C3 CHAPTER 3

PHONOLOGICAL VARIATION AND LEXICAL FORM

C3.S1 RUARIDH PURSE, MEREDITH TAMMINGA, AND YOSIANE WHITE

3.1 What is phonological variation?

C3.P1 The **mental lexicon** is where we store our knowledge of the words in our language. A reasonable starting point is to think of entries in the mental lexicon as form–meaning pairs. The lexical entry for the word *cat*, for example, pairs (a) some semantic information about a household pet that meows (**meaning**), with (b) some information about the speech sounds used to refer to it (**form**). As a first pass, we could say that this form information is stored as a string of **phonemes**: /k æ t/.[1] The form side of this pairing is what allows speakers to externalize meaningful messages to the people around them and allows listeners to retrieve the intended meanings in turn. But a string of phonemes is, of course, an abstraction: what comes out of a speaker's mouth is sound waves shaped by articulatory gestures, and what a listener encounters is a continuous and complex acoustic signal. The physiological demands of this **phonetic implementation** mean that no two instances of a word in real speech are ever exactly the same, an observation known as the lack of invariance problem. Both speakers and listeners face the challenge of connecting a word's abstract lexically stored form with the continuous and multidimensional space of the phonetic implementation.

C3.P2 Often, however, words surface with multiple forms in ways that cannot be explained by the physiological demands of speech production. In the basic case, accounting for these forms is the domain of **phonology**. For example, words sometimes appear to change form when they are combined with certain suffixes. Consider the English word *confess*, which can be combined with the suffix *-ion* to make *confession*. The stem *confess*

[1] Phonemes are the distinctive sound units of a language; see Bacović et al. (this volume). These symbols are from the International Phonetic Alphabet, a system of representing speech sounds in writing.

shows up in different forms: *confession* has a palatal [ʃ] where *confess* has an alveolar [s]. Do English speakers store both forms in their lexicon and know to choose the [ʃ] form with the *-ion* suffix? Baković et al. (this volume) lay out the standard arguments that many linguists give for saying no, and instead analyzing the [ʃ] as an **allophone** (a predictable pronunciation alternant) derived by **phonological rule** rather than stored in the lexicon. A single rule can capture the generalization that [s] becomes [ʃ] in other stems that combine with *-ion*[2] (e.g., *express*/*expression*, *compress*/*compression*, and so on). On this view, phonological rules intervene between the lexicon and the phonetic implementation, editing the target segments that go on to be articulated in speech. This is not the only available model of the relationship between the lexicon, the phonology, and the phonetics,[3] but because it is a widely accepted framework, we will build our discussion around it.

C3.P3 This chapter is about **phonological variation**. But the aforementioned phonological alternation between [ʃ] and [s] is *not* generally referred to as phonological variation[4] because the rule is **obligatory** whenever the linguistic conditions that trigger it arise. Thus, the [ʃ] allophone is fully predictable from the linguistic environment. Pronouncing *confession* as [kənfɛsjən] (without having applied the rule) is simply not a well-formed option for English speakers. What, then, is phonological variation? First, while many phonological rules are obligatory, there are also cases where seemingly similar alternations are *not* fully predictable. To continue our current example, the phrase *impress you* can be pronounced as either [ɪmprɛsju] or [ɪmprɛʃju] in connected speech. The two forms look suspiciously like the input and output of our [ʃ]-deriving phonological rule, but in this case the speaker has a choice between the options. To give a more intuitively familiar example, words ending in unstressed /ɪŋ/ can optionally be pronounced with [ɪn]: *morning~mornin'*,[5] *pudding~puddin'*, *jumping~jumpin'*, *hypothesizing~hypothesizin'*, and so on. The same logic that led us to posit a general rule capturing the obligatory pattern of [s] alternating with [ʃ] might lead us to conclude that this optional variability in [s]~[ʃ] and [ŋ]~[n] is also the product of phonological rules—just not obligatory ones. This **intraspeaker variation**, where a given speaker may say the same thing in different ways, is one kind of phonological variation that we will cover in this chapter. The choice a speaker has between the different options is called a **variable**, and the options themselves are called **variants**. Phonological variables are often influenced by social and situational factors; for example, most English speakers will share the intuition that *mornin'* is a more casual way of saying *morning*. However, quantitative

[2] More precisely, a rule that palatalizes coronals before /j/-initial suffixes.

[3] A class of prominent alternatives is usage-based approaches to phonology, such as Exemplar Theory, in which episodic traces or "exemplars," prototypically word-level exemplars, are stored in memory and form the basis for the emergence of phonological categories or generalizations. The possibility of "hybrid" abstractionist/episodic frameworks has attracted increasing attention in recent years. See Pierrehumbert (2002), Pisoni and Levi (2007), and Hay (2018) for overviews.

[4] Even though colloquially the words "vary" and "alternate" seem to mean approximately the same thing.

[5] We use the ~ notation to mean "varies with."

sociolinguistic research supports the premise of **inherent variability** (Weinreich, Labov, and Herzog, 1968): that variant choice is not fully predictable, even with a hypothetically exhaustive understanding of an utterance's social context.

C3.P4 In addition to intraspeaker phonological variation, phonology can also differ across speakers, even when they are nominally speaking the same language. We refer to these differences as **interspeaker variation**. Speakers can differ from each other in many ways, but we will give special attention to interspeaker variation that involves phonological structure and lexical form. For example, two speakers may have different phonological rules in their grammars, different stored forms in their lexicon, or different **phonemic inventories** (the set of distinctive sounds in a language). One familiar source of interspeaker variation that we discuss at some length in Section 3.1 is regional dialects. Interspeaker variation may also reflect other aspects of a language user's background, such as gender, class, or race. From the point of view of language production, one might think that interspeaker variation need not be treated as variation at all: an American English speaker is probably never going to entertain the option of pronouncing the word *got* with a retroflex consonant, [gɑʈ], the way an Indian English speaker might.[6] But from the perspective of language comprehension, differences across speakers are a major contributor to the phonological variation in the input that listeners must accommodate.

