Doing phonology in the age of big data*

Abigail C. Cohn[1] and Margaret E. L. Renwick[2]
[1]*Cornell University*, [2]*University of Georgia*

Abstract. The central question we address in this paper is how to do phonology in an emerging era of big data. The more specific question we explore is how to better use naturalistic corpus data to study phonology. We support the growing trends that are expanding the range of phenomena phonologists investigate, and enhancing the richness of detail with which investigations are conducted. Presenting case studies from English, Indonesian, and Romanian, we argue that the use of corpus data necessarily follows from the goals of the generative enterprise. At the same time, experimental and laboratory investigations are crucial to fully and systematically explore both phonological patterns and individual speaker differences, as we show with case studies of English, Italian, and Catalan. We advocate for an iterative model of phonological analysis integrating careful data elicitation with both corpus analysis and experimental methods.

1 Introduction

In this era of big data, the question we address in this paper is how to incorporate naturalistic corpus data as part of our phonological toolbox. How are the central questions phonology addresses reshaped as we enrich our methods and have access to orders of magnitude more data than we have traditionally analyzed? We focus on the important empirical contribution to phonology of greatly increased access to naturalistic data, often made available through corpora. These advances have highlighted the multidimensional nature of variation in spoken language. We note at the outset that some naturalistic corpora really fall within the sphere of big data,[1] while others do not. As discussed below, incorporation of any naturalistic data enriches our work.

While work in theoretical phonology is aimed at deducing a speaker's mental grammar, the granularity of data investigated is sometimes insufficiently detailed to consider factors that experimental work has shown are systematically manipulated on a language-specific basis by speakers and listeners. Large datasets can enhance the richness of detail available for analysis, to offer greater insight into the nature of phonological competence. We argue that the use of naturalistic corpus data necessarily follows from the goals of the generative enterprise. On the other hand, we acknowledge that corpus studies alone cannot establish relationships of causality among these phenomena, nor can they fully shed light on the individual's speech patterns, which may be more systematic than those of the community as a whole. Thus analysis of naturalistic speech is needed alongside systematically-elicited data and careful experimental investigations.

---

* Corresponding author email: acc4@cornell.edu. This paper is the result of an ongoing discussion we've had about how to "do" phonology, drawing on a number of collaborative research projects. We are well aware that different parts of the paper will appeal to different readers. We would like to thank our collaborators, and our colleagues for their feedback and comments on earlier drafts of the paper. We are especially grateful to Bob Ladd, Jim Scobbie, Sam Tilsen, Natasha Warner, and Draga Zec for discussion of and comment on this paper.

[1] We loosely define *big data* as a dataset that is too large and complex to be analyzed by hand; instead such datasets require automated or computational tools like visualization, summary, statistical modeling for analysis. While many very large datasets are naturalistic corpora, they can also be psycholinguistic in nature (ten Bosch et al. 2015), or compilations of laboratory speech or other data.

We realize that for many researchers these points are obvious and want to acknowledge the significant body of work embracing this view. Here, we highlight how including naturalistic data can deepen and enhance our understanding of phonological systems and explore how our methods of gathering and analyzing data may impinge on what we learn from these data.

Much work in phonology as well as linguistic phonetics has centered on highly systematic phenomena, in which regular patterns (the more exceptionless the better) can be identified and analyzed. This work traditionally places linguistic variability of all forms into the bins of either optionality or free variation. Additionally, phonological analysis and phonetic experimentation have focused on phenomena that are most obviously phonologically and grammatically conditioned, i.e. triggered by factors like phonotactics, prosody, morphological structure, and other grammatical aspects. These practices are still implicitly accepted despite the fact that many other types of factors are known to affect the speech stream, including pragmatics, discourse and communicative factors, speaker affect, and socio-indexical characteristics, to name a few. These factors, which have been deemed external to phonological analysis, nevertheless form part of the native speaker-hearer's linguistic competence and are realized simultaneously within the speech stream. (See Labov 2006; Foulkes & Docherty 2006; and Foulkes 2010 for discussion of this point.)

Data used for phonological analysis is often at a fairly coarse-grained level of detail, while phonetics gives us access to much finer detail; the mismatch between these two raises an issue of the granularity of data used as the basis for analysis. Also, elicited data and laboratory speech data have traditionally relied on citation forms, rather than naturalistic data, meaning that researchers miss the opportunity to incorporate into their analyses of sound patterns the many additional (including non-phonological) factors that condition speech. Further, it is often assumed that a degree of systematicity is needed for a phenomenon to be considered for analysis. Phonologists and phoneticians alike have traditionally relied on what is systematic, under-attending to many sources of variability. We argue that in order to understand systematicity, variability needs to be fully embraced, a goal that is rendered possible by the availability of sufficient naturalistic data. An important body of recent work has modeled variable phenomena, both within the phonological literature (e.g. Anttila 2002; and for recent reviews see Anttila 2012; Anttila 2018; Coetzee & Pater 2011; Coetzee 2012) and the sociolinguistic and sociophonetic literature (e.g. Guy 1997; Guy 2007; Foulkes & Docherty 2006; Foulkes 2010; Docherty & Mendoza-Denton 2012). However, our goal is not to evaluate the relative merits of these different approaches, nor to attempt to provide a model integrating different sources of variation, but rather to think about the necessary sources of input to any model – that is, the quality of the data being modeled (Guy 2014).

In this paper, we do not draw an a priori line between factors that are "relevant" to phonology and those that fall outside of the phonologist's scope. Nor do we limit analysis to speech patterns that can be deemed highly systematic and captured elegantly by rule or constraint interaction. We argue that ultimately the task of phonology is to understand the relative contributions of all factors that condition the production and perception of the speech stream. This follows from the common goal of linguistics to fully explain and model speaker/hearer competence, which necessarily includes all of these sources of variation. The simple point we develop in this paper is that this capacity can be better understood through iterative exploration, description, experimentation, and modeling of speech data.

In the remainder of this section, we offer a working definition of doing phonology, and then discuss the relationship between gradience and variability and the issue of granularity of detail with regard to speaker-hearer competence. In Section 2 we present a series of case studies showing how naturalistic data can enrich our understanding of phonological patterns. In Section 3 we illustrate phenomena related to inter- and intraspeaker variation via corpus analysis and more controlled, laboratory case studies. Building on the discussion in Sections 2 & 3, in Section 4, we advocate a methodological framework interleaving hypothesis testing with a rich array of empirical data analysis and modeling options to achieve a broader and deeper understanding of phonology and highlight the necessary next steps towards an integrated model.

1.1 Doing Phonology

There are several facets of phonology that together constitute the phonological description of a particular language, including an inventory of contrastive elements, phonotactics, allophony, and morphophonemics. We define these as follows:

(1) Facets of phonology

- Sound inventory – a set of contrastive sounds whose oppositions are used to create meaningful distinctions in a language
- Phonotactics – allowable distributions of sounds, sound sequences and prosodic structures
- Allophony – a grouping of non-contrastive sound system units into abstract higher-order units of the inventory, based predictably on contextual phonological conditioning
- Morphophonemics – alternating shapes of a morpheme which are conditioned by phonological context

Following different approaches to phonology, these facets might be treated in a unified fashion (as is the case in Generative Phonology; see Halle (1959) for an argument in this regard). Or they might be treated as separate levels or components, as was the case for American Structural Linguistics, which crucially separated the phonemic level from the morphophonemic level (see Anderson 1985 for discussion). We take as a starting point the integrated view espoused by Generative Phonology, while keeping as an open question whether these facets fully function the same way. Here we focus on *doing* phonology, following the tenets of Generative Phonology (as formalized in Chomsky & Halle 1968, *The Sound Pattern of English*), as in (2):

(2) Doing phonology

- Data are gathered via impressionistic analysis or from transcribed data, revealing or confirming intuitions of phonological patterns that are modeled via a formal phonological framework
- Observed patterns in surface forms are considered to be the confluence of hypothesized abstract underlying forms, interpreted via a system of ordered rules or ranked constraints
- The goal is to model speakers' competence via a unified account of sound inventories, phonotactics, allophony, and morphophonemics
- In this modular approach, phonology is distinct from phonetics and the lexicon

We understand Generative Phonology in its broadest sense to be the foundation of most current formal approaches to phonology, including Optimality Theory. Thus, Generative Phonology provides the broad underpinnings of our discussion of doing phonology. A fuller discussion of different approaches to phonology or the history of phonology is well beyond the scope of this paper. Here we offer a brief excursus, in the Appendix, delving into the question of "what is phonology"; and we refer the reader to Hannahs & Bosch (2018) and Goldsmith and Laks (to appear) to appear for further discussion.

Here we turn to the question of how the expansion of our analytic toolbox, including a rich array of experimental methodologies and especially naturalistic data, informs how we do phonology. In what ways have these methods shifted the questions we ask about the nature of phonology and led to new insights into phonological systems? We consider briefly motivation for and results from experimental approaches to phonology, methods loosely covered under the umbrella of Laboratory Phonology, in Section 1.2 and the implications of these data for defining speaker-hearer competence in Section 1.3, returning to the question of naturalistic data in Section 2.

1.2 Gradience vs. variability

Phonology has traditionally depended on transcribed forms, which immediately separates the phonetic reality from the level of representation being modeled.[2] Starting relatively early on in the generative enterprise, scholars worked to clarify the physical instantiation of phonological patterns, framed as understanding the relationship between phonetics and phonology (e.g. Pierrehumbert 1980). Much current research takes the relationship between phonetics and the lexicon as a lens for examining phonology, to empirically test how much phonetic detail is stored in the lexicon vs. grammatically derived. Full consideration of the relationship between phonology and phonetics is well beyond the scope of the current discussion. In brief, we take the view that there are parallels and overlaps between these areas but neither is properly reduced to or contained in the other (see Cohn 2006; Cohn 2007; Scobbie 2007; Ladd 2011; Ladd 2014; and Cohn & Huffman 2014 for a review of this literature). We agree with Pierrehumbert (2002: 103) that 'categorical aspects of phonological competence are embedded in less categorical aspects, rather than modularized in a conventional fashion' and with Scobbie (2007: 27):

> Rather, phonology and phonetics would have a transition zone, like a tidal shore ecosystem, which is defined by its dynamic transitions between seabed and land surface. Sea and land are (like cognitive and physical domains) categorically distinct, but the tides create a habitat in its own right. [. . .] Overlap does not imply loss of identity: the land and the sea are not the same and neither are phonetics and phonology.

To understand what this "transition zone" could contain in terms related to speech communication, we offer the distinction between *gradience* and *variability*. We use *gradience* in reference to the physical instantiation of a particular variable both spatially and temporally (see Cohn 2006 for discussion of several different uses of the term gradience in phonology). For example, Cohn (1993) shows that vowel nasalization in English is *gradient* in the sense that the vowel is oral or only slightly nasalized at the beginning and fully nasalized at the end, different from contrastive nasal

[2] In a framework like that outlined in (2), the imposition of labels and categories onto data reduces phonologists' attention to phonetic detail and variation, although recent approaches such as Exemplar Theory (e.g. Bybee 2000; Johnson 1997; Pierrehumbert 2001) explicitly include this finer detail. In a structural approach, however, phonetic detail is only incorporated if it rises to the level of audible salience that motivates a change in representation (e.g. narrow transcription, represented as a change of at least one feature value).

vowels in French which are fully nasalized throughout all or most of their duration. We exemplify gradience in this paper via analyses of variation in Voice Onset Time (VOT) and vowel qualities. Notably, these differences are not captured through careful transcription alone.

On the other hand, we understand *variability* (often modeled as "optionality" in phonological work) to be what Ladd (2014: 88) terms 'variation in frequency of occurrence of categorically distinct events'. For example, there is extensive discussion in the phonological and sociolinguistic literature on the *variable* occurrence of t/d deletion in English. When researchers analyze the patterns of realization they are usually studying whether [t] or [d] is present or absent in a particular phonological environment or socio-indexical context. Drawing this distinction highlights the need for a better understanding of the relationship between *gradience* and *variability*. Is it the case that t/d deletion is gradient, in the sense that t/d shortens in the same environments from which it is sometimes (variably) deleted? While initial analyses of t/d deletion relied on impressionistic judgments (Wolfram 1969), further investigations have shown that although readers may fail to *hear* a [t] or [d], it may still be detectable at the articulatory level (Browman & Goldstein 1990; Thomas 2002) and has been argued to be a Connected Speech Process (Temple 2014).

While Ladd (2014: 88) states that these two dimensions of variation are not clearly distinct from one another, Cohn (2006: 31–32) states 'It is not necessarily the case that temporal/spatial gradience and variability go hand in hand.' It is an empirical question, yet to be fully tested with data sufficiently detailed along both these dimensions, whether and to what degree they correlate with one another or whether they are orthogonal. A question still to be addressed is, are our data sources and methods sufficient to understand the nature of variation in both temporal/spatial dimensions and in terms of frequency of occurrence?

The extensive body of literature on the relationship between phonology and phonetics has highlighted the fact that introspection and impressionistic transcription, while useful starting points, do not offer a fine enough lens to capture data with sufficient detail to *see* gradience, much less distinguish it from variability (e.g. Ladd 2011). The critical need for instrumental and experimental data is one of the central tenets of Laboratory Phonology (see Solé, Beddor & Ohala 2007; and Cohn, Fougeron & Huffman 2018 for recent discussion). Likewise as amply attested in the sociophonetic literature introspection and data elicitation are similarly insufficient to study variability (e.g. Stuart-Smith, Lawson & Scobbie 2014). Both experimental data (articulatory, acoustic, psychoacoustic, neuroimaging data, etc.) and naturalistic data (interviews, corpora, etc.) are needed to advance our understanding of speech. Without these richer data sources, the limitations of our data circumscribe and may predetermine our understanding and models of phonology. With richer and messier data, and more complex conceptualizations of the systems which generate the data, we can ask about the relationship between the more categorical aspects of sound patterns and their more gradient aspects, and about the contrastive aspects of lexical representation and the finer-grained details in long-term memory. Only then can we fully address to what extent "phonetics" and "phonology" are distinct domains, and how the interaction of gradience and variability bear on their relationship.

