Ohio State University], the NIH [T32 DC00012 to Indiana University], and the John Templeton Foundation [Award 36617 to Northwestern University]. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. We would like to thank Sarah Glazier-Kipe, Rachel Monnin, and John Pate for research assistance.

Appendix

The pool of stimulus words in Experiments 1 and 2 is shown in (1). The minimally-paired new words, their corresponding study words, and the predicted vowel confusion for each word pair are shown in Table A.1.

- (1) bad, bag, balk, bat, bed, beef, beg, bet, boat, book, boot, boss, bowl, buck, bug, bull, bus, cab, calf, cap, caught, cave, chalk, cheek, chess, choose, code, coke, cook, cop, cot, cough, could, cuff, dad, date, dead, deaf, dean, dig, dock, dog, duck, dug, fate, fetch, fig, fish, food, foot, fought, gap, gas, gawk, geese, goat, good, guess, gym, had, hawk, head, heat, hitch, hoed, hog, hood, hoof, hoop, hope, huff, hug, hut, jet, juice, kiss, knead, knit, knob, lace, late, lit, loaf, log, look, loss, luck, lush, main, map, mass, match, maze, meek, mesh, mess, mob, moon, moose, mop, nab, net, niece, nod, pack, pad, page, pat, pave, peck, peg, pet, pig, pod, poke, pole, pull, put, putt, reap, red, rid, rough, rush, sack, sad, safe, sag, said, sap, sash, sat, scene, sell, set, shake, sheep, shock, shook, shoot, shop, shout, showed, shuck, sock, sought, soup, stalk, suit, talk, tape, taught, thief, tick, toad, took, touch, tub, tube, tug, wad, wage, wash, wedge, weed, wood, yacht, youth

Minimally-Paired New Word	Study Block Word	Vowel Confusion
balk	buck	$\Lambda > \text{ɔ}$
bed	bad	$\text{æ} > \text{ɛ}$
beg	bag	$\text{æ} > \text{ɛ}$
bet	bat	$\text{æ} > \text{ɛ}$
boss	bus	$\Lambda > \text{ɔ}$
cot	caught	$\text{ɔ} > \text{ɑ}$
dog	dug	$\Lambda > \text{ɔ}$
had	head	$\text{ɛ} > \text{æ}$
hog	hug	$\Lambda > \text{ɔ}$
map	mop	$\text{ɑ} > \text{æ}$
mess	mass	$\text{æ} > \text{ɛ}$
net	knit	$\text{ɪ} > \text{ɛ}$
pat	pet	$\text{ɛ} > \text{æ}$
putt	put	$\text{ʊ} > \Lambda$
pull	pole	$\text{ow} > \text{ʊ}$
red	rid	$\text{ɪ} > \text{ɛ}$
sack	sock	$\text{ɑ} > \text{æ}$
said	sad	$\text{æ} > \text{ɛ}$
sat	set	$\text{ɛ} > \text{æ}$

shuck	shook	$U > A$
-------	-------	---------

Table A.1. Minimally-paired new words, their corresponding words in the study block, and the predicted phonological confusions arising between the study and test words.

As shown in (1), many of the stimulus words formed minimal pairs with other words in the pool, given the large number of phonologically similar CVC words in English. Thus, the Minimally-Paired new words were not the only words that formed minimal pairs with other targets. An analysis of the overall neighborhood density (Nusbaum, Pisoni, & Davis, 1984) of the Minimally-Paired and Other new words revealed significantly more neighbors for the Minimally-Paired new words ($M = 22.2$, $SD = 9.0$) than the Other new words ($M = 17.0$, $SD = 6.1$), across stimulus lists ($F(1, 248) = 19.99$, $p < .001$). Similarly, an analysis of the number of phonological neighbors of the Minimally-Paired and Other new words in each stimulus list revealed significantly more experiment-internal neighbors for the Minimally-Paired new words ($M = 4.2$, $SD = 1.8$) than the Other new words ($M = 2.5$, $SD = 1.7$), across stimulus lists ($F(1, 248) = 29.05$, $p < .001$). Thus, overall neighborhood density and local neighborhood density within the experiment were included as covariates in the by-item analysis of the new words in both experiments. Given that overall neighborhood density is a significant predictor of local neighborhood density ($r^2 = .16$, $p < .001$), these two measures were included in separate models. In Experiment 1, neither effect of density was significant in any of the models, suggesting that the trial type effects observed for the new words reflect the inclusion of specific phonologically-confusable minimal pairs rather than overall or experiment-internal neighborhood density. In Experiment 2, the effect of overall neighborhood density was significant ($F_2(1, 115) = 9.28$, $p = .003$, partial $\eta^2 = .07$), but experiment-internal neighborhood density was not. The effect of overall neighborhood density was negative, as expected, because more neighbors led to lower