C3.P5 Both intra- and interspeaker phonological variation can create a range of mismatches between real utterances and the stored forms in the lexicon. These mismatches pose a substantial challenge for lexical processing, including processes of word recognition (see Magnuson and Crinnion, this volume), word production (see Kilbourn-Ceron and Goldrick, this volume), and word learning (see Creel, this volume). In addition to its processing consequences, phonological variation raises new questions of representation. We have already pointed out that there are some parallels between obligatory phonological rules and intraspeaker phonological variables, but it does not *necessarily* follow that a non-obligatory phonological rule is the right analysis for any given variable; in some cases, there might be reason to believe that the options are stored in the lexicon, or arise in the phonetic implementation. Given the complexity of the challenges posed by phonological variation, it is unsurprising that models of the mental lexicon have largely set aside variable phenomena and have instead developed on the basis of invariant forms of words in isolation. However, as many other authors in this volume point out, variation is one of the major hurdles standing in the way of modeling the comprehension and production of words in their real-world context of continuous speech. As psycholinguists set their sights on increasingly realistic and dynamic models of how the mental lexicon is structured and used, it will become correspondingly worthwhile to tackle issues of phonological variation.

[6] This excludes speakers who command both varieties and may code-switch between them, as well as perhaps a narrow set of circumstances that we think can reasonably be set aside as marginal, such as deliberately performing a different accent.

C3.S2

3.1.1 The scope and aims of this chapter

C3.P6 We have already touched on a number of important themes, each of which alone could each easily fill a chapter on phonological variation: lexicon vs. phonology vs. phonetics; intraspeaker vs. interspeaker variation; comprehension vs. production; processing vs. representation. These dimensions could also be crossed with each other to form a very large space of interacting topics: the processing of intra- vs. interspeaker variation, the mental representation of intra- vs. interspeaker variation, and so on. An exhaustive treatment of this space is clearly beyond what could be accomplished in a single chapter. We therefore pursue a smaller set of more narrowly directed aims. Our overarching goal is to make explicit the relevance of phonological variation to the study of the mental lexicon. To do so, we connect the linguistic properties of phonological variation with their observed or potential consequences for lexical representation and, especially, processing.

C3.P7 First, in Section 3.2, we review experimental work showing that phonological variation comes into play during **lexical access**, when language users "look up" words in their mental dictionaries for use in production or comprehension. This literature, which focuses mostly on spoken word recognition, demonstrates that variation sometimes but not always interferes with word recognition, that listeners are able to rapidly accommodate even unfamiliar variation, and that listeners may in fact use phonological variation as extra information to guide lexical access. While these results show that there is some relationship between phonological variation and lexical access, they leave many questions about the nature of this relationship unanswered. We suggest that the path forward should take the linguistic properties of different types of phonological variation into account. Different variables may be represented differently, and have different structural consequences, as we outline in Section 3.3. In Section 3.4, we elaborate on how such differences have the potential to impact lexical processing and conclude by sharing our optimism about the advantages that incorporating phonological variation may offer for models of the mental lexicon.

C3.S3

3.2 How does phonological variation affect lexical access?

C3.P8 Although we have noted that psycholinguistic models of the mental lexicon have largely developed on the basis of isolated citation-form words, there are a number of lines of experimental work about how different pronunciations of words might impact lexical access. The evidence from this body of work provides a number of insights. Phonological variation does not necessarily disrupt lexical access; in fact, listeners are quite tolerant of licit variation in isolated words, and can flexibly adapt to novel accents characterized by many co-occurring phonological variables. Further, there is emerging evidence that

listeners can use phonological variation and their knowledge of a speaker's accent to facilitate spoken word recognition, suggesting that understanding phonological variation may ultimately help us understand lexical access processes.

C3.S4 ### 3.2.1 Variation need not impede word recognition

C3.P9 Variation creates mismatches between the form in a listener's lexicon and the phonetic form they perceive. **Canonical** forms are careful (or even hyperarticulated) pronunciations perceived as matching the "dictionary" pronunciation, while **non-canonical** forms diverge from that ideal in some respect.[7] A reasonable hypothesis following from this point is that a non-canonical pronunciation might delay how quickly a listener can access a word's lexical entry if the input diverges from the corresponding stored lexical form, whereas canonical forms might diverge less and thus be easier to access. An early example of a study supporting the hypothesis of a **canonicality advantage** in processing is Andruski, Blumstein, and Burton (1994). This study manipulated the voice onset time (VOT) of voiceless initial consonants in English, which tends to be relatively long in isolated words compared to connected speech. In a **semantic priming** task, where prior presentation of a semantically related prime word speeds recognition of a target word, primes with longer (i.e., more canonical) VOTs generated more priming than primes with short VOTs. Interestingly, this advantage only arose when the time between the prime and target was very short, pointing to a role for variation in the early stages of spoken word recognition. LoCasto and Connine (2002) found a similar advantage for words like *camera* with a canonical pronunciation vs. a non-canonical reduced schwa; Racine and Grosjean (2000, 2005) and Racine, Bürki, and Spinelli (2014) reported an advantage for canonical word forms in French with a schwa in the first syllable (e.g., *genou* = knee). Tucker and Warner (2007) found a small facilitation effect for words pronounced with a canonical word-medial /d/ or /g/ in comparison to reduced forms of those consonants.

C3.P10 Conversely, a number of studies using similar methods with other phonological variables have not found support for a canonicality advantage, instead finding apparently equivalent facilitation for canonical and non-canonical forms. One such variable is place assimilation, where a word-final consonant adopts the place of articulation of the following segment. For example, a word ending in a coronal stop /d/ such as *wicked* can (but need not) be pronounced with a labial stop [b] when the following word is labial-initial, as in [wɪkɪb pɹæŋk] for *wicked prank*. Gaskell and Marslen-Wilson (1996) show that non-canonical place-assimilated pronunciations yield just as much priming as the non-assimilated forms, contra the predictions of the canonicality advantage. Similar effects in which non-canonical pronunciations did not impede word

[7] Ultimately, a pronunciation's "canonicality" is a social construct; see Section 3.4.1 for additional discussion.

recognition have been found for nasal flapping (Ranbom and Connine, 2007; Pitt, Dilley, and Tat, 2011; Sumner, 2013, but cf. Pitt, 2009), voicing assimilation (Snoeren, Segui, and Hallè, 2008), and final /t/ allophony (Deelman and Connine, 2001; Sumner and Samuel, 2005).

C3.P11 Even studies failing to support the canonicality advantage, though, have generally found that listeners' tolerance for non-canonical pronunciations is not unbounded. Gaskell and Marslen-Wilson (1996) argued that their place assimilation results did not just reflect listeners' tolerance for mismatch: when the [wɪkɪb] pronunciation of *wicked* was followed by *game*, where [b] could not have been generated through place assimilation, it *did* inhibit lexical access (cf. Gow, 2001). In a slightly different vein, Sumner and Samuel (2005) found that various /t/ allophones produced naturally by speakers were accessed equally by listeners, but there was no priming from a minimally contrastive nonword prime (i.e., [flus] compared to non-canonical forms of *flute*). These results suggest that successful recognition of non-canonical forms depends on the variants being (a) possible pronunciations that are (b) licensed by context. In other words, the variation needs to represent surface patterns that should be familiar to listeners from their real-world listening experiences.