We think that entirely discrete domains are unlikely, a perspective supported by Anderson's (1981) discussion of what the domain of phonology proper is, which identifies *language* as the intersection of general principles of learning, perceptual psychology, physiology, acoustics, social factors, and more. It is also unlikely that a sharp line can be reliably drawn between the more categorical and more gradient aspects of sound patterning. Some theories give a theory-internal answer (e.g. *SPE*), but it is clearly inadequate to strictly equate phonology with categorical phenomena and phonetics with gradience; for instance, the fact that some phonotactic patterns are more well-formed than others (e.g. Hay, Pierrehumbert & Beckman 2003) shows that phonology is not always categorical. Instead, as schematized by Cohn (2006; 2007), we take the position that while fine-grained phenomena are *generally* associated with phonetics and more coarse-grained phenomena are typically within the domain of phonology and greater abstraction, there is significant overlap between the two. In any investigation of speech patterns, it is necessary to understand the distribution of the relevant data. This includes distributional variation that may be due to individual speakers, or that is apparently stochastic in nature (see Section 3).

1.3 Granularity and speaker-hearer competence

The issue of granularity is also related to lexical representations, which may not be as abstract as proposed within the *SPE* model. Traditionally generative phonology takes the position that only idiosyncratic (unpredictable) information is stored in the lexicon, while predictable information is derived by rule or in later approaches constraint interaction. However, ample psycholinguistic evidence from lexical processing shows that speaker-hearers have access to fine details of pronunciation, accents, voices, etc., leading to the view espoused by Exemplar Theory that lexical representations are built up of clouds of specific utterances that include and are affected by highly specific phonetic properties (e.g. Johnson 1997; Pierrehumbert 2001; Pierrehumbert 2002; Johnson 2007). What is the relationship between the more abstract and fine-grained details of linguistic information as stored by speaker-hearers? We reject the false dichotomy of phonological representations as either abstract or episodic. The growing consensus in the literature is that both abstract and fine-grained details are part of lexical knowledge, and that accounting for this necessarily demands integrated models (Pisoni & Levi 2007; Ernestus 2014; Fink & Goldrick 2015; Pierrehumbert 2016).

In parallel to advances that support greater integration of phonetic detail into the lexicon, recent developments in sociophonetics have used naturalistic and finer-grained data to study the relationship between phonology and socioindexical aspects of language (see Labov 2006; Foulkes & Docherty 2006; Pierrehumbert 2006; Foulkes 2010 for discussion). This body of work highlights the dimension of sociolinguistic competence (e.g. Coleman 2002; Labov 2006; see Ladd 2014 for recent discussion). Taking Chomsky & Halle's (1968) definition of the generative enterprise as the speaker-hearer's 'knowledge of the grammar that determines an intrinsic connection of sound and meaning for each sentence', we believe, actually leads to a rejection of the view put forward by Chomsky (1965: 3) that 'Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogenous speech community.' Rather, we understand phonology as the interplay within a system of communication used to convey multiple dimensions of meaning (linguistic, paralinguistic and socioindexical), and each of these dimensions involves competence and usage.

As models relating phonetics to both lexical and social factors start to converge, it becomes increasingly clear that adequate phonological models cannot be built without sufficiently fine-grained and rich data – ideally data that are both *messy* and *big*. In the following sections, we further motivate the incorporation of different types of data and methods on which such modeling could be founded, providing a sketch of how to do phonology in such a multifaceted system.[3]

[3] In understanding what phonology is there is also a sociological dimension about the labeling of fields and subfields, including the relationship between phonology, laboratory phonology, and (linguistic) phonetics. We follow Cohn et al. (2018)'s position that laboratory phonology is an approach to doing phonology, not a theory, and consider

We turn now to the question of how we know which aspects of variation are "phonologically" relevant.

2 Aspects of variation and variability that are "relevant to phonology"

Language is a system of patterned variation, as richly documented in the variationist sociolinguistic tradition (Labov 1994; Labov 2001; Labov 2010; Tagliamonte 2012). Key to understanding language, not only as a cognitive system but also as a communicative system, is understanding and modeling these sources of variation (see Guy 2011 for discussion). In the case of variation in sound patterns we can ask, what is the relationship between "phonologically relevant" variation and "other variation"? Scobbie (2007: 18) highlights the complexity that the fine grained details of production and perception bring to the task of understanding what is phonologically relevant:

> Phonetically detailed studies of multiple speakers reveal the extent of language-specific control of phonetic targets (often resulting in subtle interspeaker variation) in phenomena that are firmly within the phonological canon. Such work shows the extent to which subtle, gradient, and variable (i.e. phonetic) patterns exist alongside the gross and categorical (i.e. phonological) ones previously easily detected via native speaker intuition and impressionistic transcription of individuals or small homogeneous groups of speakers.

Traditionally, using introspection and impressionistic transcription, there was a general sense that details apparent to the trained ear were relevant to the phonology and those aspects of variation not attributable to the factors thought to play a role in our particular phonological theory fell outside the phonology. From this point of view, variation not conditioned by phonological factors, "free variation", fell outside our theory. Increasingly as we broaden our understanding of linguistic competence, it has been acknowledged that the idea of "free variation" is a dodge:

> Arguably, though, the only thing that is free about free variation is that it frees the linguist up to dust their hands and say 'OK, we've analysed that!' Sociolinguists' studies of language in use have shown that variation is always more or less constrained by some factor relevant to the context in which a speaker is using their language […] Sociolinguists have shown that a lot of what appears to be free variation can be accounted for if linguists take social factors into account as well as linguistic factors (Meyerhoff 2011: 12).

Furthermore, it was often assumed that those factors understood to be part of the phonology were systematic in their realization and those patterns outside might well be variable. The problem is that if we make *a priori* assumptions about which aspects of variation are phonologically relevant, *and* we collect data that aren't sufficiently rich, we cannot determine which aspects of variation and variability are "relevant to the phonology".

If our goal is to understand speaker-hearer competence in its fullest sense, then the data we study need to incorporate this richness, both in terms of granularity and in terms of the socio-indexical and communicative properties they express. Fortunately, with recent technological,

---

experimental work of all kinds, including sociophonetics, as part of laboratory phonology; for many researchers the incorporation of corpus work and more naturalistic data is also part of laboratory phonology.

computational and methodological advances such data are becoming increasingly available and useable.

The ultimate testing ground for models of phonology is *naturalistic data*, which typically means relying on speech corpora. Through corpus analysis, it is possible to test whether effects that arise in experimental or intuition-based studies are widespread and meaningful. Since most models of phonology are built on intuition and experimentation, an extension to corpus analysis means that these models are being tested on speech collected "in the wild." Recent research, such as the case studies we describe in this paper, shows that in addition to confirmatory results, corpus analysis provides a window on phonology that experimental and intuition-based study alone cannot. However, it is important to note that corpora cannot *substitute* for experimental studies, as their contents are inherently unbalanced due to frequency distributions in natural language, and they cannot be controlled as tightly as in an experimental scenario. We present three recent cases that highlight the contributions of corpus analysis to phonology.

The first shows the benefit of a naturalistic corpus for an analysis of Romanian, in which naturalistic data show *different* trends than laboratory data. This possibility highlights the fact that controlled data offer only a partial picture of linguistic competence, while a fuller picture emerges through the investigation of naturalistic data. Second, a corpus study of Jakarta Indonesian highlights the individual speaker's role in producing surface forms for a morphophonological pattern; and third, a study of English nasal place assimilation shows that a sufficiently large naturalistic corpus allows the discovery of patterns that are regular, but rare. The latter two studies also emphasize the importance of speaker metadata.

2.1 Romanian central vowels

An illustration of the benefits of corpus data with respect to data collected in a laboratory comes from recent studies of Romanian, to answer the phonologically-motivated question: how can we be certain that a pair of sounds have a phonemic distinction? The Romanian language includes two central vowels /ɨ ʌ/, which are historical allophones now separated by a handful of minimal pairs. In a series of comparisons of the two vowels, investigating their historical evolution, synchronic distribution, and frequency profiles, Renwick (2012; 2014) argued that they are *marginally contrastive*. Marginally contrastive sounds are considered structurally separate phonemes due to the presence of at least one minimal pair, but other linguistic facts suggest the sounds are conditioned. For instance, Romanian /ɨ/ is typically stressed, pre-nasal, and never appears in post-tonic position, while /ʌ/ is typically unstressed and often word-final. (See Hall 2013; Scobbie & Stuart-Smith 2008 for other recent discussions of marginal or quasi-phonemic contrast.) As a consequence of this nearly complementary distribution, specifically because a Romanian central vowel's quality can usually be inferred by phonological or morphological context, Renwick predicted that the two sounds could be subject to phonetic merger, and that listeners might not reliably distinguish them. Such a finding would call into question the marginal evidence for contrastiveness found in the lexicon, leading to a less categorical division between sounds which are contrastive and those which are in an allophonic relationship. While a perception study provided weak evidence for the latter prediction, a laboratory investigation of the Romanian vowel space showed that /ɨ, ʌ/ were highly distinct, with no more acoustic overlap than other adjacent pairs of vowels in the system (Renwick 2012; Renwick 2014). This constituted a null result with respect to the predictions of marginal contrast.

To further test whether the Romanian contrast has the same phonetic properties as a more robust phonemic contrast (one with more minimal pairs and less phonological conditioning), Renwick and colleagues turned to a more naturalistic source of data (Vasilescu et al. 2016; Renwick et al. 2016). They analyzed the vowel space in a large phonetic corpus of Romanian broadcast speech, including both read speech and debate-style speech. In this sample of continuous speech gathered from news media, the predicted phonetic effect on /ɨ ʌ/ does occur: the high vowel /ɨ/ is lower in the vowel space, overlapping with mid /ʌ/ (see Fig. 1). The authors show that according to measures of vowel overlap, all adjacent pairs of vowels are non-overlapping in laboratory speech, but that in broadcast speech /ɨ ʌ/ overlap strongly in F1 (and to a lesser extent, so do /u o/, representing an overall reduction in vowel space size). Using the same corpus the authors show that the /ɨ ʌ/ contrast has the lowest functional load of all Romanian vowel pairs, and that in an automatic speech recognition system, word error rate would not increase significantly if the central vowels were merged; that is, the distinction does very little "work" either lexically or acoustically.

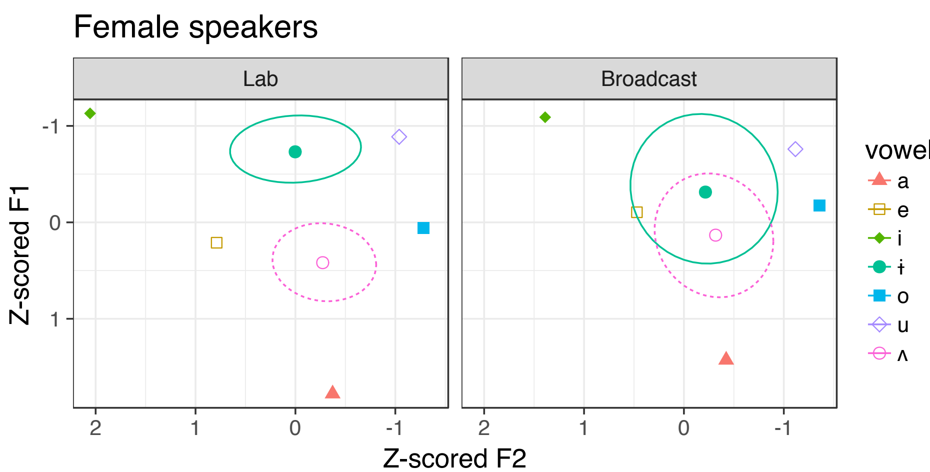

Figure 1. Comparison of Romanian vowel spaces in laboratory speech (left) and broadcast speech (right), for female speakers, with normalized F1, F2 values

Renwick and colleagues argue that these findings provide strong evidence for the marginal status of the /ɨ ʌ/ contrast. The theoretical implication is that although there is lexical evidence for a phonological distinction, it is scarce; the nearly-complementary distributions of the sounds may allow speakers to nearly merge the pair in production. Methodologically this case highlights the difference between results gained from laboratory studies and those from corpora. Specifically, corpus results illuminate patterns that follow from distributional facts about the language, but which are not elicited in the formal environment of the laboratory. Without naturalistic data, we would not see the actual, marginal realization of this contrast; on the other hand, based solely on corpus results we might infer that the contrast was being neutralized, while in fact the laboratory data confirm that native speakers are capable of retaining distinct central vowels, consistent with their structural distinction.

A drawback of this Romanian broadcast corpus is that, due to its collection methods, it lacks speaker metadata: it is possible to distinguish speakers by sex, style (reading vs. spontaneous debate) and TV program, but few other characteristics are retained. Increased metadata and broader sampling would allow us to differentiate, for instance, groups of individual speakers who either maintain a distinction between /ɨ ʌ/ or merge them, and in what social or communicative contexts. We turn next to a dataset that does permit incorporation of these factors.

2.2 Morphophonemic variability in Jakarta Indonesian

The variable realization of a morphophonemic pattern in Jakarta Indonesian provides another example of what we can learn from naturalistic data that we can't learn from careful introspection or elicited data alone. A phonological analysis of Jakarta Indonesian reveals that the active prefix (cognate with *məN-* in Standard Indonesian) shows variation in its realization before voiced stops, surfacing either as a homorganic nasal or as [ŋə], e.g. /bəli/ 'to buy', [mbəli] ~ [ŋəbəli] 'buy, active'. From a phonological perspective, the goal is to uncover the underlying phonological form and fully characterize the alternation in an appropriate formalism; in order to describe the process, however, first we must learn what the source and structure of the variation are. This case study shows that factors including the oral stop's place of articulation and the speaker's linguistic background are important predictors of the surface form, but that variation still exists at the individual level.

Most prior work on Indonesian phonology focuses on Standard Indonesian. But as a variety taught in school and prescribed by the National Language Board, it is normative and formal (and has only existed as such since 1945). It does not reflect the language acquired and spoken natively at home, and thus is not a good object of study for phonology. Rather we should turn to spoken varieties of the language, including the many vernacular varieties of Malay spoken across the Indonesian archipelago, but also "colloquial" Indonesian (see Kurniawan 2018 for discussion). Linguistic work on this topic has been sparse (though see Ewing 2005), but recently attention has turned to Jakarta Indonesian, a rapidly developing variety of colloquial Indonesian, spoken by an increasingly large group of native speakers in and around Indonesia's capital. With millions of native speakers, the need for documentation of this emerging variety is acute. Furthermore, there is evidence that Jakarta Indonesian is emerging as a generalized urban variety in other urban areas in Indonesia. Another issue is the relationship between Jakarta Indonesian and Betawi (Malay), a vernacular variety of Malay historically spoken in and around Jakarta. Both Jakarta Indonesian and Betawi form a dialect continuum with Standard Indonesian.

