Modularity in phonology

Thomas C. Purnell, Department of Linguistics
Eric Raimy, Department of English
Joseph Salmons, Department of German
University of Wisconsin — Madison

Abstract
A fundamental question for linguistic theory is how language is organized in cognition: Is language processed within a single cognitive system or by a set of discrete subsystems? With regard to the study of sound systems, the former position has been pursued in much recent work — 'exemplar' models (for instance, Pierrehumbert 2001, Bybee 2001) and 'classic' Optimality Theory (Prince & Smolensky 2004, McCarthy 2002), while the latter position is that of classic generative phonology (Chomsky & Halle 1968) and 'stratal' OT (Kiparsky 2006). Despite a flood of recent work in linguistics and cognitive science on modular approaches, no coherent advanced introduction to a contemporary modular perspective is available. This book fills the gap with reference to the broader context of speech sounds in linguistics and cognition.

Table of Contents

Estimated length: 75,000 words *en toto*

Introduction

This book introduces the idea of MODULARITY with respect to sound systems and explores it in the broader context of grammar and cognition. Modularity affords an opportunity for the analysis of sound systems to explicitly integrate aspects of morphology, phonetics, psycholinguistics, and sociolinguistics with traditional formal approaches to phonology. By treating each of these topic areas independently, as having their own specific computations and representations, we can produce novel explanations of phonological phenomena in an emergent manner. Treating general sound phenomena as emergent from multiple modules reduces the explanatory burden of each component and allows a deeper and more elegant understanding of sound patterns in language as a whole.

Our assumption is that the readers of this book will have been exposed to the basics of linguistics and phonology by having taken an introduction to linguistics course and an overview course in phonology. As such, we expect some familiarity with terms and concepts introduced by Kenstowicz 1994, Roca & Johnson 1999, Gussenhoven & Jacobs 2005, Odden 2005, Hayes 2008, or similar texts. We further assume that students will have been exposed to basic morphological concepts such as inflection and derivation, as well as having experience with basic descriptive phonetics, e.g., Voice Onset Time. Beyond that level, we will introduce concepts and principles as needed. While we will naturally draw on a broad range of examples from the world's languages, we will rely across chapters on a small set of phenomena, including the laryngeal phonology of English (that is, distinctions like $t \neq d$), variation and change in vowel patterns of contemporary American English (such as the 'Northern Cities Shift and 'low back merger'), and the phonology of inflectional and derivational forms, illustrated with English verbs.

Each chapter introduces a key concept in understanding phonological theory and practice. Chapter 1 introduces the entire book and outlines our goals. Chapter 2 discusses how architectural issues must be considered in developing an analysis of any phonological phenomenon. Chapter 3 provides a discussion of what is meant by a module and the role that interfaces between modules play. Chapters 4-8 present how 'classic' areas of linguistic theory are interpreted in a modular framework, in phonology but also phonetics and morphology. Each chapter presents the structure of relevant modules and the work that a particular module does in the explanation. The bulk of these chapters discuss which parts of phonological phenomena should be accounted for by which module and why this is the case.

Chapters 9-11 survey pressing contemporary issues in phonology from a modular perspective. Each presents topics that are viewed by many as particularly difficult for classic phonological theory, but we show that these areas are actually complementary to and should be integrated into phonological theory. A core aspect of these later chapters is

outlining what parts of these issues are directly 'linguistic' and should be accounted for within a modular linguistic grammar and which parts are 'non-linguistic' and should be accounted for by interaction with general cognitive processes or distinct non-linguistic modules. In earlier research, much effort was invested in arguing that language was built simply on general cognition (in much 'functionalist' research, for instance) or arguing that even details of grammar were 'hard-wired' into Universal Grammar (as in some generative work). In contrast, like an increasing number of researchers, we assume a robust role for both in language, and see it as a goal of linguistics and cognitive science to determine what phenomena belong to which realm and how the two interact.