C3.P12 There is some evidence that the effect of context in facilitating lexical access of different pronunciations may be gradient. Tucker (2011), while finding a general advantage for canonical pronunciations, also found that the predictability of a non-canonical pronunciation in context improved its acceptability and speeded response times. This is supported in the domain of production by findings that French speakers produce words with non-canonical schwa omission faster as the relative frequency (i.e., predictability) of this form, compared to the canonical form, goes up (Bürki, Ernestus, and Frauenfelder, 2010). Another aspect of context that appears to modulate the recognition of non-canonical forms is speech style. In a study probing the conflicting results on whether nasal flapping (e.g., [splɪntə̬]~[splɪnə̬] for *splinter*) exhibits a canonicality advantage, Sumner (2013) observed equivalent priming from naturally produced canonical and non-canonical primes, but no priming from a non-canonical variant spliced into a carefully articulated word frame. In the latter case, the non-canonical variant was not licensed by the context of other acoustic cues to speech style within the word; Sumner proposes that careful and casual speech styles induce different processing modes, which may not be efficient for dealing with variants that are incongruous with the style. However, Bürki, Viebahn, Racine, Mabut, and Spinelli (2018) found that lexical decision latencies for words with and without schwa-reduction were equivalent across careful and casual speech styles.

C3.P13 The full set of results in this literature are challenging to reconcile completely, but that is to be expected for an area of such active inquiry. What we can take away at present is that hearing a word pronounced in a non-canonical way does not necessarily disrupt recognition of that word, and that listeners' ability to reconstruct or predict the variants from the surrounding context may facilitate the processing of phonological variation. However, this ability is contingent on the variation being consistent with listeners' social, stylistic, and linguistic experiences in the real world.

C3.S5

3.2.2 Listeners adapt even to pervasive and unfamiliar variation

C3.P14 The mixed support for the canonicality advantage hypothesis suggests that, on the whole, listeners are quite good at coping with variation so long as it is limited and familiar. But we might further inquire how variation influences word recognition processes when listeners encounter many different variable features at once, or when those features are not part of a listener's own production repertoire. Both of these situations are a consequence of interspeaker variation, particularly between regional varieties of a language. These regional sub-varieties of a language are commonly called **dialects**, and we can refer to the full set of a dialect's pronunciation features as an **accent**. Of course, the term "accent" can also be used to talk about the pronunciation of second-language speakers of a language. The notion of accent variation is intuitively familiar to listeners, who generally think of them in holistic terms; for example, a speaker might be said to have "a Southern accent" or "a French accent" even though the listener is unlikely to identify the cluster of features that give rise to that percept.

C3.P15 Naïve listeners have shown in perceptual categorization experiments that they are broadly able to identify where speakers are from, and that they can use specific acoustic-phonetic properties of talkers' speech to make their judgments (van Bezooijen and Gooskens, 1999; Clopper and Pisoni, 2004, 2006, 2007). This skill is not limited to adult listeners. Children as young as 12 months are sensitive to dialect differences while listening to speech (Schmale, Cristià, Seidl, and Johnson, 2010). In fact, by four years of age listeners are able to group similar talkers together and distinguish them from dissimilar talkers, with further major developmental improvements in classifying talkers by region happening between the ages of 7 to 11 (Jones, Yan, Wagner, and Clopper, 2017; Evans and Lourido, 2019).

C3.P16 A number of studies have shown that listeners have highly flexible word recognition processes that allow them to quickly adapt to both regionally accented (Best, Shaw, and Clancy, 2013; Maye, Aslin, and Tanenhaus, 2008, i.a.) and foreign accented speech (Clarke and Garrett, 2004; Bradlow and Bent, 2008; Witteman, Weber, and McQueen, 2013, 2014; Vaughn, 2019; Imai, Flege, and Walley, 2003; Bent and Frush Holt, 2013). Specifically, the evidence suggests that listeners experience a temporary disturbance when encountering a new accent, but this is normalized at an early stage of processing and improves over time (Floccia, Goslin, Girard, and Konopczynski, 2006; Goslin, Duffy, and Floccia, 2012). Adaptation is even possible for laboratory-created accents, so long as listeners are given enough exposure to them (Weatherholtz, 2015). Further, listeners' expectations about what accent they are going to hear from a novel talker can impact how successfully they adapt (Vaughn, 2019).

C3.S6

3.2.3 Listeners may use phonological variation to guide lexical access

C3.P17 If phonological variants do not dramatically inhibit word recognition and listeners can quickly come to understand different accents (even if we are not sure how they do it),

does the study of the mental lexicon really need to deal with phonological variation? The conclusion that real-world, appropriately contextualized variation never derails lexical access is probably premature; in Section 3.4 we will discuss cases where we think such issues are quite plausible. However, the canonicality advantage hypothesis does not capture the only possible way in which phonological variation might be relevant to lexical access. In fact, in this section we turn to the mounting experimental evidence that phonological variation can actually provide information to listeners that may help guide their lexical access processes.

C3.P18 Many of the studies in Section 3.2.1 not only fail to show negative consequences for non-canonical pronunciations but also that listeners actually use phonological variants to anticipate upcoming words (Bürki, 2018; Gow, 2001, 2002; Lahiri and Marslen-Wilson, 1991; Tucker, 2011, i.a.). In an eye-tracking study, Mitterer and McQueen (2009) presented Dutch listeners with words in sentential context, and with a deleted word-final /t/ to make them ambiguous (e.g., *tast* = 'touch' sounds like *tas* = 'bag'). Listeners used probabilistic knowledge of the likelihood of /t/ deletion in different following contexts to anticipate which image to look toward.