To provide an empirical foundation for the study of Jakarta Indonesian, the Betawi-Jakarta Corpus (BJC) of the Jakarta Field Station (Gil & Tadmor 2015), a naturalistic corpus of speech in Jakarta, was created. This grammatically annotated and phonetically transcribed corpus consists of conversations and interviews with 65 adult native speakers of both Jakarta Indonesian and Betawi. Its 28 hours of audio contain a total of 75,079 spoken utterances, searchable by orthography, phonetic transcription, morphological structure, etc.; and it includes speaker metadata.

Kurniawan (2015) used the BJC to investigate both the patterning and source(s) of variation noted above. As illustrated in (3), the prefix surfaces as [ŋ-] (3a) before root-initial vowels, as [ŋə-] (3b) before liquids and glides, and as Ø before nasals (3c). We observe a pattern of "nasal substitution" with voiceless stop initial roots, whereby the root initial consonant is replaced by a homorganic nasal (3d).

(3) Patterning of *N-* prefix in Jakarta Indonesian (the suffix +*in* has no influence)

a. Root-initial vowels: [ŋ-]

Root	*Active*	
aŋkat	ŋaŋkat	‘to lift’
obat+in	ŋobatin	‘to medicate’
ekor	ŋekɔr	‘to follow’
isi	ŋisi	‘to fill’

b. Root-initial liquids and glides: [ŋə-]

Root	*Active*	
lamar	ŋəlamar	‘to propose’
rusak	ŋərusak	‘to destroy’
jakin+in	ŋəjakinin	‘to believe’
wabah	ŋəwabah	‘to be epidemic’
harus+in	ŋəharusin	‘to require’

c. Root-initial nasals: Ø

Root	*Active*	
makan	makan ‘to eat’	
nilai	nilai	‘to grade’
ɲaɲi	ɲaɲi	‘to sing’
ŋaŋgur	ŋaŋgur	‘to do nothing’

d. Root-initial voiceless consonants: [N-] “nasal substitution”

Root	*Active*	
pilih	milih	‘to choose, to vote’
tulis	nulis	‘to write’
kasih	ŋasih	‘to give’
sapu	ɲapu	‘to sweep’
t͡ɕari	ɲari	‘to seek’

This suggests an underlying form of /ŋ-/, or /N-/, corresponding to Standard Indonesian /məŋ-/ or /məN-/. However, the analysis becomes more complicated when we observe the variable realization of the prefix before root-initial voiced obstruents where both nasal assimilation (4i) and the form [ŋə-] (4ii) are observed.

(4) Active forms with root-initial voiced obstruents: both assimilation and [ŋə-]

bəli	i. mbəli	‘to buy’
	ii. ŋəbəli	
dapət	i. ndapət	‘to get’
	ii. ŋədapət	
d͡ʑawab	i. ɲd͡ʑawap	‘to answer’
	ii. ŋəd͡ʑawap	
guntiŋ	i. ŋguntiŋ	‘to cut with scissors’
	ii. ŋəguntiŋ	

What is the source of this variation, and what are its implications for the phonological analysis of this observed pattern? Is the variation conditioned by phonological properties, such as place of articulation, or lexical properties, either token frequency or specific lexical forms? Or is it due to a stylistic difference or a dialect difference (Jakarta Indonesian vs. Betawi?), or speaker specific differences? Once these factors are taken into consideration, is the pattern systematic or stochastic? In a similar pattern of nasal assimilation and substitution in closely related Tagalog, based on dictionary and corpus data, Zuraw (2010) finds that the variation is largely lexicalized, variable on a word-by-word basis.

Kurniawan (2015) addresses these issues using the Betawi Jakarta Corpus. He analyzed the observed patterns for the realization of the prefix using data from speakers who had a relatively high number of tokens of active (N-prefixed) forms and who could clearly be identified as Jakarta Indonesian (JI) speakers (6) or Betawi (BM) speakers (15) based on the corpus metadata. As reported by Kurniawan (2015) and further discussed by Cohn & Kurniawan (2016), the number of tokens by speaker, by place of articulation, etc. show a large range, so we consider the findings as proportions out of 1. Looking at results by place of articulation across all speakers with a relatively high number of relevant forms, we see in Table 1 that place of articulation does seem to affect the observed patterns:

Table 1. Proportion of nasal assimilation (homorganic clusters) across place of articulation. Root-initial labial and palatal places favor an assimilated nasal allomorph over [ŋə-], while velar and alveolar consonants are preferentially preceded by [ŋə-] rather than [n ŋ].

Labial	Alveolar	Palatal	Velar
0.55	0.27	0.55	0.45

Most notably, it is more common to use the un-assimilated [ŋə-] form for alveolars and equally likely to use either form for the other places of articulation. Does dialect background contribute to the observed patterns? Looking at the data by Betawi vs. Jakarta speakers pooled across place of articulation (Table 2), again the results are suggestive, with Betawi speakers using more assimilated forms than Jakarta speakers.

Table 2. Proportion of nasal assimilation (homorganic clusters) form by variety. Betawi speakers favor allomorphs [m n ɳ ŋ], while Jakarta speakers favor un-assimilated [ŋə-].

Betawi	Jakarta
0.57	0.23

However if we compare Betawi vs. Jakarta speakers' outcomes by place of articulation, we see that the data are not evenly distributed. For labials, Betawi speakers use assimilated forms much more than Jakarta speakers. For alveolars both Betawi and Jakarta speakers tend to use unassimilated [ŋə-] forms more than assimilated forms, but the number of tokens, especially for Jakarta speakers, is very small.

Table 3. Inter-speaker and place of articulation differences: proportion of nasal assimilation (homorganic clusters) across place of articulation and speaker variety (Kurniawan 2015: 156, (8)). Speakers' preferences for allomorph type depend both on root-initial place of articulation and on their linguistic background.

	Labial	Alveolar	Palatal	Velar
BM	0.72	0.27	0.55	0.51
JI	0.14	0.29	0.5	0.33

Both place of articulation and speaker background appear to condition variation, but we can quickly see that neither of these is deterministic. We might think that another contributing factor is lexicalization, with particular forms typically produced with one or the other variant. Or we might think that token frequency affects the degree of variability. However, neither of these possibilities, while potentially contributing factors, is deterministic because the same lexical form can be realized with both variants by the same speaker, as illustrated in Table 4. Further, this observed variation does not appear to be determined by stylistic differences, since the whole corpus is based on conversations recorded in informal settings.

Table 4. Intra-speaker variation for single lexical items (number of tokens) (Kurniawan 2015: 157, (9)).

BM Speakers	Underlying representation	Homorganic clusters	Epenthetic schwa
DADBTW	/N+gajəm/ 'to eat'	ŋgajəm (1)	ŋəgajəm (1)
AFRBTJ	/N+daptar/ 'to register'	ndaptar (1)	ŋədaptar (1)
SALBTW	/N+gəletak/ 'to lie down'	ŋgəletak (1)	ŋəgəletak (1)
MLYBTJ	/N+batɕa/ 'to read'	mbatɕa (2)	ŋəbatɕa (2)
SIRBTJ	/N+ d͡ʑual/ 'to sell'	ɲd͡ʑual (2)	ŋəd͡ʑual (1)
JI Speakers	Underlying representation	Homorganic clusters	Epenthetic schwa
EXPOKK	/N+dəŋər+IN/ 'to listen'	ndəŋərin (1)	ŋədəŋərin (1)

The annotations for this corpus include metadata allowing us to separate Jakarta speakers from Betawi speakers, and thereby to identify variation related to native dialect. However, the ability to identify individual speakers reveals variation even at that level. This level of annotation detail reveals that factors of *phonological context*, *native dialect*, and *individual speaker* help determine when each variant is likely to occur in everyday, conversational speech. These results highlight what we can learn from integrating naturalistic data into our other modes of analysis. In the absence of data annotated with this level of detail, a phonologist might be tempted either to model the observed variation as stochastic or free, or to consider only a single speaker's data and potentially over- or underestimate the likelihood of each variant for other speakers. Considering multiple factors adds greater precision to our models of phonological variation.

Despite this enriched detail, the BJC does not permit us to draw firm conclusions regarding the relative contributions of context, dialect and individual on the morphophonemic realization of this prefix. Although the BJC contains over 75,000 utterances, the available token counts illustrate how *small* the data become at the level of individual speaker. This corpus is a very useful tool for investigation, and it is large by the standards of traditional phonological analysis, contains nowhere near enough data once we start to look at specific phonological factors affecting observed patterns, combined with the relative contributions of linguistic and socio-indexical factors. We could strive to develop larger corpora, or to use results from corpus analysis as a jumping-off point for more focused experimental work. We pursue both possibilities here, beginning with the benefit of a truly large corpus for modeling English phonology, and we report on a follow-up experiment on nasal prefixes in Jakarta Indonesian in Section 3.2.

2.3 Nasal assimilation in the Audio British National Corpus

A goal of phonology is to have explicit formal accounts of alternations, accounts which neither over- nor under-apply with respect to the patterns observed in spoken language. Another, more specific goal is to understand the domain of a particular alternation: for instance, does it only occur locally to adjacent segments, or can it apply in a long-distance fashion, or across word boundaries? This case study shows how a sufficiently large dataset can help achieve these goals, even in the face of a rare and variable alternation.

A corpus such as the 5-million word Audio British National Corpus (Audio BNC) (Coleman et al. 2012) exhibits characteristics that are ideal for in-depth modeling of the factors that condition sound patterns: it is very *large* and *naturalistic*, and is accompanied by *speaker metadata*. A positive consequence of the intersection of the first two characteristics is that such a corpus permits the analysis of words and phonological phenomena that are understudied due to their rarity, which results either from infrequent occurrence or by restriction to very naturalistic or informal speech registers. For example, it is worth noting that the Romanian corpus described in Section 2.1 contains approximately 7 hours of speech, vs. more than 1200 hours in the Audio BNC. The Romanian corpus entirely lacks certain phonological structures which are known to be possible; for instance, /ɨ/ can appear in word-final position, in a subclass of infinitive verb forms, but zero tokens of that vowel were found word-finally in the broadcast corpus, simply because the relevant infinitives are very rare. A larger corpus is expected to exhibit a greater range of infrequent forms.

Additionally, in a large corpus it is possible to conduct searches that are highly restricted to particular phonological contexts; the larger the corpus, the more contexts, in terms of both token and type frequency, are likely to occur. This means that frequent contexts are represented by more tokens, but also that a greater range of contextual patterns (types) can be sampled. Typically large-scale studies abstract away from individual differences, for example by normalizing acoustic data, conducting statistical analysis with mixed-effects models including random effects for speaker, or simply by averaging across speakers. Given a sufficient amount of data (for instance, hundreds or thousands of tokens), these techniques provide a quantitatively reliable illustration of trends in a corpus. However, with the addition of metadata that is also available for portions of the Audio BNC, it is possible both to generalize at the level of the speech community, and to show that individual speakers reflect specific phonological patterns.

In an analysis of the Audio BNC, Coleman, Renwick and Temple (2016) provide evidence of place assimilation by English nasals that are underlyingly velar or labial. This is a striking phonological finding, since it is unpredicted according to prevailing views that only coronals can assimilate, which in turn were used for phonological models of coronal underspecification (Avery & Rice 1989).[4] As an alternative, Coleman et al. (2016) propose a model of probabilistic phonetic underspecification according to which nasals at all places of articulation can assimilate, but at different rates or to varying degrees. The evidence for labial and velar assimilation by nasals in word-final position comes principally from acoustic data sampled across hundreds of tokens, which show that the formant values of labial and velar nasals differ significantly according to the place of articulation of the following consonant: for example, the F2 values of underlying [m] are lower before [b] than before [t], suggesting considerable coarticulation or assimilation with the following consonant. If categorical assimilations occur, the authors expect to find a *bimodal* distribution of acoustic values, with one peak corresponding to realizations of labial [m], and other to assimilated [n]. When the data are pooled across speakers, a unimodal distribution occurs, which Coleman et al. (2016: 443) argue is due to large variability across speakers, obscuring any underlyingly bimodal distribution. However, by analyzing tokens of /m/ from a *single* speaker in the corpus, they are able to show that bimodality can occur. As shown in Fig. 2, among the tokens of /m/ before /t, d/ (pairs of words like *seem to*, which speaker "Fred" was heard to pronounce as *seen to*), there is a strong F2 mode around 1200 Hz, corresponding to a labial place of articulation. But Fred also produced a subset of tokens with F2 values between 1800 – 2300 Hz, which corresponded to audibly assimilated nasals. While the distribution of /m/ before /p, b/ is statistically unimodal (according to a Hartigans' dip test, D = 0.0067, $p \sim 0.99$), the pre-coronal test cases like *seem to* are not (D = 0.0565, $p < 0.05$).

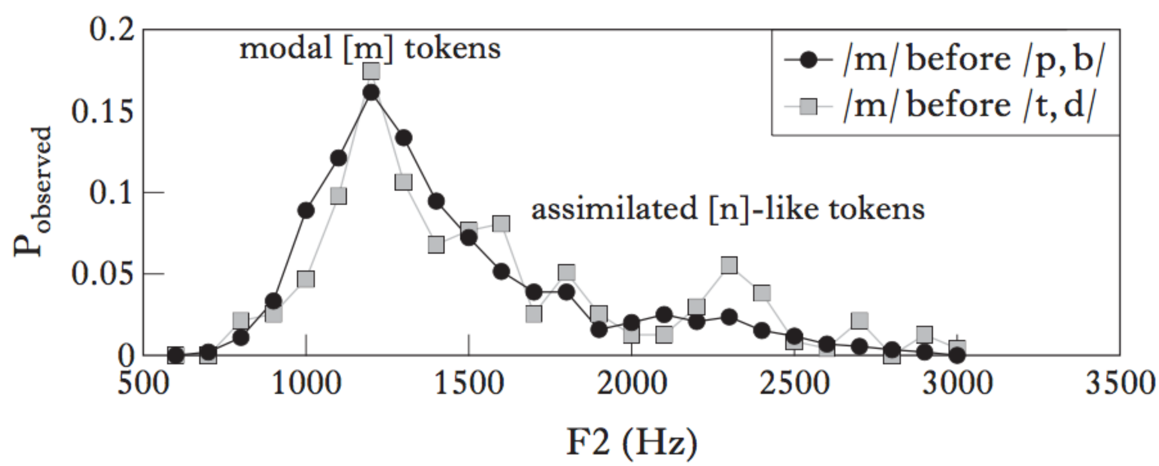

Figure 2. Histograms of F2 frequency of /m/ in all 1670 tokens produced by speaker "Fred". 235 tokens of /m/ preceded coronal consonants. Reproduced from Coleman, Renwick & Temple (2016, Fig. 7).