Chapter 12 concludes the book, drawing together the threads woven through the preceding chapters and providing a coherent overview of the whole model. It also outlines how a modular approach to phonology not only increases our understanding of phonology itself but also allows phonology to make a core contribution to cognitive science in general. Sound patterns in human language provide some of the most robust phenomena that can be investigated in empirical manners which allow scientists to probe the acquisition and use of knowledge.

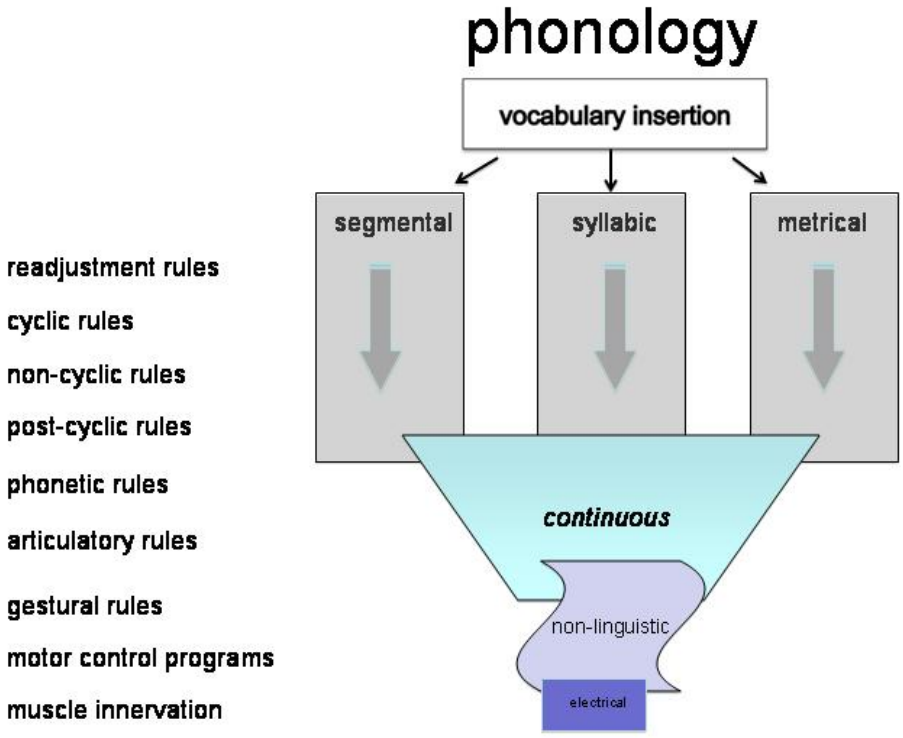
Chapter synopses

1.0 Introduction: Parsimony in phonology

Phonology is often defined as the scientific study of sound patterns in human language. This definition, while overbroad in some sense, is beneficial because it covers all topics having to do with sound patterns of human language, such as different branches of phonetics (the study of the acoustics, articulation, and perception of human speech), phonology (in a narrow sense), and sociophonetics. Nevertheless, the goal of this chapter is to argue that an inclusive definition of the broader notion of phonology does not imply a loss of boundaries among the subparts (Ohala 2005). In contrast to much recent research, we argue for a more nuanced and complex architecture for investigating how cognitive processes operate over sound units. Defining these subareas of phonology broadly conceived provides a good starting point for initial hypotheses about what modules exist.

Figure 1 below presents a general picture of the architecture for phonology that we will investigate in the book. A key aspect of this model is distinct and multiple phonology-internal modules which allows for different sources of explanation to be developed for different phenomena. Our approach relies on both serial (ordered) and parallel processing within the overall system. This is in sharp contrast to the far coarser debates that have raged over the last two decades about monostratal versus serial approaches in phonology, where much work in 'classic' Optimality Theory has insisted that all of phonology and even phonetics and morphology are all processed in parallel, whereas early generative approaches were simply serial (see §2.0).