C3.P19 Besides cues from intraspeaker variation in the input, listeners can also use knowledge of a speaker's background to guide lexical access. Listeners presented with a word like *pants*, which has a different dominant meaning in the United Kingdom (*pants* = undergarment) vs. the United States (*pants* = trousers), used the accent of the speaker to facilitate semantic access to the congruent meaning of the word (Cai, Gilbert, Davis et al., 2017). Accents can also help listeners decode non-canonical pronunciations in isolated words. American listeners hearing /r/-final words pronounced either with consonantal /r/ in a General American accent (e.g., *slender* as [slɛndɚ]), or with vocalized /r/ in either a British or New York City accent (e.g., [slɛndə]) were able to use their knowledge of the British English accent to support their understanding of /r/-vocalized words in isolation in a priming task. However, /r/-vocalized words were much harder to recognize in the New York City accent (Sumner and Kataoka, 2013). Sumner, Kim, King, and McGowan (2014) argue that this is because certain language varieties are more salient or idealized than others, which facilitates lexical access.

C3.S7 3.2.4 Dimensions of phonological variation in lexical access

C3.P20 The body of work discussed in this section makes a strong start at examining the ways in which phonological variation affects lexical access. We suggest that a useful next step in understanding how variation is represented and processed could be to consider the ways in which phonological variation is not monolithic. One possible reason that the experimental studies discussed in this section have not reached a consensus on these issues is that they manipulate different phonological variables, which in turn have different representations and may interact differently with lexical access processes. While the

canonicality advantage literature has acknowledged and explored the point that gradient phonetic variation—such as the VOT manipulation of Andruski et al. (1994)—is probably different in some respect from variation between different phonemes—such as the variation between /d/ and /b/ in Gaskell and Marslen-Wilson (1996)—there is much more that could be said about the dimensions along which we could classify different kinds of phonological variation. In the next part of this chapter, we provide a non-exhaustive overview of some such dimensions, with an eye to how incorporating a more linguistically complex understanding of phonological variation might facilitate psycholinguistic research on the mental lexicon.

C3.S8

3.3 The complexity of phonological variation

C3.P21 We have just proposed that in the study of the mental lexicon, our ability to surmount the difficulties of phonological variation may be dependent on a detailed empirical understanding of the variability itself. In this section, we therefore outline the ways phonological systems can differ between speakers, as well as how the types of variation produced by individual speakers can be classified based on their representation. We primarily draw our examples from varieties of English due to the abundance of work on variation in Englishes and the fact that the majority of readers will have some frame of reference for these varieties. However, the typology of phonological phenomena that we describe is relevant to all language varieties. Taking representational differences into account can better position us to make realistic predictions about how phonological variation impacts lexical access. We begin by focusing on the linguistic descriptions in this section, and then expand on their possible processing consequences in Section 3.4.

C3.S9

3.3.1 Interspeaker variation

C3.P22 Different people speak differently, even when they are speaking what is ostensibly the same language. Many of the studies mentioned in Section 3.2 explored how different accents are processed. However, research of this type has not generally asked about the specific linguistic and (with some exceptions) social properties of accents. Here we cover four elements that characterize accents and can differ between them: representations for specific words, the phonetic realization of phonemes, the structure of the phonemic inventory, and allophonic processes.

C3.P23 In some cases, interspeaker phonological variation is lexically specific. A person from New York City and a person from London could both say the English word *tomato*, but the New Yorker is likely to say something like [təˈmeɪroʊ], while the Londoner is likely to say something like [təˈmɑːtoʊ]. Both are speaking English, and both are using the same

word in reference to the same (prototypically) edible red fruit, but the pronunciations are different: in most varieties of American English, the second syllable of *tomato* has a vowel like that in the word FACE[8], but in most British Englishes, the vowel in this syllable matches the one in PALM. This is not a phonological difference that generalizes to other words; the British English pronunciation of *mate* is not [mɑ:t]. Therefore, the same word must be represented with different phonemes depending on the variety. Lexically specific phonological differences can be found at a smaller scale, too. So far, we have juxtaposed accents such as "American English" vs. "British English." However, the Atlas of North American English (Labov, Ash, and Boberg, 2006) actually identifies seven major dialect regions across the United States and Canada (and British English may be even more internally diverse). Similar distinctions can be made between many of these accents for certain words; for example, for speakers in the northern half of New Jersey, the preposition *on* contains the LOT vowel while in southern New Jersey it has the THOUGHT vowel (Coye, 2009).

C3.P24 In addition to some lexically specific differences, a common difference between accents concerns the across-the-board phonetic realization of phonemes. In fact, the primary criteria used in the Atlas of North American English (Labov et al., 2006) to draw boundaries between dialect regions concern differences in vowel quality. For example, speakers from the eastern Great Lakes region in the north of the United States produce LOT with a fronted vowel compared to speakers from elsewhere. A fronted LOT is one that has, over time, developed into a vowel more like the one that most other American English varieties use in TRAP. In this LOT-fronting variety, TRAP is also shifted from an earlier position, so that LOT does not overlap with TRAP and the two vowels are still distinct. When two vowel shifts seem to push or pull each other along in the same direction like this, it is called a **chain shift**. Chain shifts can result in dramatic differences in the realization of phonemes while keeping all the categories intact. The two chain links we have described here (LOT fronting and TRAP tensing) are part of a larger chain called the Northern Cities Shift, which characterizes the accent of the Inland North dialect region (see Figure 3.1). A number of US dialects are involved in chain shifts like this one, which Labov (2012) argues are leading to greater regional differentiation than homogenization in the United States.

C3.P25 Beyond the phonetic properties of various phonemes[9], we can also ask about the number of phonemic categories in an accent and how they are organized. The main mechanisms by which a language variety comes to have a different number of phonemes are **mergers** and **splits**. There are many dialects in which some two phonemes have

[8] In dialectology and sociolinguistics, words in small capital letters are conventionally used to represent vowel phonemes, whatever the actual pronunciation of this vowel in a given variety. This kind of representation is called a lexical set (Wells, 1982) because it picks out the set of words that, historically, share the same phoneme.

[9] It could be argued that even if all the phonemes of two systems have quite different phonetic properties, it does not constitute a phonological difference so long as every phoneme in one system finds a structural equivalent in the other.