The authors argue that the Audio BNC data prove the existence of a phonological process that others have failed to recognize, and that is in fact excluded by their formalizations; that is, nasal

[4] For an alternative modeling perspective on coronal assimilation, which integrates perception results with noisy Harmonic Grammar, see Coetzee (2016).

assimilation can occur at *all* places of articulation in English, and its scope can cross word boundaries. The Audio BNC data motivate a proposed modification to phonological theories that exclude assimilation in non-coronal contexts; this modification permits non-coronal assimilation, but also accounts for its relative rarity.

With a sufficiently large corpus, it is possible not only to seek differences in central tendencies across many speakers, but also to verify the occurrence of phonological phenomena at the speaker-specific level, determining whether the apparent phonetic pattern is an artifact of pooling results from a diverse group of speakers. The Audio BNC has the added benefit, due to its large size, of containing phenomena and phonological contexts that are rare. Still, even this corpus containing millions of words has its limitations; within the corpus, the combination of a nasal velar adjacent to another velar consonant across a word boundary is very rare, producing only 112 usable tokens across the entire corpus (Coleman, Renwick & Temple 2016: 429). A follow-up study of the acoustics of word-final velar stops, using the Santa Barbara Corpus (DuBois et al. 2000), also encountered limitations of "vanishingly small" data (Lipani 2017). As discussed further in Section 4, we propose that this type of situation can be addressed by combining large corpus studies with more targeted experimental work.

2.4 Desiderata of corpora for studying phonology

We have shown that studying corpus data allows us to uncover sound patterns and their variation more accurately than using elicited data alone. These data-rich resources are powerful tools that can illuminate the relationship between competence and performance from angles that are obscured by impressionistic techniques. Here, we outline what is desired in a corpus used for phonological research, along with benefits and pitfalls of such an undertaking. A corpus used for phonological research ideally consists of *naturalistic speech data*, meaning that it captures unelicited, spontaneous, continuous native speech in a communicative context. While such a dataset might exist in transcribed form only, the categorization of sounds required by even very close transcription obscures finely detailed variation and can be unreliable.[5] A corpus for phonological research needs audio files which are transcribed either orthographically or phonetically (or both), in a way that links them in time, for example via Praat TextGrids (Boersma & Weenink 2017) or via Elan – Linguistic Annotator (https://tla.mpi.nl/tools/tla-tools/elan/). The corpus needs metadata, minimally to separate one speaker from another, but ideally including relevant demographic information (sex/gender, age, ethnicity, socioeconomic class, education level, occupation, location, linguistic background, etc.). Additionally, if multiple communicative contexts or speech registers are included in the corpus, these should be identified; for example the Audio BNC is split into an informal portion collected by volunteers who carried tape recorders during portions of their everyday interactions, and a more formal (but still unscripted) portion collected during e.g. interviews and religious services. As language use can easily vary across contexts and between different interlocutors in the same setting, researchers need information about these factors.

A consequence of collecting speech "in the wild" is that the resulting corpus tends to be highly unbalanced, with some forms occurring exponentially more often than others, following a Zipfian distribution (Zipf 1935). On the one hand, this means that rare or previously unnoted forms may

[5] An impressionistically transcribed, or even orthographic, corpus can be sufficient for some phonological and morphophonemic patterns, such as the variation in allomorphy in Finnish nicely documented and modeled by Anttila (1997).

crop up, particularly those that are restricted to informal registers, as found in the Audio BNC case study. On the other hand, some phenomena are very well represented while others may fail to occur. Similarly, naturalistic recording conditions result in a significant amount of noise in the signal (including outliers, which can be checked manually or filtered automatically; overlapping speech, annotation errors, and literal noise), meaning that not all data can be processed and analyzed. Noisiness is both a drawback and an advantage, because under everyday speech conditions, where competence and performance meet, imbalance and noise are part of reality.

Another consideration is that the size of large speech corpora means that analysis by hand is typically not possible. Thus rather than conducting analyses in which the phonologist listens to and categorizes each token, researchers who carry out analyses of large corpora use automatic processing techniques, to extract and analyze a large number of speech tokens from an even larger sample, leading to statistical modeling rather than detailed description, as discussed for the Audio BNC. Access to large corpora is improving thanks to resources like the Linguistic Data Consortium (https://www.ldc.upenn.edu/) and the increasing presence of corpora freely available via the Internet. Usability of corpora is also greatly augmented by techniques like automatic forced alignment, which aligns a transcription in time with the corresponding audio signal, typically at the level of the segment, to produce a searchable database in much less time than would be required with manual annotation (Schiel 1999; Yuan & Liberman 2008; Young et al. 2009; Gorman, Howell & Wagner 2011; Rosenfelder et al. 2011; Bigi & Hirst 2012; Reddy & Stanford 2015; Goldrick et al. 2016; Kisler et al. 2016; McAuliffe, Socolof, et al. 2017; McAuliffe, Stengel-Eskin, et al. 2017). Apart from the issues outlined above, corpora are labor-intensive both to create and to analyze, although once the data are collected and annotated, the work becomes much easier; for recent examples of corpus analysis, see Stuart-Smith et al. (2015) and Sonderegger et al. (2017). This approach extends to less commonly studied languages, as well: meaningful analysis is possible even with a small corpus (DiCanio et al. 2013; DiCanio et al. 2015; Bennett, Tang & Sian 2018); and we want to emphasize that even limited naturalistic data greatly enrich our understanding of phonological phenomena.

There are several considerable advantages of phonetic corpus research for phonologists. First, naturalistic speech permits the study of communication "in the wild," which in sufficiently large quantities is likely to contain rare or elusive (but often systematic) phenomena. The accompanying audio recordings provide access to continuous phonetic data, in which variation can be observed *in situ* and can be linked to multiple conditioning factors.

Nonetheless, we acknowledge that corpora are not and should not be the definitive source of data for phonological modeling. If the corpora available are insufficient to answer a particular research question, or if a balanced dataset is required to test a specific hypothesis or establish causality, or if less noisy recordings are essential, then corpus data alone will not be sufficient.

We advocate a multifaceted approach that investigates spoken communication through methodologies including impressionistic work, controlled data, and naturalistic data (see Warner 2012; Scobbie & Stuart-Smith 2012 for discussion of methods for phonologically-motivated research with these types of data sources). An initial illustration of this workflow is shown by the Romanian case study; distributional facts about the language (the highly-complementary distribution of /ɨ ʌ/) were noted through phonological and dictionary analysis as well as historical research, spurring laboratory studies of both production and perception, with confirmation of the distributional observations via a corpus study. Another aspect of phonological variation highlighted by the case study of Indonesian is the relevance of *individual speaker variation*, which may be studied via corpora but can also be elicited under controlled circumstances. As shown above by the project on nasal place assimilation across word boundaries, naturalistic data can bring to light new generalizations, in this case about place assimilation of /m ŋ/, suggesting gradient and variable patterns. If a corpus is studied speaker by speaker we may well find categoriality and greater systematicity than seen in the gradient and variable patterns observed across the same corpus. Since phonology strives not only to capture communicative speech patterns at the group level but also as an expression of the *individual's competence*, analysis of data speaker by speaker is essential. We turn to this topic in the next section.

3 Inter- vs. intra-speaker variation

In order to achieve our goal of a more nuanced and rich theory of linguistic competence, we need to remember that language is ultimately a system of communication, and that any seemingly stochastic pattern may indeed directly represent variability, or may be the interaction of two or more more systematic patterns. Yet, as illustrated in Section 2, the sources of variation can be very different, and identifying the source is crucial to accurately modeling the variation. Even the most "regular" (morpho-) phonological patterns that we understand as conditioned by well-known phonological factors (neighboring sounds, syllable position, word position, prosodic structure) might have some exceptions (e.g. regular plural marking in English and exceptions such as *mice, children, ox,* etc.). To say these patterns occur either 100% or 0% of the time is thus a slight fiction. This is different, however, from patterns that might occur 75% or 60% or 50% of the time.

While on the face of it, our standard phonological models cannot account for such stochastic patterns, there is an increasingly rich body of literature modeling variation and offering insight into how such variation fits into formal models. First there is a long standing tradition of modeling variation in sociolinguistics using variable rules and other approaches (Guy 2014). Moreover, recently a number of formal phonological models have been developed that integrate variation into formal models. For recent reviews of this work see Coetzee & Pater (2011), Coetzee (2012), Coetzee & Kawahara (2013) and Anttila (2012; 2018). However, the point we are making here is that we need to be sure the data themselves are valid and reliable, and provide a sufficiently rich description. We need to determine if the observed patterns are truly variable, or whether the pattern results from the interaction of multiple systematic factors (as considered in a recent study of French schwa by Bayles, Kaplan & Kaplan 2016). Such formal models bridge towards adequate and accurate "computational models of the processes of speech production and recognition" (Coleman 2011: 599), which researchers have sought across generations of generative phonology.

Consider the simple example of phonologically conditioned variation in English marking, where the regular plural suffix is realized as [-z] after voiced sounds and [-s] after voiceless sounds, unless the sound is a sibilant in which case the plural is realized as [-ɨz] or [-əz]. If we counted up all the regularly plural marked forms in a corpus, at first blush we would see a stochastic pattern, but we would quickly realize that the forms are largely in complementary distribution and we would attribute the observed variation to phonological conditioning factors. Some patterns of variation are also clearly lexicalized, as it the case for "irregular" plural forms such as Ø, ablaut and so forth. It is equally important to look at the degree of systematicity in patterns in the pragmatic and socio-indexical domains, which might in the past have been dismissed as "free variation". We focus here on the latter domain, asking most fundamentally: are observed patterns due to inter- or intra-speaker variation? In the most clear-cut cases inter-speaker variation might be attributable to a systematic dialect difference, while a pattern of intra-speaker variation might be attributable to a stylistic difference. Our point is a basic one: the source of variation cannot be identified unless we can look at data speaker by speaker, and both experimental work and naturalistic data require this level of analysis. We review here several cases highlighting the need to incorporate speaker-specific information in order to understand and model variation.

We first review a striking demonstration of this point by Scobbie (2006) based on voicing contrasts in the Shetland Isles (3.1). We then return to the question of intra- vs. interspeaker variation observed in the active prefix in Indonesian (3.2). We finally explore further the difference among facets of variation by contrasting the vowel systems of two fairly closely related languages, Italian and Catalan (3.3).

3.1 Shetland Isles voicing

In order to characterize phonological inventories and alternations, most researchers rely on a set of distinctive phonological features, whose precise universality, nature, and grounding are themselves a subject of study. Scobbie (2006) sheds light on the phonetic realization of Voice Onset Time (VOT) relative to stops, illustrating their language-specific nature. In so doing, he offers a striking example of the need to differentiate inter- vs. intra- speaker variation, looking at VOT as a cue to the voicing contrast in speakers in the Shetland Isles. Based on laboratory data from 12 participants, he finds extensive variation in the realization of /b/, which form a continuum from fully voiced to voiceless to moderately aspirated. This seems to call into question the robustness of the voiced/voiceless opposition as a coherent phonological distinction, insofar as VOT is not a reliably categorical acoustic correlate, as shown in Figure 3.

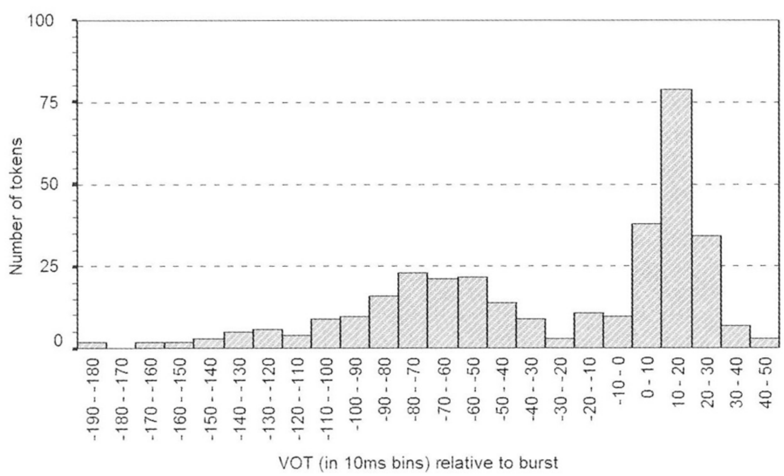

Figure 3. Histogram of Voice Onset Time (ms) for /b/, all subjects pooled; reproduced from Scobbie (2006, Fig. 2)

However, Scobbie (2006: 367) observes ‘The Voice Onset Time (VOT) cue to the /p/-/b/ voicing contrast in Shetland Isles English was found to demonstrate a high degree of interspeaker variation, which was non-arbitrary on two counts.’ Once we consider the results by speaker and even more interestingly, take note of the linguistic background of the participants’ parents, a much more systematic and striking pattern emerges. The subjects’ background is reproduced in Table 5 and the results by speaker are shown in Figure 4.

Table 5. Subject identifiers and age details coded by sex and group; reproduced from Scobbie (2006, Table 1).