1
2
3
4 recognition accuracy. All of the reported effects in section 3.5.3 are independent of this
5
6 significant effect of overall neighborhood density on recognition of new words in the test block.
7
8
9

10 11 **References** 12

- 13
14 Bent, T., & Holt, R. F. (2013). The influence of talker and foreign-accent variability on spoken
15
16 word identification. *Journal of the Acoustical Society of America*, 133, 1677-1686.
17
18
19 Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, 74, 1-
20
21 15.
22
23
24 Campbell-Kibler, K. (2012). Contestation and enregisterment in Ohio's imagined dialects.
25
26 *Journal of English Linguistics*, 40, 281-305.
27
28
29 Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit
30
31 memory for voice intonation and fundamental frequency. *Journal of Experimental*
32
33 *Psychology: Learning, Memory, and Cognition*, 20, 521-533.
34
35
36 Clopper, C. G., & Bradlow, A. R. (2008). Perception of dialect variation in noise: Intelligibility
37
38 and classification. *Language and Speech*, 51, 175-198.
39
40
41 Clopper, C. G., Carter, A. K., Dillon, C. M., Hernandez, L. R., Pisoni, D. B., Clarke, C. M.,
42
43 Harnsberger, J. D., & Herman, R. (2002). The Indiana Speech Project: An overview of
44
45 the development of a multi-talker multi-dialect speech corpus. *Research on Spoken*
46
47 *Language Processing Progress Report No. 25*, 367-380. Bloomington, IN: Speech
48
49 Research Laboratory, Indiana University.
50
51
52
53 Clopper, C. G., Pierrehumbert, J. B., & Tamati, T. N. (2010). Lexical neighborhoods and
54
55 phonological confusability in cross-dialect word recognition in noise. *Laboratory*
56
57 *Phonology*, 1, 65-92.
58
59
60
61
62
63
64
65