C3.F1

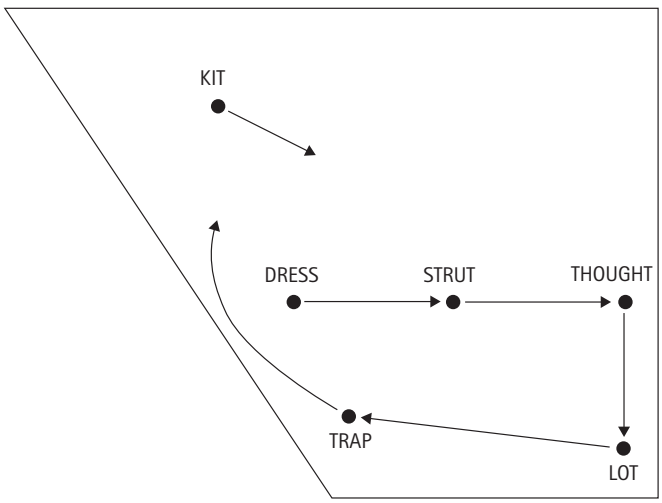

FIGURE 3.1 Vowel change trajectories comprising the Northern Cities Vowel Shift.

merged to form a single category. A well-known example in English involves the vowels in LOT and THOUGHT. In some dialects, the vowels in these words remain distinct. However, many dialects have undergone a merger in which these lexical sets have combined into a single large lexical set, reducing the overall number of contrastive phonemic categories in the inventory (for American English, see Labov et al., 2006, p. 58). Conversely, when one phoneme splits into two,[10] the number of contrastive phonemic categories increases. While mergers and splits constitute qualitative differences in the structure of the phonemic inventory, these effects are often not salient to the speakers themselves. Two speakers of American English varieties might disagree about whether words like *cot* and *caught* are pronounced the same, but they are unlikely to comment on or even notice this disagreement (Labov, 1994, p. 344).

C3.P26 Without reorganizing the underlying phonemic inventory, accents may exhibit different allophonic processes that affect how phonemes are realized in certain contexts (Labov, Fisher, Gylfadottir, Henderson, and Sneller, 2016; Sneller, 2018). This means rules like the variable palatalization process outlined in Section 3.1 may or may not exist in a given dialect. A useful example may be found in rhoticity: the pronunciation of /r/ when there is no following vowel (Scobbie, 2006). Most varieties of English spoken in the United States, Canada, Scotland, and Ireland are rhotic, meaning /r/ is always a consonant, typically an approximant like [ɹ]. However, many varieties of English spoken in England, Wales, Australia, and New Zealand are non-rhotic, meaning there is an obligatory allophonic rule turning /r/ into a vowel when it is followed by a consonant

[10] The mechanism for splitting is more complicated. Normally, it takes place in two stages: (1) an allophonic alternation stage, in which one phoneme is realized two different ways according to linguistic context, and (2) the loss of the triggering environment for allophony, so the different realizations are no longer in complementary distribution and must be reanalyzed as contrastive.

or a pause.[11] For other varieties still, most notably certain regional varieties of English from areas of the United Kingdom (Wells, 1970; French, 1989; Blaxter, Beeching, Coates, Murphy, and Robinson, 2019) and the United States (Labov, 1972, 2001; Feagin, 1990; Carmichael and Becker, 2019), consonantal or vocalic /r/ are both possible in instances of the same context. For speakers of these dialects, it makes sense to posit a variable rule that probabilistically turns non-prevocalic /r/ into a vowel.

C3.P27 Every person who has acquired language has acquired a particular accent, even if that accent happens to be held up as a prestigious or "standard" way of speaking. Prestigious language varieties are often assigned properties of neutrality or universality, but there is no objective linguistic basis for such a designation. For example, "standard" or "mainstream" American English is a non-uniform collection of varieties typically associated with white speakers from the Midwest or non-urban Northeast of the United States. Not only is the perspective that these varieties are objectively "neutral" born from racist and classist ideology, it conceals assumptions about linguistic representation and processing that should be interrogated. In Section 3.4.1, we unpack some possible consequences of failing to account for interspeaker variation.

C3.S10 3.3.2 Intraspeaker variation

C3.P28 We have just seen that a group of people all speaking the same language cannot be treated as uniform. It would also be inaccurate to characterize each individual as an invariant member of this heterogeneous group. Indeed, it is not an exaggeration to say that every utterance involves a number of decisions to produce words in one way and not another. We have already encountered a number of examples where an individual can produce the same utterance in different ways in Sections 3.1 and 3.2, noting that the issue of how these options are represented may not be straightforward. In reality, different types of intraspeaker variation probably work differently in this respect. In a recent review of phonological variation from a cognitive perspective, Bürki (2018) lays out some dimensions for classifying variable phonological phenomena, such as distinguishing between deletions, insertions, and substitutions. Here we build on that foundation by outlining some dimensions of structural classification that could prove relevant for word recognition and other elements of processing. Phonological variation is complex and pervasive, and attending to these complexities may prove fruitful in designing experiments and interpreting seemingly-incongruous results.

C3.P29 In at least some cases of intraspeaker variation, the most parsimonious analysis seems to be to allow for variation inside the lexicon itself, disrupting the basic notion of a form–meaning pair by ascribing multiple forms to a single meaning. For example, many speakers of American English can pronounce the word *economic* with

[11] We could entertain the possibility that some set of vowel phonemes (NEAR, SQUARE, START, NORTH, FORCE, CURE, LETTER) just have different phonetic properties in these dialects. However, this analysis is not as successful at capturing patterns of intervocalic /r/ (e.g., *he is* vs. *here is*).

an initial vowel like that in FLEECE or DRESS. Since those same speakers are not free to interchange those vowels otherwise, we might think that their lexicon contains two different forms for *economic*. This resembles the *tomato* example in Section 3.1 in that it is lexically idiosyncratic, but speakers actually produce both forms. On the other hand, many patterns of variation apply more generally, affecting all words with the relevant phonological properties. The same representational desiderata that motivate abstraction in the invariant phonology can be applied here. In Section 3.1, we suggested that such patterns could be accounted for through phonological rules that are stipulated to apply with some probability rather than obligatorily. Such a rule is called a **variable rule** and has long been a prominent way of thinking about phonological variation in quantitative sociolinguistics (Weinreich et al., 1968; Cedergren and Sankoff, 1974; but cf. Fasold, 1991). While the original variable rules were patterned on the phonological rule formalisms of Chomsky and Halle (1968), the modeling of phonological variation across a range of formal frameworks is a vibrant area of active research (for overviews, see Coetzee and Pater, 2011; Nagy, 2013).