Parents born and raised in	Shetland	Scotland	England
Male Shetlanders	S1: age 20	S5: age 16	S9: age 20
	S3: age 21	S7: age 17	S11: age 19
Female Shetlanders	S2: age 30	S6: age 17	S10: age 17
	S4: age 22	S8: age 17	S12: age 17

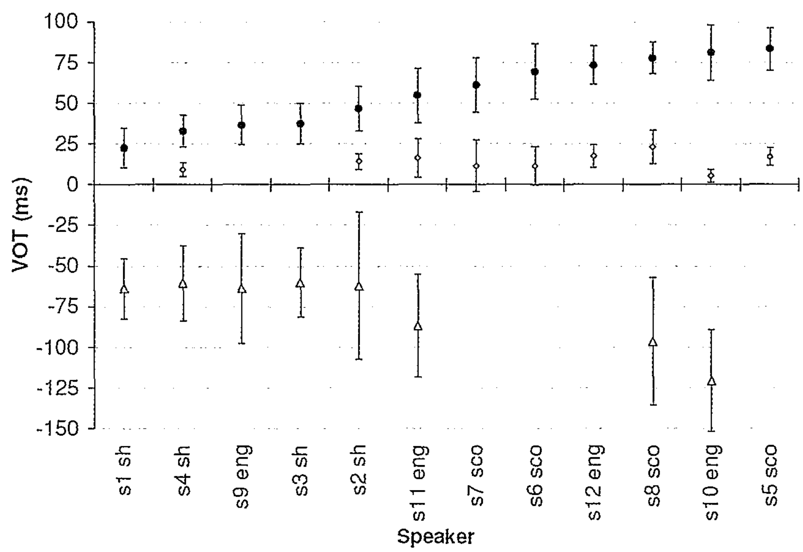

Figure 4. Mean Voice Onset Time (ms) relative to burst; reproduced from Scobbie (2006, Fig. 3). "sh" indicates parents born and raised in Shetland; "eng" indicates English parents; "sco" indicates Scottish parents. Filled circles indicate VOT for /p/, while diamonds and triangles indicate positive and negative means for /b/, respectively.

Figure 4 shows that the continuous range of VOT values for /b/, at the interspeaker level, hides what is actually a systematic /p/-/b/ contrast for each speaker. Some speakers show a clear voiced-voiceless contrast (particularly those whose parents were also born and raised in Shetland, cf. Table 5); while others use an unaspirated – aspirated (short lag – long lag) contrast. Some speakers show a fairly wide range of values within categories, or even a bimodal difference within a category, possibility due to the influence of the mixed dialect system they are exposed to; but crucially every one of the speakers has a robust system of contrast, very different from the picture suggested by the combined data presented in Figure 3. Additional explanatory power comes not only from considering speakers individually, but also from metadata regarding their parents' origins as an indicator of linguistic background. This illustration of how a particular contrast is phonetically realized shows that apparent variability at the community level may obscure motivated systematicity at the individual level, and as Scobbie (2006: 367) concludes, they show that 'indexical and phonological cues are simultaneously present [...] and that both categorical and gradient approaches to phonology must be pursued.'

With this example in mind, we return to the case of the Indonesian active prefix.

3.2 More on nasal prefixes in Jakarta Indonesian

As noted above in Section 2.2, the realization of the active prefix before voiced verbal stems is variable in Jakarta Indonesian and Betawi. The corpus results show speaker differences that might be consistent with a dialect difference and also show intra-speaker variation even for the same lexical item, so this phenomenon cannot be attributed to particular lexical items. Since the data in the corpus were insufficient to fully tease apart these factors, Kurniawan (2015) conducted a production experiment to better understand the source(s) of variation in this case. The goal is still to understand the factors conditioning each variant, in order to formalize and predict when each will occur, to deepen our phonological understanding of Jakarta Indonesian.

In the experiment, participants were presented with a sentence in the passive voice and asked to produce the same sentence in the active voice. These sentences included existing verbs beginning with [b d d͡ʑ g] (all found in the corpus to show variation), with distractors starting with [p t k], and each verb embedded in two different sentences. The sentences were constructed in a way expected to cue informal speech, consistent with forms used in Jakarta Indonesian. An example is shown in (5).

(5) Stimuli in study of nasal prefixes in Jakarta Indonesian

Speaker hears:

uang	itu	udah	*dibalikin*	Toni
money	that	already	Pass-return-Ben/Caus	Toni

'That money was returned by Toni'

Expected response:

Toni udah *mbalikin/ngebalikin* uang itu.

'Toni has returned the money.'

Eight speakers (3 male, 5 female), between the ages of 25 – 45 years old participated in the experiment which was conducted in Ithaca, NY and all eight participants identified as Jakarta Indonesian speakers.

The overall results are shown in Figure 5. Extensive variation is readily apparent at each place of articulation, though the unassimilated forms account for the majority of forms in each case. Next it is noteworthy that despite the use of sentences coding informal speech, a small percentage of forms were produced with the form of the prefix from Standard Indonesian, /məŋ-/; and particularly interesting is the extensive use of bare-stemmed forms which are unexpected in this grammatical context, but have been observed anecdotally. In particular, speaker S1 made extensive use of bare stemmed forms, so this speaker is excluded from the speaker specific analysis that follows.

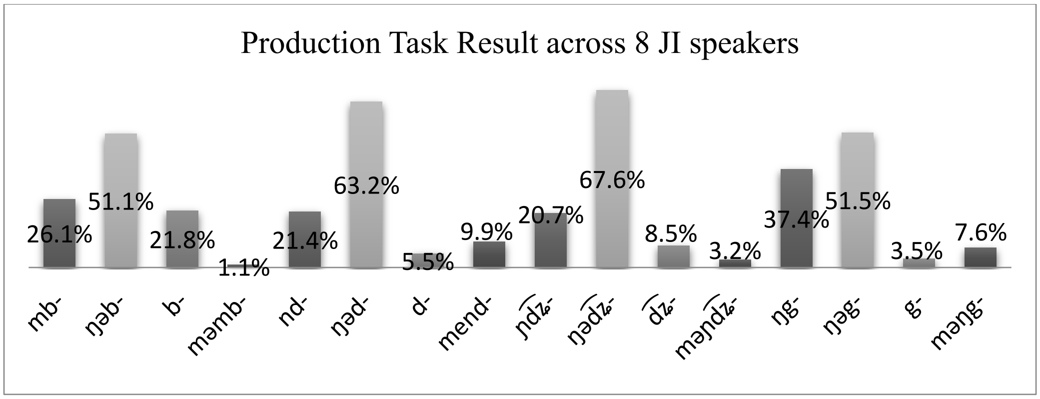

Figure 5. Production task results across 8 Jakarta Indonesian speakers; reproduced from Kurniawan (2015: 160 (11))

Looking at the results by speaker and pooled for place of articulation, we observe three distinct types of speakers (labeled type A, B, & C), as shown in Figure 6.

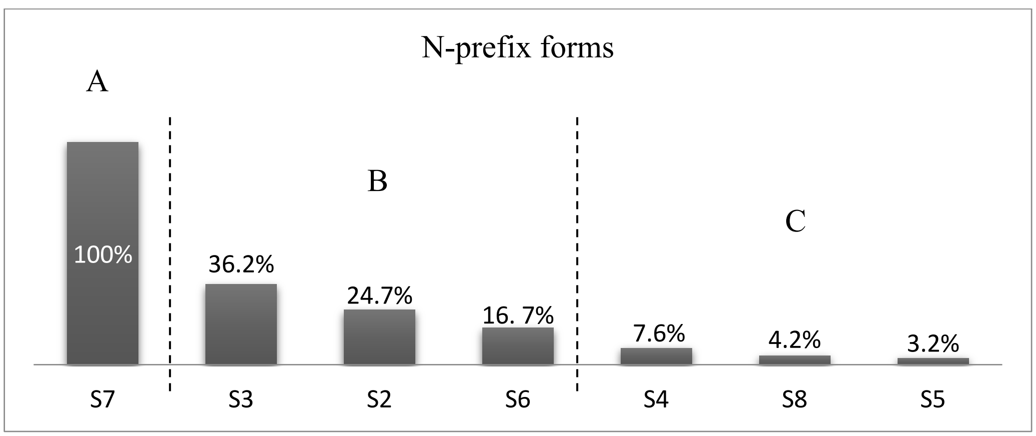

Figure 6. Production task results by individual Jakarta Indonesian speakers, pooled across place of articulation, percentage of assimilated forms; reproduced from Kurniawan (2015: 161 (12))

First, type A is exemplified by one speaker, S7, who used nasal assimilated forms 100% of the time. (This is suggestive of the pattern seen in the corpus results for Betawi speakers, who at least for the labial place of articulation showed a strong preference for nasal assimilated forms). On the other hand, type C consisted of three speakers (S4, S8, S5) who used almost solely unassimilated [ŋə-] forms (with assimilated forms ranging from 3-8% of productions). Finally, type B includes three speakers (S3, S2, S6) who showed considerable variability (ranging from 16-37% nasal assimilated forms).

This tripartite grouping can be tied back to the observations seen in the naturalistic corpus (Section 2.3), confirming that inter- and intra-speaker variation occur in both the corpus data and production experiment. The production experiment shows that individual speaker differences are a very important factor in explaining variation in allomorph selection. All of these participants are speakers of Jakarta Indonesian, but unlike Scobbie's findings on individual variation in VOT, these results do not tie directly to dialect variation, nor can they be explained by degrees of formality. Perhaps additional factors, such as discourse context and addressee, are at play. These studies of spoken Indonesian show that an iterative method of laboratory and naturalistic investigations can offer greater empirical understanding of the sources of variability, highlighting these ingredients' relevance to an integrated model. This case also illustrates a case where higher-level morphophonemic alternations can be studied without a more detailed acoustic analysis.

3.3 Marginal contrast in Romance mid vowels

A third case study focusing on data from individual speakers is motivated, like the studies of Romanian, by an ostensibly *marginal contrast* between vowels, this time in Italian and Catalan. These contrasts are worthy of study because they may be prime candidates for phonological change over time; it has been shown that contrasts with low functional load have an increased likelihood of merger (Wedel, Kaplan & Jackson 2013; Wedel, Jackson & Kaplan 2013). The studies of Italian and Catalan share the perspective that marginal contrastiveness is not simply a distributional fact about pairs of sounds, but a phenomenon with consequences for speech production and perception, linked to an increased likelihood of phonetic overlap or perceptual confusion (Nadeu & Renwick 2016; Renwick & Ladd 2016; Renwick et al. 2016). However, it is also clear that the sources and manifestations of marginal contrasts are diverse, and one question is whether a contrast must be marginal across an entire speech community, or whether one factor in phonemic robustness (Renwick 2014) is the implementation of a contrast *at the level of individual speakers*. If a contrast is robust for individuals but implemented differently *across* individuals, two consequences follow: first, a phonological model built at the level of the community does not reflect reality for any individual's grammar (the presence of interspeaker variability may cause the contrast to appear weaker than it actually is). Second, the presence of variability in a larger speech community may weaken individuals' contrast representations, increasing the likelihood of diachronic contrast neutralization.

In the Romance languages Italian and Catalan, the behavior of mid vowels encourages us to ask: what are the consequences for phonology if speakers disagree on *which phoneme to use* in a particular word? Studies of these languages explore marginal contrasts from the individual's perspective, using an experimental methodology in which the acoustics of speakers' productions are directly compared to their phonological intuitions about particular sounds. Both investigate the strength of phonological contrasts between pairs of higher and lower mid vowels /e ɛ, o ɔ/, in Italian (Renwick & Ladd 2016) and in Catalan (Nadeu & Renwick 2016; Renwick & Nadeu 2018). While most speakers have accurate intuitions about their own speech, a variety of intuition-production mismatches do occur.

In Standard Italian, a few minimal pairs are said to distinguish the open and close mid vowels, as in /venti/ 'twenty' vs. /vɛnti/ 'winds.' However in a particular word like *cento* 'one hundred,' speakers may produce different vowels, e.g. either [ˈtʃento] or [ˈtʃɛnto]. This *can* be related to dialect differences, as differences in lexical selection of mid vowels are often a highly salient indicator of speaker dialect; but variability is found within a single dialect, as well. Similarly in Catalan, these vowels are contrastive under stress, as in /be/ 'good' vs. /bɛ/ 'lamb'; /os/ 'bear' vs. /ɔs/ 'bone'; yet inter- and even intraspeaker variation is observed, leading scholars to report two normative pronunciations for some words (Recasens 1993). In Italian and Catalan, the mid vowel contrasts thus seem less robust than others: they neutralize in unstressed position, though in different ways; vowel height is partially phonologically conditioned, and varying across dialects; and the higher mid and lower mid vowels may be undistinguished orthographically.

In parallel studies of both languages, Renwick and colleagues tested for the "particular closeness" among mid vowels, as described for French by Trubetzkoy (Trubetzkoy 1939; Ladd 2006). It was hypothesized that phonological intuitions of mid vowel height would vary across speakers, and that speakers' own judgments might not always match their pronunciation. Native speaker participants (17 Italian, 14 Catalan) first produced a set of words containing a target stressed mid vowel, as well as the other 5 stressed vowels. They subsequently judged the height of the same target mid vowels, providing speaker-specific phonological judgments to accompany target words. The judgments revealed considerable interspeaker variation: very few words were judged uniformly by all speakers. The acoustics of individual vowel productions, measured in F1 and F2 values, were plotted and compared to speakers' judgments. Acoustic mid vowel tokens were also classified using a k-means clustering algorithm to provide an objective measure of height class. K-means clustering results were compared with speaker intuitions to identify cases of mismatch, where vowels' F1/F2 values placed them in the opposite vowel cloud from that matching the speaker's intuition.

3.3.1 Phonological closeness in Italian mid vowels

Results showed that in Italian, most speakers have clear phonetic distinctions between high and low mid vowels: speakers' /e ɛ/ and /o ɔ/ show little overlap in the vowel space. In fact, the mid vowel pairs /e ɛ/ and /o ɔ/ tend to be farther apart in F1 space than they are from adjacent point vowels /i a u/; Renwick and Ladd (2016: 14) thus argue that mid vowels are *phonetically* distinct, and that identifiable mismatches are not the result of phonetic overlap between mid vowel clouds. Though phonological awareness for all speakers is generally high, Renwick and Ladd observe cases where production and speaker judgment fail to match. Among the judgments themselves, however, it is possible to observe systematic patterns that identify individual speakers, and in fact group them according to their regional variety of Italian.