Figure 1: A hypermodular view of the speech chain



In order to familiarize the reader with this overall architecture and the information flow through the multiple submodules in Figure 1, we will guide the reader through a discussion of adjacency, or what happens when sounds are next to each other. Adjacency of sounds is important for each of part of the process of converting morphosyntactic features into acoustic landmarks that are perceptible. We do not assume here a position where phonology occurs in the lexicon and phonologically specified morphemes are manipulated by the syntax (as, for instance, in Kiparsky's Lexical Phonology), but adopt a position closer to Halle & Marantz's Distributed Morphology, where morphosyntactic features retrieve sound shapes to be used in the processing stream. Such vocabulary insertion, then, involves linearization or the placement of sounds in a temporal sequence and may trigger readjustments of items in that linear sequence (e.g., English ablaut). Much of the following phase entails traditional accounts of derivational phonology (that is, where cyclic rules → non-cyclic rules → post-cyclic rules) with the exceptions that (a) syllabification, stress computation and feature manipulation occur partly in parallel and (b) information from these independent submodules is shared across those boundaries. Consequently, the aspect of phonology as the regulator of function sequence (that is, rule ordering) arises from the regulation of information exchange between the segmental, syllabic and metrical submodules in Figure 1.

Suppose, for example, that an American English 'r-dropping' speaker is generating a number of word forms that include the lexeme TO TEAR ('to rip', as in *tear*, *tears*, *tearing*, *torn*, etc.).[1] If the morphosyntactic features trigger a readjustment for the past participle form, then ablaut occurs. If the features are for third person singular agreement, then /z/ is added. If the speaker belongs to a speech community such as African American English (AAE), the /z/ may be blocked because speakers draw on information outside formal phonology regarding speech register appropriateness (e.g., by familiarity and perceived ethnicity, Baugh 1988). Adding to the past participle and third person forms at this early phase, if the morphosyntactic features indicate the progressive aspect marker, then /-ɪŋ/ or /-ɪn/ morphemes are added, again with the choice of variant mediated by sociolinguistic factors of formality and social class (Campbell-Kibler 2006). Throughout this processing phase, syllabification is going to generate forms where /ɹ/ is postvocalic in a cluster (*tears*, *torn*), postvocalic at the end of the word (the AAE form and the 'do nothing' default *tear*) or postvocalic but a syllable onset (*tearing*, *tearin'*). The rule of /ɹ/-deletion for the 'r-dropping' dialect speaker is the product of the segmental submodule and must be coordinated (which is to say, temporally timed) to occur after syllabification in order to prevent deletion from applying to the progressive forms. Moreover, if the 'r-dropping' dialect also involves the process of breaking (or /ɹ/-vocalization, [t^{h}ɪə], as in Sledd 1966), then the function of breaking has to be regulated to occur before the function of /ɹ/-deletion. The progressive forms and cluster forms reflect the issue of adjacency at this stage: /ɹ/ being followed by a vowel in the former case, and between a vowel and another consonant in the latter.

As the sound shapes move through this process towards articulation, further adjacency changes take place. For one, if a speaker belongs to a speech community where final obstruent 'devoicing' (or fortition, we will argue) occurs in the third person agreement form (effectively *tear*[s] for *tear*[z], Purnell et al. 2005, Iverson & Salmons forthcoming), then phonetic implementation occurs in the interface between the output of traditional phonology and preparation of the gestural score. This implementation, like vocabulary insertion or rule operation, can be mediated by the cognitive interface with social rules, and provides evidence of ongoing interactions between the larger phonological processes and socio-cultural information. A second adjacency process that may occur depending on articulation rate with the result of the word final velar articulation is made in the region of the alveolar ridge, particularly if the following word begins with an alveolar (e.g., as in *teari*[n] *two strips*). Because this coarticulation effectively mirrors the formality- and social-class-shaped choice of /-ɪn/ mentioned earlier, we use this to highlight to the reader apparent systemic ambiguities.

[1] Discussion of the paradigm of *to tear* should not be interpreted as the speaker making reference to the paradigm of the word (e.g., Blevins 2006) and is only employed here as an explanatory tool. Also note that the full manuscript will lay out the technical distinction between 'lexeme' and 'morpheme.'