C3.P30 In addition to the possibility of representing variation within the lexicon, we must contend with other non-phonological levels of structure. For example, there is some debate around whether phenomena like variable palatalization really constitute the same kind of process as their invariant counterparts, or instead come about as a consequence of how a sequence of segments is sometimes executed in the phonetics. More specifically, when producing the sequence *press you*, speakers may not always perfectly separate the alveolar articulation of [s] and the palatal articulation for [j]. Instead, these articulatory movements are produced simultaneously and the segments are **coarticulated** in a way that can acoustically resemble [ʃ]. In order to consider whether a process is phonological or phonetic, we must consider the properties of these two modules (see e.g., Pierrehumbert, 1990; Cohn, 1993 for in-depth discussions). One generally held distinction between phonological and phonetic operations concerns the properties of categoricity and gradience.[12] Phonological processes are typically held to take, as both input and output, some finite number of discrete categories that the language user stores in memory, e.g., /s/, [s], [ʃ]. The phonetics, on the other hand, control all the continuous and infinitely subdivisible dimensions of the physical instantiation of language, e.g., all possible configurations of an idealized target [s]. As mentioned earlier, this is one difference between Andruski et al.'s (1994) manipulation of VOT and the other, more categorical, variables laid out in Section 3.2.1.

C3.P31 Exploring the properties of categoricity and gradience in production, Zsiga (1995) uses electropalatography to investigate how speakers articulate underlying /ʃ/ (*fresher*), word-internal derived [ʃ] (*pressure*), and word-final derived [ʃ] (*press you*). She concludes that variable palatization does not result in precisely the same [ʃ] as in underlying /ʃ/ or obligatorily derived [ʃ]. The latter two were indistinguishable from each other, although this is not necessarily always the case for obligatory derivation

[12] Not to be mistaken for invariance versus variability.

either (e.g., Port, Mitleb, and O'Dell, 1981; Ernestus and Baayen, 2006). Similarly, Ellis and Hardcastle (2002) looked at variable place assimilation in sequences like *green card*, which could be produced with a velar nasal [ŋ] or an alveolar nasal [n] (compare obligatory word-internal assimilation, e.g., *enter*, *amber*, *prank*). They find that while some speakers variably produce a fully velar [ŋ], others retain some residual alveolar articulation. Importantly, just because the phonetic realization of, for example, palatalization in *press you* is different from that in *pressure*, it does not necessarily follow that the former is not phonological. The different types of palatalization cannot be represented with a single process anyway, since one is obligatory and one is variable.

C3.P32 As a larger point, phonological processes need not be **structure preserving**. That is, they do not have to result in sounds that are already part of the underlying inventory (Scobbie, 1995; Bermúdex-Otero, 2010). For instance, most speakers of American English will produce postvocalic /t/ and /d/ as a flap, [ɾ], before unstressed vowels (e.g., *city*, *writer*, *rider*), but /ɾ/ is not generally considered to be a sound that is available for underlying lexical representations because it only occurs in certain predictable environments (Kiparsky, 1979; Kahn, 1980). Structure non-preserving allophony is licit whether it **neutralizes** the contrast between multiple underlying segments (e.g., *latter* and *ladder* become homophones), or there is only one possible underlying representation of a surface segment, like the word-final /t/ manipulations overviewed in Section 3.2.1 (Deelman and Connine, 2001; Sumner and Samuel, 2005). A proper diagnosis of categoricity in this sense, then, does not require the recreation of phonological categories that occur elsewhere in the system. Rather, it is only necessary that the various instances of forms be distributed in discrete categories, whatever the precise nature of these categories turns out to be.

C3.P33 To further exemplify this point, consider British English /t/-glottaling and TH-fronting.[13] Most speakers of British Englishes can optionally realize /t/ as [t] or as a glottal stop, [ʔ], (Stuart-Smith, 1999; Fabricius, 2002) word-finally or before unstressed syllables in the same word (e.g., *bottle*, *butter*, *bat*). Like for flapping, /ʔ/ is not typically considered part of the underlying inventory, that is, the process is structure non-preserving. Plus, instances of British English /t/-glottaling are not equally distributed across phonetic continua like the gradual reduction of a coronal gesture or constriction of the glottis. Rather, there is variable, but categorical, selection between discrete coronal and glottal closure options (Heyward, Turk, and Geng, 2014). In contrast, TH-fronting is a variable feature of many varieties of British English (Kerswill, 2003) and African American Language (Green, 2002; Sneller, 2020), whereby underlying interdental fricatives /θ ð/ can be realized as labiodental fricatives [f v]. Since labiodental fricatives are required for the representations of other words (e.g., *free*, *vine*), TH-fronting is a structure preserving process. As with all structure preserving processes, it is also neutralizing, since the contrast between interdentals and labiodentals is lost when it applies (e.g., *three* becomes homophonous with *free*). The dimensions of structure

[13] TH here stands in for both voiced /ð/ and voiceless /θ/ interdental fricatives.

preservation and neutralization are relevant for processing because they impinge on matters of the nature and number of stored segments, and potentially the readiness with which underlying forms are retrieved.

C3.P34

	Structure preserving	Structure non-preserving
Neutralizing	TH-fronting	Flapping
Non-neutralizing	—	/t/-glottaling

C3.P35 Of course, the invariance problem means that it is often not a straightforward task to identify categories. Phonetic variation is a constant, even if it is not solely responsible for a pattern of variation. In practice, this means that phonological categories manifest as individual *distributions* and not individual *points* in phonetic space, since instances where some category is the intended target will inevitably be perturbed by phonetic variation. Moreover, phonological processes are often developed from the stabilization of more gradient phonetic variation (Bermúdez-Otero, 2007), and these physiologically motivated phonetic phenomena can remain and co-exist even after a phonological process is established from them (Bermúdez-Otero, 2013). This means that phonological variables are often accompanied by diachronically related phonetic variation that resembles and even conceals them.

C3.P36 Another factor in determining the structural properties of a potential phonological variable involves the contexts in which it occurs and, specifically, how it interacts with morphology. As mentioned in Section 3.1, words ending in unstressed *-ing* can variably be pronounced with [ɪŋ] or [ɪn]. As it turns out, the form with a coronal nasal appears much more often in verbal forms, where *-ing* is a suffix (e.g., *working~workin'*), than in nouns (e.g., *awning~awnin'*). Even without detailed evidence of categories in phonetic space, many grammatical theories do not allow the phonetics to be directly affected by morphology like this (Fodor, 1983; Bermúdez-Otero, 2010). This kind of a relationship between morphology and phonetics is prevented by the concept of **modularity**, which relegates different kinds of operations to separate parts of the grammar that are strictly ordered: by the time a word or utterance gets to the point of phonetic implementation, its internal morphological structure is no longer relevant. If we apply modular reasoning to variable processes, we have to conclude that variation in *-ing* is not phonetic. However, there is a different source of ambiguity present in these kinds of variables. Specifically, we can also account for morphologically conditioned patterns in variation by saying the morphology has some capacity for variation itself. Just as [t] and [ʔ] are possible allophones of /t/, perhaps /ɪŋ/ and /ɪn/ are possible **allomorphs** of *-ing*. Thus, it is not clear whether an instance of the word *workin'* was rendered in this form by way of phonological process or if the speaker selected an alternative form of the *-ing* suffix with a coronal nasal.