Here, we view Renwick and Ladd's (2016) data from a new angle. Using the dictionary's assignment (DeMauro 2000) of mid vowel quality as a normative standard of comparison, each judgment (100/speaker) was assigned a value of 1 if it matched the dictionary, and 0 otherwise. This produced a matrix with 100 columns (1 column/word) and 17 rows (1 for each speaker), which was subjected to hierarchical clustering via the `hclust()` algorithm in R (cf. Baayen 2008; and see Clopper 2012 for an overview of clustering methods in a laboratory phonology approach). This analysis produces a tree whose branches terminate in speaker labels. Speakers who are grouped more closely in the tree are "clustered" together, which in this instance means that they provided similar phonological intuitions for specific words. The results are shown below in Figure 7, which shows strong evidence that speakers from the same region (i.e. North vs. Central vs. Northeast) give similar judgments. These intuitions largely match speakers' individual production patterns (see Renwick & Ladd 2016 for details). It is crucial to note that these patterns are *not visible* when data are collapsed across speakers, because no two speakers had identical phonological intuitions or production patterns. In other words, we cannot assume that phonological patterns and phonolexical mappings are identical across speakers, and we need methods that take into account differences in individuals' grammars.

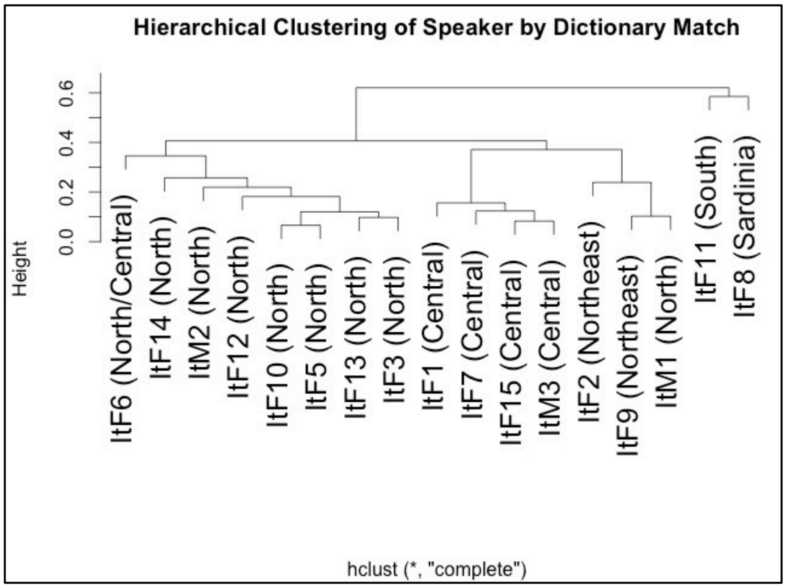

Figure 7. Italian speakers grouped by hierarchical clustering, based on whether their judgments of mid vowel quality matched the prescriptive standard

3.3.2 The relationship between vowel production and phonological intuitions in Catalan

In Catalan, substantial variation is found in the phonetic implementation of the front and back mid vowel contrasts (Nadeu & Renwick 2016). All speakers maintain a height distinction between /e ɛ/ and /o ɔ/ according to Pillai scores, a metric used to quantify overlap in F1,F2 space (Hay, Warren & Drager 2006); however, some speakers' mid vowels were highly distinct, while for others the vowels' distributions overlapped considerably.

Figure 8 provides a comparison, for three speakers, of vowel spaces mapped according to speakers' judgments (SpeakerV) vs. a standard imposed by the experimenters (ExperimenterV), which shows both the divergence between the two standards and also the diversity among individual speakers. For most speakers, the individual speaker's judgments were a better match for the acoustic realization. In Fig. 8, f05 and f14 largely agree with ExperimenterV, but are better judges of their own pronunciation: for speaker f05, several tokens labeled /e/ by ExperimenterV are realized in the [ɛ] cloud, and vice versa. The same occurs among the back vowels. Speaker f14 represents a case of moderate overlap, and is a better judge of her pronunciation than ExperimenterV. The situation is not the same for f10: in her SpeakerV plot, both front mid vowel ellipses are realized in the same acoustic space, and likewise for the back vowels. This speaker does not appear to have accurate intuitions of her own mid vowel height, although she does produce clouds that occupy the appropriate phonetic regions. These speakers are representative of other speakers in the study: some have highly distinct mid vowels and are excellent judges of their own pronunciation, while others have more phonetic overlap and are less-precise judges of their own speech.

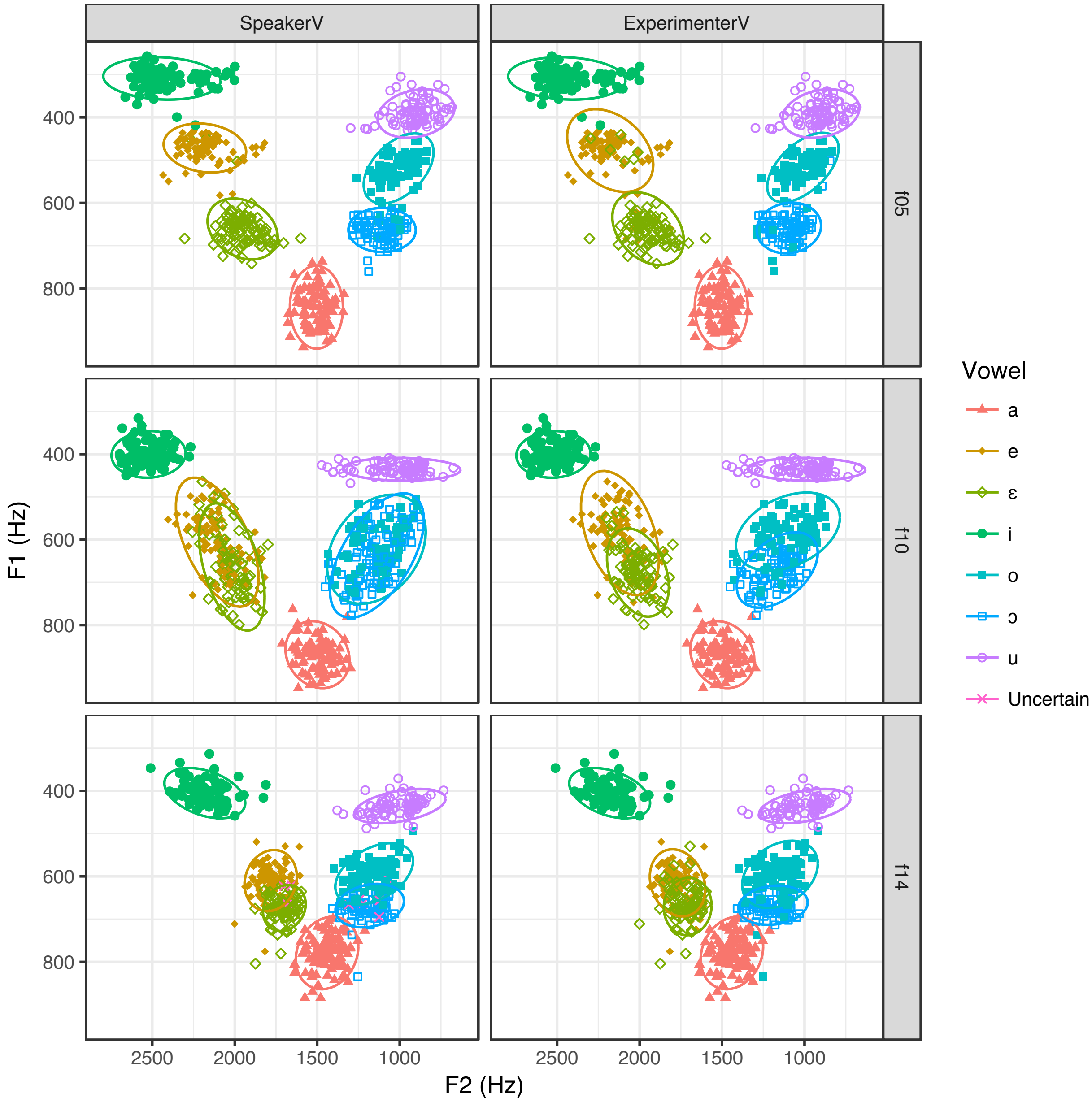

Figure 8. Vowel spaces of three Catalan speakers: f05, f10, f14. For each speaker, the data points represented are the same in the two plots. However, mid vowel height is determined on the left by the speaker's judgment (SpeakerV) and on the right by the experimental standard (ExperimenterV). Ellipses are based on 95% confidence intervals and, for the SpeakerV plots, they exclude any "Uncertain" tokens. Each point represents one repetition of a target vowel. Reproduced from Nadeu & Renwick (2016, Fig. 2)

The crucial point is that even within dialects, the lexical distribution of the vowels is variable; a Fleiss' Kappa test was used to evaluate uniformity of judgments across speakers, and its result of 0.50 indicates only "moderate" agreement according to the Landis scale (Landis & Koch 1977). Neither the amount of phonetic overlap nor the number of misjudgments correlates with speakers' dominance in Catalan (vs. Spanish), evaluated with a bilingualism questionnaire (Birdsong, Gertken & Amengual 2012). This is significant because it has been argued that as dominance in Spanish increases, speakers' sensitivity to the mid vowel contrasts decreases (Pallier, Bosch & Sebastián-Gallés 1997; Bosch, Costa & Sebastián-Gallés 2000; Bosch & Sebastián-Gallés 2003; Navarra, Sebastián-Gallés & Soto-Faraco 2005; Bosch & Ramon-Casas 2011). However, in this study no such relationship was found, suggesting that speakers' phonolexical representations may vary independent of language background. A subsequent internet-based survey of mid vowel intuitions (Renwick & Nadeu 2018), found that a core of variably-judged words remains, and also that among a more diverse participant group, socioindexical factors can be relevant for phonological intuitions: Catalan dominance correlates with confidence in height judgments, while the selection of high vs. low mid vowel is conditioned by speaker age. Taken together, these studies reiterate the potential for individual variation to occur alongside identifiably conditioned variation. While the mid-vowel contrasts in Italian and Catalan are phonetically distinct and *can* be reliably implemented, the variation in *how* they are implemented across and within speakers suggests that they may be unstable, making them potential candidates for diachronic phonological change.

3.4 Interim conclusions

These studies each highlight the complex and nuanced ways that variation may be observed and instantiated across speakers and within speakers. Studies that focus on hypothesis testing often shy away from considering individuals in favor of pooled data that are amenable to statistical analysis, but this obscures the view of crucial, systematic detail, and an understanding of how models are implemented at the individual level. Within a broader intellectual shift that also affects psychology, greater awareness of individual speaker differences has grown in the experimental phonetic and laboratory phonology literature which can help resolve questions that impressionistic phonology alone cannot account for. For example, individuals' intuitions of syllable count have been found to correlate with production patterns in English words like *fire* (Tilsen & Cohn 2016), and syllabification of Italian initial /sC/ clusters is best understood via articulatory data analyzed at the individual level (Hermes, Mücke & Grice 2013). Ultimately only with such attention to speaker specific patterns of behavior, and with insights from naturalistic data will we be able to tease apart the complex interactions between linguistic, para-linguistic and socio-indexical factors. This more nuanced understanding provides a firmer empirical foundation for our phonological analyses.

4 Integrating methods for phonology

We have discussed a series of case studies that elucidate the advantages of broadening our methodological scope and embracing variation in doing phonology. Only by examining "messy" data, can we understand the patterns of variation and sources of that variation, the factors that condition it and the degree of systematicity it exhibits. While we have focused on this particular set of illustrations, we note that these studies fit in a growing trend of large-scale and finely detailed analyses whose goal is to enrich our understanding of linguistic competence and communication. We know that for many we are preaching to the choir, but we hope that others will be encouraged to engage with research methods described here. For some we have crossed a line by suggesting competence and performance cannot be strictly delineated. This does not mean that there is no such thing as competence distinct from the collective experience of performance, but we do think it is important to move beyond modeling "ideal speaker-listener, in a completely homogenous speech community" (Chomsky 1965: 3). (See Guy 2018 for recent discussion of this issue.)

Crucially, a fuller understanding of the factors that contribute to linguistic competence arises from use of a rich range of methods. First, production and comprehension of variation are intrinsic parts of competence in a communicative phonological system, and merit inclusion in our models, and cannot be set aside as "optional" or "free." Second, modeling a full range of variables, not only those that seem phonologically relevant in a traditional sense, is needed. Native speakers deal equally easily with variation triggered by phonotactics as when socioindexical or pragmatic properties are the cause, and as these all contribute to the resultant surface form of an utterance, all these dimensions need to be included in phonological analysis. This requires working with naturalistic speech, in which the range of factors at play is not limited and we hear how people *actually* talk. Third, phonology needs to embrace data that include fine phonetic detail including gradient phenomena, below the level of intuition and IPA transcription — because these "details" are known to be both systematic and relevant to our understanding of competence at the phonological, lexical and grammatical levels. Addressing fine detail typically means working with acoustic, articulatory, or perceptual data, which offer sufficient granularity of data and also have the advantage of reducing observer biases that often occur in transcribed data. From a logistical and methodological standpoint, it is becoming much easier to gather, store, and analyze large amounts of data needed to investigate fine phonetic detail. In some cases, for example corpus studies, the relevant data may already be freely available.