An advantage of this neo-generative approach is that different sources of explanation combine to create 'emergent' explanations of more complex phonological data. Our understanding of phonology in general is increased when we can posit simple explanations for different observations which when combined account for complex observations in a straightforward manner, as seen in the various derivations involving TO TEAR above. As such, we will use the term Hypermodular Phonology (HMP) to refer to the general modular sound system. The term FORMAL PHONOLOGY refers to phonological aspects of processing and regulating sounds which correspond to the parallel segmental, syllabic and metrical paths in Figure 1. We recognize, however, that invoking modularity in analyses is a responsibility and cannot be done haphazardly. What modules are present in the brain is an empirical question and linguistic data is only one line of evidence in answering this question. Parsimony of analysis must be determined across the whole system and not on isolated, narrowly defined problems. We argue that linguists can only determine whether a modular analysis is required by developing such analyses and then evaluating the quality of the resulting explanations.

2.0 Architecture in hypermodular phonology

The question of architecture in HMP is basically the question of how many modules exist, what the modules are and how the modules interact. The two dominant views of formal phonology in the last 50 years provide opposing views on this question. *The sound pattern of English* (SPE, Chomsky & Halle 1968) represents one pole of the architecture scale in that sequential modules feed representations into each other in a 'top down' fashion. Optimality Theory (OT, Prince & Smolensky 2004, McCarthy & Prince 1995) provides an example on the opposite side of the scale where only a single module is posited for all sound related processes (a position now abandoned in McCarthy 2008). The main distinction between these two models is in the flow of information and what type of information is available to the formal phonological mechanism.

One commonality between the SPE and OT models of phonology is that they are both *single processor* models. A single processor model does not allow for independent computation of different aspects of the domain of operations, in this case sounds. SPE does not allow for variable computation because of its strict serial modular architecture. OT, too, does not accommodate variable computation because its 'parallelism' is in the informational sense. On modular perspectives in general, one of the few recent explicit discussions is found in *Phonetically-based Phonology* (Hayes, Kirchner & Steriade 2004), and it represents a view dramatically different from that developed in the present volume. In fact, both SPE and this new model of Hayes *et al.* fail to take advantage of parallel distributed processing.

We develop the position that HMP must include both a regulatory and procedural aspect. The regulatory dynamic manages information flow among parallel and serial processes, while the procedural dynamic behaves like a bank of functions operating over

a narrow set of elements, including for example distinctive features, timing nodes, metrical grids. Parametric variation in the regulation and operation of phonology accounts for the variety of sound systems.

In short, we argue that phonological theory can benefit by investigating not only modularity per se, but what aspects of HMP are dependent on or independent of each other. One likely set of candidates for parallel separation and processing are segmental and intonational phonology. Intonational phonology deals with the overall pitch patterns in sentences and is likely calculated strictly off of the syntactic representation prior to vocabulary insertion. This allows the intonational domains to be calculated separately from the segmental ones. A new aspect of architecture under this view is that different streams of computation and representation must be integrated at a distinct point in the overall calculation.

3.0 What is a module?

We define a MODULE as a distinct set of representations and computations that support a unique task. This narrow formal definition of a module suits the purposes of the demonstrations in this book. We also interpret this definition of MODULE in two distinct ways: cognitively and descriptively. The utility of this type of definition is demonstrated in chapters 4-8, where we use it to account for the broadest possible range of sound patterns within a single overarching framework.

Fodor (1983) provides the classic definition of modularity in cognitive processes. We adopt a neo-Fodorian version of modularity as a basis for our general modular approach (cf. Coltheart 1999). This way of interpreting modularity is cognitive in our view in that it makes particular claims about actual aspects of the cognitive system which supports phonology. In contrast to this view of modularity, there is also the descriptive use of modularity. We explicate how Marr's (1982) view of cognitive science is a useful form of descriptive modularity. Marr's core proposal about cognitive science is that computational, algorithmic and implementation levels of analysis can and should be done separately. Only by having independent explanations of a cognitive module at the computational, algorithmic and implementation levels is a complete understanding achieved. The application of these notions to sound systems has barely been explored to date.