C3.P37 Ultimately, these levels of structure (phonetics, phonology, morphology) are not always neatly separable, and what look like cases of a single variant may actually have a

number of different sources. Furthermore, different phonological processes interact with the underlying inventory of sounds in different ways. Understanding these structural properties, more broadly, helps us to discern the kind of mechanics that are at play when someone produces phonological variation, and the nature of the processing tasks required to interpret the linguistic signal as it is perceived.

C3.S11

3.3.3 Non-linguistic factors in phonological variation

C3.P38

Our overview thus far may have given the impression that phonological variation merely injects uncertainty into multiple levels of structure. In reality, phonological variation is systematically conditioned by a number of factors. Weinreich et al. (1968) refer to this systematicity as **orderly heterogeneity**. There are observable linguistic differences correlated with many social dimensions such as gender identity (Eckert, 1989; Kiesling, 2004; Zimman, 2009), race/ethnicity (Fought, 1999; Hoffman and Walker, 2010; Bucholtz, 2011; King, 2016; Holliday, 2019), class or socioeconomic status (Labov, 1966; Rickford, 1986; Eckert, 1988; Labov, 1990), and sexual orientation (Gaudio, 1994; Moonwoman-Baird, 1997; Bucholtz and Hall, 2004). All of these elements of a person's identity are relevant to the language they produce and are negotiated with regard to the particular setting and audience of an utterance. The interplay between various social dimensions is particularly well exemplified in classic observations of the social stratification of phonological variation. Variants whose rate of use is correlated with a language user's socioeconomic status are typically also correlated with formality. An early demonstration of this effect can be found in Labov's (1966) investigation of the social stratification of rhoticity in New York City. Speakers increasingly used consonantal realizations of non-prevocalic /r/, the variant associated with American English speakers of a higher socioeconomic status, as they performed tasks inducing more linguistic self-monitoring and a more formal style.

C3.P39

Results like these suggest that speakers understand how phonological variation is socially stratified and can draw on this knowledge to inform their linguistic choices. As such, inter- and intraspeaker variation are not divorced from one another but closely intertwined. This basic premise is the foundation of "Third Wave Sociolinguistics" (Eckert, 2012), which focuses on individuals' capacity to dynamically construct and perform their identities by using linguistic features that have garnered particular social meanings according to the context they appear in. The fact that individuals have detailed knowledge of how phonological variation is socially organized is further demonstrated by studies that explicitly elicit listener judgments of an interlocutor (Campbell-Kibler, 2009, 2011). These show broad agreement in terms of what social information can be inferred from certain linguistic behaviors. All of this is to say that language is inherently social, and any linguistic processing must occur in tandem with processing of social information.

C3.S12

3.4 Consequences of phonological variation for models of the mental lexicon

C3.P40 Section 3.3 provided a detailed, though not comprehensive, look at the possible grammatical underpinnings and structural outcomes of different types of phonological variation. Equipped with this background, we now turn to consider some additional ways in which these structural aspects of phonological variation might be relevant to processing (that is, beyond the widely explored questions of non-canonicality that we discussed in Section 3.2). We find it useful to frame some of these questions in terms of their methodological implications, in order to highlight the inescapability of these issues even for studies that are not designed to address phonological variation. However, the issues we spell out are not *merely* methodological, since the potential differences between a model talker and a participant that we will outline could equally well be thought of as potential differences between two real-world interlocutors.

C3.S13

3.4.1 When processing meets mismatching representations

C3.P41 A practical consideration that we have only touched on briefly so far is that experimental research in spoken word recognition must choose what form of each word to use as a stimulus. The usual practice is to create stimuli using a model talker whom the researchers judge to sound "standard," producing word stimuli with the form that is taken to be "canonical" in some respect. As we discussed in Section 3.2, canonical forms may or may not have a special representational status, but even if they do, we must recognize that standard accents and canonical forms are inherently socially constructed. And even once we acknowledge that these are social constructs, there are cases where different options are equivalently standard or canonical in different varieties of the language. Choices that feel like neutral defaults to a researcher, then, simply have no guarantee of either matching what is in any given participant's mental lexicon or reflecting the bulk of their real-world listening experience. And as our discussion in Section 3.3 makes clear, there are many dimensions along which the form chosen by the researcher might not align with participants' mental representations.

C3.P42 Many of the under-explored issues we identify have to do with mismatches between language users with or without some merger: in other words, representational differences across individuals. Consider the word competitors that are entertained and eliminated over time in cohort-based word recognition models (Marslen-Wilson, 1987; see Magnuson and Crinnion, this volume, for more detail). In these models, when a listener hears an initial sound such as [k], they generate a list of possible word candidates starting

with /k/: *capture*, *kick*, *cotton*, *continent*, *caution*, *cough*, and so on. But if the second sound is [ɑ], a listener with a merged LOT-THOUGHT class (most often realized phonetically as [ɑ]) might eliminate only *capture* and *kick* from this (partial) list, while the listener with a LOT-THOUGHT distinction might additionally eliminate *caution* and *cough* (because that listener expects those words to contain /ɔ/, not /ɑ/). A related lexical property like neighborhood density,[14] which has been shown to influence word identification ease and accuracy (Vitevitch, Luce, Pisoni, and Auer, 1999; Vitevitch and Luce, 2016), might exhibit similar types of differences depending on whether it is calculated over LOT-THOUGHT merged or LOT-THOUGHT distinct lexicons.[15] Interestingly, this suggests that dialects differing along phonemic inventory lines such as the LOT-THOUGHT merger might offer a useful opportunity to study homophone representation (Swinney, 1979; Caramazza, Costa, Miozzo, and Bi, 2001) because they may allow a minimal comparison between speakers for whom some word pairs are and are not homophones. Another merger-related question is whether nonwords used as experimental stimuli might sometimes have an unrecognized real word status to listeners with different lexical or phonological mental representations. A researcher might construct an intended nonword *frind* without taking into account that to a listener who has a *pin/pen*-merger (where DRESS and KIT are merged before nasal consonants) this would be a perfectly good instance of the real word *friend*. The real-world flipside of this question is whether a pronunciation like [frɪnd] from a Southern-accented talker might be processed as a nonword by a non-Southern listener.