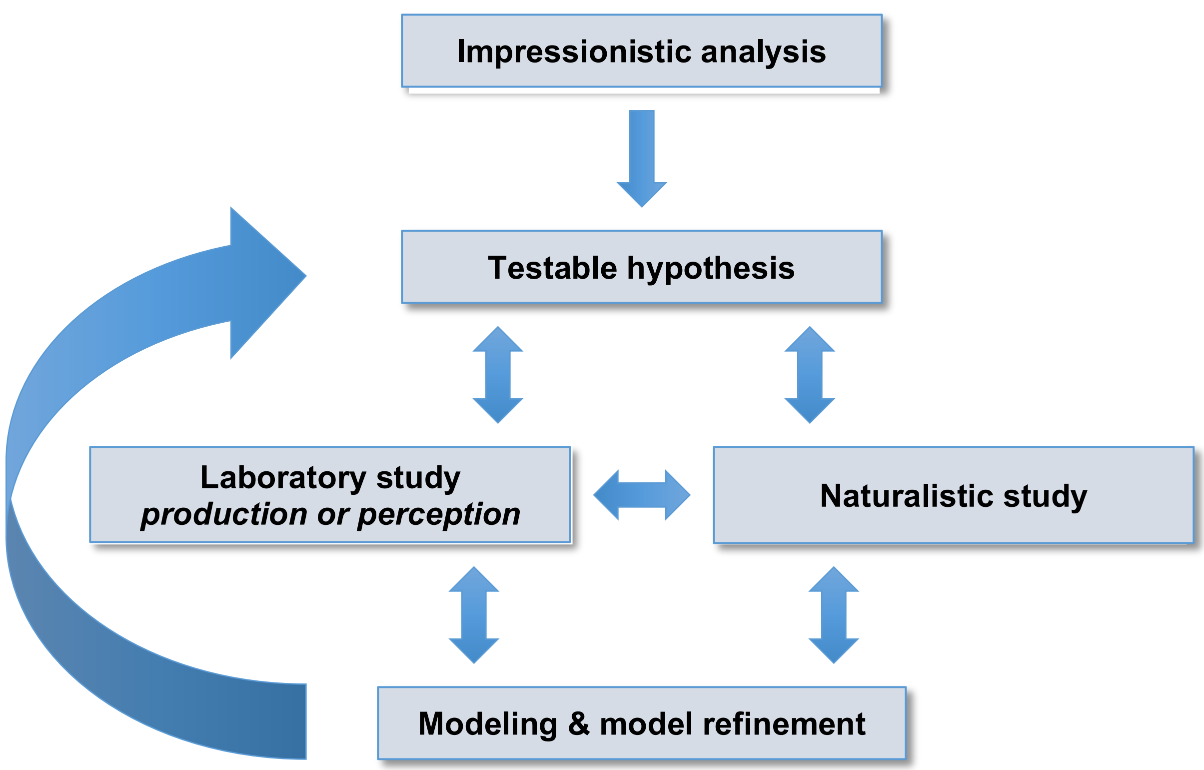

Figure 9. Proposed iterative approach to phonological study and modeling; arrows indicate directions of feedback between studies and methods

Taken together, the case studies in §2 and §3 capture a range of methods, *all of which* we argue remain relevant to the enterprise of phonology. We advocate for an approach to investigating phonology that is methodologically *iterative*, and which incorporates *modeling* not only of phonology as a formal system (in the sense of *SPE* or OT), but which also elucidates the cognitive and communicative processes that are central to speech. This approach is summarized in Figure 9.

Its starting point is the combination of impressionistic observations, phonological intuitions and description that have been a hallmark of phonological analysis. These observations serve to generate a *testable hypothesis* that gives rise to a study with sufficient fine-grainedness, data, metadata, etc.: this could be a laboratory study, or a naturalistic (corpus) study, depending on the nature of the phenomenon and hypothesis under investigation. As indicated by the double-ended arrows in Fig. 9, the development of hypotheses, controlled studies and naturalistic investigations can be iterative. This is illustrated most fully by the Romanian study of marginal contrast, but also by Indonesian, in which a corpus study has been followed up by a controlled study; additionally the study of Catalan mid vowels has led to a wider survey of phonological intuitions, and the next step for investigating Italian mid vowels is to evaluate their realization across a naturalistic, regionally balanced corpus with speaker metadata.

We have not attempted to discuss what kinds of integrated models would account for both the systematic and variable patterns that together constitute the competence of speaker-hearers. We do not delve into the nature of modeling here beyond making the simple point that to address the task of the nature of phonology and speaker-hearer's competence, we need messier data and a fuller toolbox to understand that data. Following a thorough investigation of a phonological phenomenon, building empirically grounded models can test the accuracy and generative capacity of our increasingly rich descriptions. These multifaceted approaches increasingly being adopted in the field lead to enriched and deeper understandings of the nature of phonological competence.

References

Anderson, S.R. 1981. Why phonology isn't "natural." *Linguistic Inquiry* 12(4). 493–539.

Anderson, Stephen R. 1985. *Phonology in the Twentieth Century: Theories of Rules and Theories of Representations*. Chicago: University of Chicago Press.

Anttila, Arto. 1997. *Variation in Finnish phonology and morphology*. Stanford, CA: Stanford University Ph.D. Dissertation.

Anttila, Arto. 2002. Variation and phonological theory. In J. K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 206–243. Malden, MA: Wiley-Blackwell.

Anttila, Arto. 2012. Modeling phonological variation. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology* (Oxford Handbooks in Linguistics), 76–91. Oxford; New York: Oxford University Press.

Anttila, Arto. 2018. Variation in Phonology. *Oxford Research Encyclopedia of Linguistics*. doi:10.1093/acrefore/9780199384655.013.368.

Avery, Peter & Keren Rice. 1989. Segment structure and coronal underspecification. *Phonology* 6(2). 179–200.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bayles, Andrew, Aaron Kaplan & Abby Kaplan. 2016. Inter-and intra-speaker variation in French schwa. *Glossa: a journal of general linguistics* 1(1). 19. doi:http://doi.org/10.5334/gjgl.54.

Bennett, Ryan, Kevin Tang & Juan Ajsivinac Sian. 2018. Statistical and acoustic effects on the perception of stop consonants in Kaqchikel (Mayan). *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 9(1). 9. doi:http://doi.org/10.5334/labphon.100.

Berko, Jean. 1958. The Child's Learning of English Morphology. *WORD* 14(2–3). 150–177. doi:10.1080/00437956.1958.11659661.

Bigi, Brigitte & Daniel Hirst. 2012. SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. *Speech Prosody*, 19–22. Shanghai (China): Tongji University Press. http://www.sppas.org/.

Birdsong, D., L. M. Gertken & M. Amengual. 2012. Bilingual language profile: An easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin*. https://sites.la.utexas.edu/bilingual/.

Boersma, Paul & David Weenink. 2017. *Praat: Doing phonetics by computer [Computer program], Version 6.0.30*. http://www.praat.org.

Bosch, L. F. M. ten, L. W. J. Boves, B. Tucker & M. T. C. Ernestus. 2015. DIANA: towards computational modeling reaction times in lexical decision in North American English. *Proceesings of Interspeech 2015*, 1576–1580. Dresden, Germany.

Bosch, Laura, Albert Costa & Núria Sebastián-Gallés. 2000. First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology* 12(2). 189–221.

Bosch, Laura & Marta Ramon-Casas. 2011. Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics* 39(4). 514–526.

Bosch, Laura & Núria Sebastián-Gallés. 2003. Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech* 46(2–3). 217–243.

Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. In John Kingston & Mary Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376. Cambridge: Cambridge University Press.

Bybee, Joan L. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, 65–85. Stanford, California: CSLI Publications.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: M.I.T. Press.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Clopper, Cynthia G. 2012. Clustering and Classification Methods. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology* (Oxford Handbooks in Linguistics), 678–691. Oxford; New York: Oxford University Press.

Coetzee, Andries W. 2012. Variation: where laboratory and theoretical phonology meet. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology* (Oxford Handbooks in Linguistics), 62–75. Oxford; New York: Oxford University Press.

Coetzee, Andries W. 2016. A comprehensive model of phonological variation: Grammatical and non-grammatical factors in variable nasal place assimilation. *Phonology* 33(2). 211–246. doi:https://doi.org/10.1017/S0952675716000117.

Coetzee, Andries W. & Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language & Linguistic Theory* 31(1). 47–89.

Coetzee, Andries W. & Joe Pater. 2011. The place of variation in phonological theory. In John A. Goldsmith, Jason Riggle & C. L. Alan (eds.), *The Handbook of Phonological Theory*, vol. 75, 401–434. 2nd edn. Oxford: Wiley-Blackwell.

Cohn, Abigail C. 1993. The status of nasalized continuants. In Huffman, Marie K. & Krakow, Rena A. (eds.), *Nasals, Nasalization and the Velum* (Phonetics and Phonology 5), 329–367. San Diego: Academic Press.

Cohn, Abigail C. 2006. Is there Gradient Phonology? In Gisbert Fanselow, Caroline Fry, Ralf Vogel & Matthias Schlesewsky (eds.), *Gradience in Grammar: Generative Perspectives*, 25–44. Oxford: Oxford University Press.

Cohn, Abigail C. 2007. Phonetics in phonology and phonology in phonetics. *Working Papers of the Cornell Phonetics Laboratory* 16. 1–31.

Cohn, Abigail C., Cécile Fougeron & Marie Huffman. 2018. Laboratory Phonology. In Anna Bosch & S. J. Hannahs (eds.), *The Routledge Handbook of Phonological Theory*, 504–529. London: Routledge.

Cohn, Abigail C. & Marie Huffman. 2014. Interface between phonology and phonetics. In Mark Aronoff (ed.), *Oxford Bibliographies Online: Linguistics*. New York: Oxford University Press.

Cohn, Abigail C. & Ferdinan Kurniawan. 2016. Sources of phonological variation: Evidence from a naturalistic corpus of Jakarta Indonesian. Manchester, UK.

Coleman, John. 2002. Phonetic representations in the mental lexicon. In J. Durand & B. Laks (eds.), *Phonetics, Phonology and Cognition*, 96–130. Oxford: Oxford University Press.

Coleman, John. 2011. Phonology as Computation. In John A. Goldsmith, Jason Riggle & C. L. Alan (eds.), *The Handbook of Phonological Theory*, vol. 75, 596–630. 2nd edn. Oxford: Wiley-Blackwell.

Coleman, John, Ladan Baghai-Ravary, John Pybus & Sergio Grau. 2012. Audio BNC: the audio edition of the Spoken British National Corpus. *Phonetics Laboratory, University of Oxford*. http://www.phon.ox.ac.uk/AudioBNC.

Coleman, John, Margaret E. L. Renwick & Rosalind A. M. Temple. 2016. Probabilistic underspecification in nasal place assimilation. *Phonology* 33(3). 425–458. doi:10.1017/S0952675716000208.

DeMauro, Tullio. 2000. *Il dizionario della lingua italiana*. Torino: Paravia.

DiCanio, Christian, Hosung Nam, Jonathan D. Amith, Rey Castillo García & D. H. Whalen. 2015. Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec. *Journal of Phonetics* (The Impact of Stylistic Diversity on Phonetic and Phonological Evidence and Modeling) 48. 45–59. doi:10.1016/j.wocn.2014.10.003.

DiCanio, Christian, Hosung Nam, Douglas H. Whalen, H. Timothy Bunnell, Jonathan D. Amith & Rey Castillo García. 2013. Using automatic alignment to analyze endangered language data: Testing the viability of untrained alignment. *The Journal of the Acoustical Society of America* 134(3). 2235–2246. doi:10.1121/1.4816491.

Docherty, Gerard & Norma Mendoza-Denton. 2012. Speaker-Related Variation–Sociophonetic Factors. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology* (Oxford Handbooks in Linguistics), 43–60. Oxford; New York: Oxford University Press.

DuBois, J. W., W. L. Chafe, C. Meyer & S. A. Thompson. 2000. Santa Barbara Corpus of Spoken American English Part - I (LDC2000S85). Linguistic Data Consortium, Philadelphia.

Ernestus, Mirjam. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142. 27–41.

Ewing, Michael C. 2005. Colloquial Indonesian. In Alexander Adelaar & Nikolaus P Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 227–258. London, New York: Routledge.

Fink, Angela & Matthew Goldrick. 2015. The influence of word retrieval and planning on phonetic variation: Implications for exemplar models. *Linguistics Vanguard* 1(1). 215–225.

Foulkes, Paul. 2010. Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology* 1(1). 5–39.

Foulkes, Paul & Gerard Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34(4). 409–438.

Gil, David & Uri Tadmor. 2015. The MPI-EVA Betawi-Jakarta database. A joint project of the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology and the Center for Language and Culture Studies, Atma Jaya Catholic University.

Goldrick, Matthew, Joseph Keshet, Erin Gustafson, Jordana Heller & Jeremy Needle. 2016. Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition* 149. 31–39. doi:https://doi.org/10.1016/j.cognition.2016.01.002.

Goldsmith, John A. & Bernard Laks. to appear. Generative phonology: its origins, its principles, and its successors. In Linda Waugh, John Joseph & Monique Monville-Burston (eds.), *The Cambridge History of Linguistics*, 1–31. Cambridge: Cambridge University Press.

Gorman, Kyle, Jonathan Howell & Michael Wagner. 2011. Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics* 39(3). 192–193.

Guy, Gregory. 2007. Variation and phonological theory. In R Bayley & C Lucas (eds.), *Sociolinguistic Variation: Theories, Methods and Applications*, 5–23. Cambridge: Cambridge University Press.

Guy, Gregory R. 1997. Violable is variable: Optimality theory and linguistic variation. *Language Variation and Change* 9(3). 333–347.

Guy, Gregory R. 2011. Variability. In Marc van Oostendorp, Colin J Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 1–24. Oxford: Blackwell.

Guy, Gregory R. 2014. Linking usage and grammar: Generative phonology, exemplar theory, and variable rules. *Lingua* 142. 57–65. doi:https://doi.org/10.1016/j.lingua.2012.07.007.

Guy, Gregory R. 2018. Variable grammars. In Naomi L Shin & Daniel Erker (eds.), *Questioning Theoretical Primitives in Linguistic Inquiry: Papers in honor of Ricardo Otheguy*, 45–66. Amsterdam: John Benjamins Publishing Company.

Hall, Kathleen Currie. 2013. A typology of intermediate phonological relationships. *The Linguistic Review* 30(2). 215–275.

Halle, M. 1959. Questions of linguistics. *Il Nuovo Cimento (1955-1965)* 13(2). 494–517. doi:10.1007/BF02724681.

Hannahs, Stephen J. & Anna Bosch. 2018. *The Routledge Handbook of Phonological Theory*. New York, NY: Routledge.

Hay, Jennifer, Janet B. Pierrehumbert & Mary E. Beckman. 2003. Speech perception, well-formedness and the statistics of the lexicon. In John Local, Richard Ogden & Rosalind Temple (eds.), *Papers in Laboratory Phonology VI: Phonetic Interpretation*, 58–74. Cambridge: Cambridge University Press.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4). 458–484.

Hermes, Anne, Doris Mücke & Martine Grice. 2013. Gestural coordination of Italian word-initial clusters: the case of 'impure s.' *Phonology* 30(01). 1–25.

Hulst, Harry van der & Nancy A. Ritter. 2009. The SPE-heritage of Optimality Theory. *The Linguistic Review* 17(2–4). 259–290. doi:10.1515/tlir.2000.17.2-4.259.

Hyman, Larry M. 1975. *Phonology: Theory and Analysis*. New York: Holt, Rinehart and Winston.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullenix (eds.), *Talker variability in speech processing*, 145–165. San Diego: Academic Press.

Johnson, Keith. 2007. Decisions and mechanisms in exemplar-based phonology. In Maria-Josep Solé, Patrice Speeter Beddor & Manjari Ohala (eds.), *Experimental approaches to phonology*, 25–40. Oxford: Oxford University Press.