An important but often misunderstood notion about modularity is that modules are closed. Rather than assuming that there is a single phonological module that is closed off from interaction with information between submodules and outside formal phonology, the derivations of TO TEAR (above) suggest the existence of a set of smaller modules that interact frequently and can operate in some cases in a serial and in others in a parallel manner (see Jackendoff 1997). Our understanding of a closed module does not exclude interactions between submodules.

In this chapter, we outline our understanding of intra-phonology modularity or HMP. Specifically, there is sufficient evidence that segmental phonology, syllabic phonology and metrical phonology are computed in separate processing streams (see again Figure 1), which are then integrated at different points in the derivation. This claim aligns HMP with Carruther's (2008) claim of massive modularity in the mind and much recent work on biological sytems (Gilbert, Opitz & Raf 1996, Klingenberg 2008).

4.0 Morphology

Morphology is the linguistic module which acts as an interchange between syntactic and phonological representations (Halle & Marantz 1993), and in many ways is the most controversial module within generative linguistics. On our view, the morphology module is responsible for three main things on the phonological side: the encyclopedia of vocabulary items which are the phonological representations for morphemes, the establishment of thematic classes among morphemes and how vocabulary items are concatenated to each other according to syntactic structure.

The encyclopedic aspect of morphology is that it contains all of the vocabulary items for a given grammar. Vocabulary items are the phonological aspects of memorized morphemes and serve as the underlying representation which begins a modular derivation. An aspect of this encyclopedia is that any exceptions or 'minor rules' that a morpheme is subject to is also stored so must be encoded in a phonological representation in some manner (Lightner 1974, Halle & Nevins 2009).

In order to begin demonstrating a modular analysis of a complete 'phonological derivation', we provide a discussion of regular and irregular verb forms in English based on proposals in Yang (2002). This discussion will focus on affix ordering in verbs, the HMP representation of the past tense suffix in English and how readjustment rules account for irregular verb forms. The function of the morphology is to create a representation for the formal phonology module to operate on.

5.0 Cyclic Phonology

The cyclic phonology module has unique properties because it demonstrates aspects of parallel processing, as we will explore in detail in the full manuscript. Classic 'cyclicity effects' are produced by each cyclic morpheme initiating a new processing stream. These parallel processing streams interact in a limited way and produce the overwriting of old cycles by new cycles.

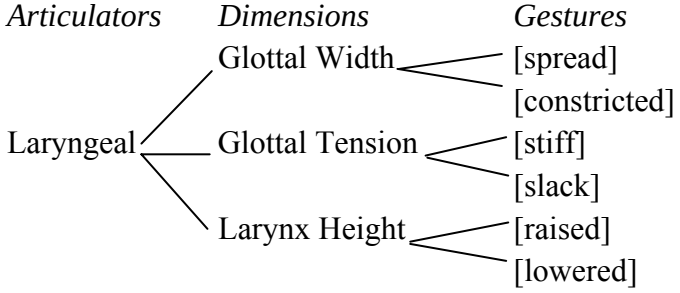
Two main demonstrations of cyclic effects will be presented in this chapter. The first is in the segmental phonology of the past tense suffix in English. There are three realizations of the past tense suffix: /d/, /t/ and Ø. These three suffixes behave differently in two ways; first, /t/ but not /d/ or Ø triggers cyclic shortening of the vowel in the verb and two, the suffix /t/ triggers laryngeal assimilation (e.g. *leave* ~ *lef-t*) while /d/ and Ø do not. These distinct behaviors of the different exponents of the past tense morpheme in

English provide evidence that cyclic vs. non-cyclic behavior of an affix has to be encoded on vocabulary items themselves and not morphemes.

A 'devoicing' pattern associated with the /t/ realization of the past tense in certain varieties of English begins discussion of the nature of distinctive features in phonological representations. We present the Toronto School of Contrast approach to distinctive features which places contrast as determined by phonological activity as the determining factor in representation. Since the /t/ triggers laryngeal assimilation in the cyclic rules (i.e. it is phonologically active), it must be marked while the /d/ will be unmarked (because it is phonologically inactive).