C3.P43 Is it plausible that such representational mismatches intervene in processing in this way, given the literature we already surveyed in Section 3.2? This premise is supported by the observation that misunderstandings arising from phonemic inventory differences are not uncommon in everyday conversational interaction (Labov, 1994 p. 324–327), even in seemingly disambiguating communicative contexts. It is also supported by a number of **cross-dialectal comprehension** studies targeting specific sources of representational mismatch (Labov, Karen, and Miller, 1991; Flanigan and Norris, 2000; Labov, 2010). In other words, although our Section 3.2 discussion emphasized evidence that listeners are eventually able to overcome the challenges of phonological variation, we should not conclude that those challenges do not arise at all. And methodologically speaking, we cannot safely assume that these issues are minor in scope or easily avoidable. For example, the presence vs. absence of the LOT-THOUGHT merger divides the geographic territory of US English approximately in half across a number of non-contiguous dialect regions (Labov et al., 2006, p. 59).

C3.P44 Beyond these interspeaker differences, we highlighted issues of structure preservation and neutralization in Section 3.3 because intraspeaker variation can give rise to parallel issues when it is structure preserving or neutralizing. For example, the variable deletion of word-final /t/ and /d/ in consonant clusters, which happens in every English dialect we are aware of, can generate homophony (*past/pass, mold/mole*) and erase morphological

[14] Neighborhood density measures capture how many other words in the lexicon have a similar phonological shape; see Magnusson and Crinnion, this volume.

[15] Exactly how these differences play out will depend on the calculation method and the treatment of homophones within that method.

information (*jumped/jump*) but does not always do so (*act, spent*). These inconsistent lexical consequences pose challenges for listeners and researchers alike. Intraspeaker variation that is structure non-preserving, on the other hand, introduces variants that inherently signal their derived status by virtue of not existing in the phonemic inventory, a signal that could in principle trigger shifts in how lexical access processes proceed. Nonetheless, structure non-preserving intraspeaker variation can additionally generate homophony when it is neutralizing (*latter/ladder*), reintroducing some of the issues around cohort competitors and neighborhood density that we discussed for more basic merger examples. Empirical questions about how any particular variable is represented, such as whether it involves incompletely neutralized phonetic variants or an alternation between discrete phonemic units, take on new importance in light of the possibility that they may give rise to different processing consequences.

C3.S14 ### 3.4.2 When social information comes into play

C3.P45 Work in the emerging area of **sociolinguistic cognition** (Campbell-Kibler, 2010; Loudermilk, 2013; Chevrot, Drager, and Foulkes, 2018) suggests that listeners' navigation of inter- and intraspeaker phonological variation during lexical access is guided by their experience with the social influences on variation such as we discussed in Section 3.3. There is robust evidence that listeners can use social information to make inferences about a speaker's likely linguistic system.[16] They can then take into account whatever knowledge they have of that system, instead of relying on their own system, when identifying sounds and words produced by that speaker (Niedzielski, 1999; Strand, 1999; Hay and Drager, 2010; D'Onofrio, 2015; Hay, Walker, Sanchez, and Thompson, 2019). This kind of reverse-engineering appears to be possible based on the presence of linguistic features that tend to co-occur (such as different features of a Southern accent) without extra social information (Dahan, Drucker, and Scarborough, 2008), although it remains an open question whether these effects are mediated by social inference (Wade, 2020). For example, in the *frind/friend* example above, even a Southern-accented participant might be able to use the model talker's non-Southern accent to infer that the model talker did not intend the word *friend.* Of course, the fact that this listener might ultimately reach the correct conclusion about the intended nonword status of the stimulus does not rule out the possibility that they first retrieved *friend* and then backtracked, or were otherwise delayed in ways that a non-Southern listener might not have been. The time course of reasoning about an interlocutor's differing linguistic system, and how it interacts with possibly more basic lexical access mechanisms, is certainly not well established. Furthermore, while there is evidence that listeners can also use linguistic and social information to help guess the intended word when intraspeaker variation is in play (Mitterer and McQueen, 2009; Casasanto, 2008), inherent variability means that incorporating such information can favor, but not guarantee, the correct outcome.

[16] It should be noted that listeners' ability to use such information is modulated by listeners' social attitudes (Kang and Rubin, 2009).

C3.P46 The fact that phonological variation encodes social information in speech, then, makes it an exciting frontier for understanding the mental lexicon. On the comprehension side, phonological variation is not simply noise that listeners must factor out. Rather, it is a rich source of structured information about speakers themselves and their likely behavior. Recent approaches to incorporating this kind of social information into processing models include the dual-route approach to socially weighted encoding proposed by Sumner et al. (2014) and the ideal adapter model of Kleinschmidt and Jaeger (2015; see also Kleinschmidt, 2019). On the production side, producing phonological variation demands that speakers dynamically shape their own speech to be fluid, connected, and socially and contextually appropriate; Babel and Munson (2014) give a useful overview on production issues related to variation.

C3.P47 However, this same social sensitivity is a reason that experimental work on variation needs to proceed with caution. It is easy to lose sight of the fact that a laboratory on a college campus is itself a social setting, one that for most people is far-removed from everyday life. Research that attempts to investigate socially meaningful phonological variation in the lab, but does not take into account the social properties of the experimental context itself, runs the dual risk of not only drawing scientifically unwarranted conclusions but also propping up the marginalization of "nonstandard" varieties.

C3.S15 3.4.3 Toward new advances in modeling the mental lexicon

C3.P48 The issue of phonological variation arises throughout this volume. Magnuson and Crinnion (this volume) point to the many sources of talker variation and a wide range of phonological processes producing deviations from canonical form as major challenges for current models of spoken word recognition. Creel (this volume) highlights the difficult challenge that pervasive variability poses to word learners. Kilbourn-Ceron and Goldrick (this volume) end their survey of word production by noting that we know very little about how words are produced in sentential contexts (as, indeed, words nearly always are). It appears that phonological variation poses one of the major obstacles preventing word recognition and production models from being able to cope not only with diverse talkers and social contexts but also with connected speech at all. Improving our understanding of phonological variation and its relationship to the mental lexicon thus promises to facilitate the modeling of word recognition in connected speech input and word production in context. The problems at hand are far from simple, but we believe that turning toward a view of lexical access as an inherently socially situated process offers the promise of bringing models of the mental lexicon and its use into a more detailed alignment with what human language users know and do.