Kenstowicz, Michael J. & Charles W. Kisseberth. 1979. *Generative phonology: description and theory*. San Diego: Academic Press.

Kisler, Thomas, Uwe D. Reichel, Florian Schiel, Christoph Draxler, Bernhard Jackl & Nina Pörner. 2016. BAS Speech Science Web Services – an Update on Current Developments. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia.

Kurniawan, Ferdinan. 2015. Nasal Assimilation in Jakarta Indonesian. *Austronesian Formal Linguistics Association (AFLA) 21 conference proceedings*, 149–165. Canberra: Australia National University.

Kurniawan, Ferdinan. 2018. *Phonological variation in Jakarta Indonesian: An emerging variety of Indonesian*. Ithaca NY: Cornell University Ph.D. Dissertation.

Labov, William. 1994. *Principles of linguistic change. Volume I: Internal Factors*. Oxford, UK: Blackwell.

Labov, William. 2001. *Principles of linguistic change. Volume II: Social Factors*. Oxford, UK: Blackwell.

Labov, William. 2006. A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics* 34(4). 500–515.

Labov, William. 2010. *Principles of linguistic change. Volume III: Cognitive and Cultural Factors*. Oxford, UK: Blackwell.

Ladd, D. Robert. 2006. "Distinctive phones" in surface representation. In Louis Goldstein, D. H. Whalen & Catherine T. Best (eds.), *Laboratory Phonology 8*, 3–26. Berlin; New York: Mouton de Gruyter.

Ladd, D. Robert. 2011. Phonetics in Phonology. In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The Handbook of Phonological Theory*, 348–373. Oxford: Wiley-Blackwell. doi:10.1002/9781444343069.ch11.

Ladd, D. Robert. 2014. *Simultaneous structure in phonology*. Oxford: Oxford University Press.

Landis, J. Richard & Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1). 159–174.

Lipani, Lisa. 2017. Word-final velar place assimilation in English. *Proceedings of the Linguistic Society of America* 2(24). 1–15. doi:10.3765/plsa.v2i0.4071.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2017. Montreal Forced Aligner: trainable text-speech alignment using Kaldi. *Proceedings of Interspeech 2017*. Stockholm, Sweden.

McAuliffe, Michael, Elias Stengel-Eskin, Michaela Socolof & Morgan Sonderegger. 2017. Polyglot and Speech Corpus Tools: a system for representing, integrating, and querying speech corpora. *Proceedings of Interspeech 2017*, 3887–3891.

Meyerhoff, Miriam. 2011. *Introducing sociolinguistics*. 2nd edn. London: Routledge.

Nadeu, Marianna & Margaret E. L. Renwick. 2016. Variation in the lexical distribution and implementation of phonetically similar phonemes in Catalan. *Journal of Phonetics* 58. 22–47. doi:10.1016/j.wocn.2016.05.003.

Navarra, Jordi, Núria Sebastián-Gallés & Salvador Soto-Faraco. 2005. The perception of second language sounds in early bilinguals: new evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance* 31(5). 912–918.

Pallier, Christophe, Laura Bosch & Núria Sebastián-Gallés. 1997. A limit on behavioral plasticity in speech perception. *Cognition* 64(3). B9–B17.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. Cambridge, MA: Massachusetts Institute of Technology Ph.D. Dissertation.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan L Bybee & Paul J Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 137–157. Amsterdam: J. Benjamins.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7*, 101–140. Berlin; New York: Mouton de Gruyter.

Pierrehumbert, Janet B. 2006. The next toolkit. *Journal of Phonetics* 34(4). 516–530.

Pierrehumbert, Janet B. 2016. Phonological Representation: Beyond Abstract Versus Episodic. *Annual Review of Linguistics* 2(1). 33–52. doi:10.1146/annurev-linguistics-030514-125050.

Pisoni, David B. & Susannah V. Levi. 2007. Some observations on representations and representational specificity in speech perception and spoken word recognition. In Gareth M. Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*, 3–18. Oxford: Oxford University Press.

Prince, Alan & Paul Smolensky. 2004. Optimality Theory: Constraint interaction in generative grammar.

Recasens, Daniel. 1993. *Fonètica i fonologia*. Barcelona, Spain: Enciclopèdia Catalana.

Reddy, Sravana & James N. Stanford. 2015. Toward completely automated vowel extraction: Introducing DARLA. *Linguistics Vanguard* 1(1). 15–28. doi:https://doi.org/10.1515/lingvan-2015-0002.

Renwick, Margaret E. L. 2012. *Vowels of Romanian: Historical, Phonological and Phonetic Studies*. Ithaca, NY: Cornell University Ph.D. Dissertation.

Renwick, Margaret E. L. 2014. *The Phonetics and Phonology of Contrast: The Case of the Romanian Vowel System*. Berlin, Boston: De Gruyter Mouton.

Renwick, Margaret E. L. & D. Robert Ladd. 2016. Phonetic Distinctiveness vs. Lexical Contrastiveness in Non-Robust Phonemic Contrasts. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1). 1–29. doi:10.5334/labphon.17.

Renwick, Margaret E. L. & Marianna Nadeu. 2018. A Survey of Phonological Mid Vowel Intuitions in Central Catalan. *Language and Speech* 1–41. doi:10.1177/0023830917749275.

Renwick, Margaret E. L., Ioana Vasilescu, Camille Dutrey, Lori Lamel & Bianca Vieru. 2016. Marginal Contrast Among Romanian Vowels: Evidence from ASR and Functional Load. *Proceedings of Interspeech 2016*, 2433–2437. San Francisco, CA. http://www.isca-speech.org/archive/Interspeech_2016/pdfs/0762.PDF.

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini & Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite. http://fave.ling.upenn.edu.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris: Payot.

Schiel, Florian. 1999. Automatic phonetic transcription of non-prompted speech. *Proceedings of ICPhS 1999*, 607–610.

Scobbie, James M. 2006. Flexibility in the face of incompatible English VOT systems. In Louis Goldstein, D. H. Whalen & Catherine T. Best (eds.), *Laboratory Phonology 8*, 367–392. Walter de Gruyter.

Scobbie, James M. 2007. Interface and overlap in phonetics and phonology. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press.

Scobbie, James M. & Jane Stuart-Smith. 2008. Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In Peter Avery, B. Elan Dresher & Keren Rice (eds.), *Contrast in Phonology: Theory, Perception, Acquisition*, 87–114. Berlin: de Gruyter.

Scobbie, James M. & Jane Stuart-Smith. 2012. Socially stratified sampling in laboratory-based phonological experimentation. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology* (Oxford Handbooks in Linguistics), 607–620. Oxford; New York: Oxford University Press.

Silverman, Daniel. 2010. Usage-based phonology. In Nancy C Kula, Bert Botma & Kuniya Nasukawa (eds.), *The Bloomsbury Companion to Phonology*, 369–394. London: Bloomsbury.

Solé, Maria-Josep, Patrice Speeter Beddor & Manjari Ohala (eds.). 2007. *Experimental approaches to phonology*. Oxford: Oxford University Press.

Sonderegger, Morgan, Max Bane & Peter Graff. 2017. The medium-term dynamics of accents on reality television. *Language* 93(3). 598–640.

Stuart-Smith, Jane, Eleanor Lawson & James M. Scobbie. 2014. Derhoticisation in Scottish English: a sociphonetic journey. In C Celata & S Calamai (eds.), *Advances in sociophonetics*, vol. 15, 59–96. Amsterdam: John Benjamins.

Stuart-Smith, Jane, Morgan Sonderegger, Tamara Rathcke & Rachel Macdonald. 2015. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology* 6(3–4). 505–549.

Tagliamonte, Sali. 2012. *Variationist sociolinguistics: change, observation, interpretation* (Language in Society). Oxford: Wiley Blackwell.

Temple, Rosalind. 2014. Where and what is (t, d)? A case study in taking a step back in order to advance sociophonetics. In C. Celata & S. Calamai (eds.), *Advances in Sociophonetics*, 97–136. Amsterdam: John Benjamins.

Thomas, Erik R. 2002. Sociophonetic Applications of Speech Perception Experiments. *American Speech* 77(2). 115–147.

Tilsen, Sam & Abigail C. Cohn. 2016. Shared Representations Underlie Metaphonological Judgments and Speech Motor Control. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1)(14). 1–33. doi:http://doi.org/10.5334/labphon.52.

Trubetzkoy, N. S. 1939. *Principles of Phonology*. (Trans.) Christiane Baltaxe. Berkeley: University of California Press.

Vasilescu, Ioana, Margaret E. L. Renwick, Camille Dutrey, Lori Lamel & Bianca Vieru. 2016. Réalisation phonétique et contraste phonologique marginal : une étude automatique des voyelles du roumain. *Actes de la conférence conjointe JEP-TALN-RECITAL 2016*, vol. 1, 597–606. Paris.

Warner, Natasha. 2012. Methods for studying spontaneous speech. In Abigail C. Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology* (Oxford Handbooks in Linguistics), 621–633. Oxford; New York: Oxford University Press.

Wedel, Andrew, Scott Jackson & Abby Kaplan. 2013. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech* 56(3). 395–417.

Wedel, Andrew, Abby Kaplan & Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128(2). 179–186. doi:10.1016/j.cognition.2013.03.002.

Wolfram, Walter A. 1969. A Sociolinguistic Description of Detroit Negro Speech. Urban Language Series, No. 5. http://eric.ed.gov/?id=ED028431.

Young, Steve J., Gunnar Evermann, M. J. F. Gales, D. Kershaw, G. Moore, J. J. Odell, D. G. Ollason, D. Povey, V. Valtchev & P. C. Woodland. 2009. *The HTK book (for HTK Version 3.4)*. http://www.ee.uwa.edu.au/~roberto/research/speech/local/htk/htkbook.pdf.

Yuan, Jiahong & Mark Liberman. 2008. Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America* 123(5). 3878.

Zipf, George Kingsley. 1935. *The psycho-biology of language*. Oxford, England: Houghton, Mifflin.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory* 28(2). 417–472.

Appendix: What is phonology?

To consider how to do *phonology*, we need a working definition of what it is. As a starting point, let's take Hyman's (1975: 2) definition in his textbook *Phonology: Theory and Analysis*: 'Phonology has been defined as the study of sound systems, that is, the study of how speech sounds *structure* and *function* in language.' While Hyman mentions both structure and function in his definition, the generative linguistic approach has focused primarily on structure, often to the exclusion of function, following the tenet of separation between competence and performance (as laid out by Chomsky 1965; and building on the distinction between *langue* and *parole* in de Saussure 1916). Certain other approaches to phonology are more functionally oriented (see Silverman 2010 for discussion).

Central to the study of phonology is the question of what kind of system phonology is. Is it a declarative or procedural system, or both? Again different approaches give different answers to this question. Anderson (1985) frames the history of phonology in the twentieth century in terms of the balance of "rules" and "representations"; theories of phonology can be seen as privileging one over the other. We understand these terms in a broad sense of computation or process, on the one hand, vs. representation or structure (see Goldsmith & Laks to appear for discussion of this point considering the recent history of the field of phonology).

The generative approach to phonology, which has been a dominant view over the past 50 years, was introduced by Chomsky and Halle and explicitly codified in their seminal *The Sound Pattern of English* (1968). *SPE* and subsequent work taking a generative approach aim to answer the question: what does a speaker-hearer know if they "know" the phonology of a language? Chomsky and Halle (1968: 3–4) define the generative enterprise as follows:

> One fundamental factor involved in the speaker-hearer's performance is his knowledge of the grammar that determines an intrinsic connection of sound and meaning for each sentence. We refer to this knowledge—for the most part, obviously, unconscious knowledge—as the speaker-hearer's "competence." . . .
>
> . . . The person who has acquired knowledge of a language has internalized a system of rules that determines sound-meaning connections for indefinitely many sentences.
>
> . . . we use the term "grammar" to refer both to the system of rules represented in the mind of the speaker-hearer, a system which is normally acquired in early childhood and used in the production and interpretation of utterances, and to the theory that the linguist constructs as a hypothesis concerning the actual internalized grammar of the speaker-hearer.

From this perspective *phonology* is the study of such knowledge. To consider this view more concretely, let's take Kenstowicz & Kisseberth's (1979: 25) definition of 'the fundamental question of phonology: What mechanisms and principles must the theory of grammar contain so that the correct phonetic representation can be assigned to the utterances in any human language in such a way as to reflect the native speaker's internalized grammar as closely as possible?'

At this level of discussion, we also understand Optimality Theory to be a generative theory (see van der Hulst & Ritter 2009). That is, the central goal of the theory is to understand what speaker-hearers know, whether modeled as rules and representations (Chomsky & Halle 1968) or violable constraint interaction (Prince & Smolensky 2004).

Central to the generative approach is a modular view whereby phonology is distinct from phonetics and the lexicon. Building on the view laid out in *SPE*, Kenstowicz & Kisseberth (1979: 26) discuss this in terms of the "null" hypothesis under which 'no phonological component of the grammar would be required: The syntactic rules, which specify the order of the morphemes in the sentence, and the lexicon, which specifies the pronunciation of each morpheme, would jointly produce a pronunciation of each sentence.' Evidence against the null hypothesis, that lexical selection and syntactic concatenation are sufficient to produce correct spoken forms, comes from multiple realizations of morphemes ("allomorphs"), particularly cases of productive selection of appropriate allomorphs based on phonological context. Because spoken language provides evidence for productive generalization, Kenstowicz & Kisseberth reject the null hypothesis and argue that forces of phonology, separately from the lexicon and syntax, do come to bear on surface forms.

Key to this argument is evidence from productivity, or as framed in Berko's (1958) classic article "The child's learning of English morphology," *wug* testing:

(1) Berko (1958: 165)

1. Plural. One bird-like animal, then two. "This is a wug /wʌg/.
Now there is another one. There are two of them. There are two ____."

While our methodologies and data have evolved and some of the basic assumptions have been questioned and rethought, 50 years after the publication of *SPE*, the fundamental sense of phonology as a generative system remains at the core of what phonology is.

Department of Linguistics
Cornell University
203 Morrill Hall
159 Central Ave.
Ithaca, New York 14853
acc4@cornell.edu

Department of Linguistics
University of Georgia
240 Gilbert Hall
210 Herty Dr.
Athens GA 30602
mrenwick@uga.edu