- 1
2
3
4 Clopper, C. G., & Walker, A. (in press). Effects of lexical competition and dialect exposure on
5
6 phonological priming. *Language and Speech*.
7
8
9 Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition.
10
11 *Quarterly Journal of Experimental Psychology*, 26, 274-284.
12
13
14 Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb
15
16 speech processing? *Journal of Experimental Psychology: Human Perception and*
17
18 *Performance*, 32, 1276-1293.
19
20
21 Garrett, A., & Johnson, K. (2012). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins*
22
23 *of sound change: Approaches to phonologization* (pp. 51-97). Oxford: Oxford University
24
25 Press.
26
27
28 Gilbert, R. C., Chandrasekaran, B., & Smiljanic, R. (2014). Recognition memory in noise for
29
30 speech of varying intelligibility. *Journal of the Acoustical Society of America*, 135, 389-
31
32 399.
33
34
35
36 Goh, W. D. (2005). Talker variability and recognition memory: Instance-specific and voice-
37
38 specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,
39
40 31, 40-53.
41
42
43 Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and
44
45 recognition memory. *Journal of Experimental Psychology: Learning, Memory, and*
46
47 *Cognition*, 22, 1166-1183.
48
49
50 Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological*
51
52 *Review*, 105, 251-279.
53
54
55 Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48, 865-692.
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of
5
6 English words. *Journal of the Acoustical Society of America*, 29, 296-305.
7
8
9 Impe, L., Geeraerts, D., & Speelman, D. (2008). Mutual intelligibility of standard and regional
10
11 Dutch language varieties. *International Journal of Humanities and Arts Computing*, 2,
12
13 101-117.
14
15
16 Jacewicz, E., & Fox, R. A. (2012). The effects of cross-generational and cross-dialectal variation
17
18 on vowel identification and classification. *Journal of the Acoustical Society of America*,
19
20 131, 1413-1433.
21
22
23 Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K.
24
25 Johnson & J. W. Mullennix (eds.), *Talker variability in speech processing* (pp. 145-165).
26
27 San Diego: Academic Press.
28
29
30 Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity
31
32 and phonology. *Journal of Phonetics*, 34, 485-499.
33
34
35 Labov, W. (1998). The three dialects of English. In M. D. Linn (ed.), *Handbook of dialects and*
36
37 *language variation* (pp. 39-81). San Diego: Academic Press.
38
39
40 Labov, W., & Ash, S. (1997). Understanding Birmingham. In C. Bernstein, T. Nunnally & R.
41
42 Sabino (eds.), *Language variety in the south revisited* (pp. 508-573). Tuscaloosa, AL:
43
44 University of Alabama Press.
45
46
47 Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English*. New York: Mouton
48
49 de Gruyter.
50
51
52 Lewis-Peacock, J. A., & Norman, K. A. (2014). Competition between items in working memory
53
54 leads to forgetting. *Nature Communications*, 5(5768), 1-10.
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word
5
6 recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.
7
8
9 Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in
10
11 adverse conditions: A review. *Language and Cognitive Processes*, 27, 953-978.
12
13
14 Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there
15
16 is as much to learn from natural “oddities” as artificial normality. *Perception and*
17
18 *Psychophysics*, 70, 1235-1242.
19
20
21 McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive*
22
23 *Psychology*, 18, 1-86.
24
25
26 McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity
27
28 effects in spoken word recognition. *Journal of Experimental Psychology: Learning,*
29
30 *Memory, and Cognition*, 31, 306-321.
31
32
33 McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form.
34
35 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 539-553.
36
37
38 Newman, E. L., & Norman, K. A. (2010). Moderate excitation leads to weakening of perceptual
39
40 representations. *Cerebral Cortex*, 20, 2760-2770.
41
42
43 Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic
44
45 variables. *Journal of Language and Social Psychology*, 18, 62-85.
46
47
48 Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*,
49
50 52, 189-234.
51
52
53 Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive*
54
55 *Psychology*, 47, 204-238.
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4 Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier Mental Lexicon:
5
6 Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress*
7
8 *Report No. 10* (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana
9
10 University.
11
12
13
14 Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric
15
16 transfer in episodic memory for spoken words. *Journal of Experimental Psychology:*
17
18 *Learning, Memory, and Cognition*, 26, 1228-1244.
19
20
21 Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes
22
23 and recognition memory for spoken words. *Journal of Experimental Psychology:*
24
25 *Learning, Memory, and Cognition*, 19, 309-328.
26
27
28 Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J.
29
30 Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure* (pp. 137-
31
32 157). Amsterdam: John Benjamins.
33
34
35
36 Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.),
37
38 *Laboratory phonology 7* (pp. 101-139). Berlin: Mouton de Gruyter.
39
40
41 Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, 34, 516-530.
42
43
44 Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in
45
46 spoken word recognition. *Journal of Memory and Language*, 57, 273-298.
47
48
49 Ranbom, L. J., Connine, C. M., & Yudman, E. M. (2009). Is phonological context always used to
50
51 recognize variant forms in spoken word recognition? The role of variant frequency and
52
53 context distribution. *Journal of Experimental Psychology: Human Perception and*
54
55 *Performance*, 35, 1205-1220.
56
57
58
59
60
61
62
63
64
65

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 915-930.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, 4(1015), 1-13.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322-338.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60, 487-501.

Walker, A. (2014). *Crossing oceans with voices and ears: Second dialect acquisition and topic-based shifting in production and perception*. Doctoral dissertation, Ohio State University.