At the same time, we also adopt the position of 'laryngeal realism', as proposed by Iverson & Salmons 1995 and many others, and as crucially developed by Avery & Idsardi 2001. On this view, the distinction between /t/ and /d/ in languages like English is carried by a phonological specification on the fortis (or traditional 'voiceless') member of the pair while the other member of the pair is phonologically unspecified. As illustrated in Figure 2 below, this is captured by the laryngeal 'dimension' of Glottal Width, or 'GW', corresponding to the gestures of spread or constricted glottis, the articulatory means of instantiating distinctions within a dimension. This contrasts with true 'voicing' languages like French or Spanish, which phonologically specify the voiced member of pairs rather than the voiceless (so, *d* is marked and *t* is not, laryngeally). In such languages, the dimension of Glottal Tension includes the gesture slack vocal folds, producing voicing.

Figure 2: Geometry of Laryngeal Representation in Dimensional Theory (Avery & Idsardi 2001).



This simple system allows the full known cross-linguistic typology shown in Figure 3, where the empty brackets reflect the laryngeally unspecified series.

Figure 3: Laryngeal typology (Iverson & Salmons 2003, Avery & Idsardi 2001)

	/p ~ b›/	/b/	/p^{h}/	/p'/	/b^{h}/
Hawaiian	[ ]				
K'ekchi	[ ]			[constr]	
Spanish	[ ]	[voice]			
English	[ ]		[spread]		
Thai	[ ]	[voice]	[spread]		
Hindi	[ ]	[voice]	[spread]		[spread] [voice]

We will introduce Avery & Idsardi's Dimensional Theory in this chapter and develop it further in later chapters, highlighting its important ramifications for phonetics, motor control and sound change in particular.

6.0 Non-cyclic Phonology

The non-cyclic phonology module begins with the last cyclic processing stream that works on the representation. Processing at this point in the derivation is narrowed to a single processor for each of the relevant phonological modules (e.g. segmental, syllabic, metrical). The narrowing of the processing stream is one reason why non-cyclic rules are more surface true.

Further aspects of the past tense forms of English verbs and stress in English are investigated. For the past tense forms, we will show that there is an additional non-cyclic laryngeal assimilation process in English. The cyclic and non-cyclic versions of the laryngeal assimilation rule can be distinguished by the direction of their operation (cyclic = progressive and non-cyclic = regressive) and the phonological activity of these vocabulary items in other processes (e.g. cyclic vowel shortening). This particular example demonstrates a case where 'false parsimony' can be avoided, as argued by Salmons (forthcoming). It might be tempting to try to collapse the cyclic and non-cyclic versions of the laryngeal assimilation rule but it would obscure the different phonological activeness of the different past tense vocabulary items and unnecessarily complicate the laryngeal assimilation rule. As already noted above, we argue that simplicity of generalizations must be evaluated across the whole system and not locally.

The non-cyclic aspect of English stress will also demonstrate that morphological theme classes must persist into this module. There are two theme classes based on the distribution of secondary stress in English, providing clear support for the need for this. We will review the claims made in Lexical Phonology about the characteristics of different types of processes (i.e. lexical vs. post-lexical) and map them to the cyclic vs. non-cyclic model in this book.

7.0 Phonetics

The non-cyclic phonology produces impoverished representations that must be further specified by the phonetics interface between formal phonology and motor control. Inversely, there is the question of how gradient contrasts in acoustic representations are converted into discrete phonological features. To develop these points, we will further investigate the distinction between English /t/ and /d/ from a more directly phonetic point of view.

The phonemes /t/ and /d/ in English are acoustically distinct from each other in different ways depending on context. For most varieties of American English, word initial position in stressed syllables is the most straightforward context, where /t/ and /d/ can be contrasted simply in voice onset time: /t/ typically shows considerable aspiration, ca. 60-70 milliseconds (Weismer 1979). In other variable contexts like location (word medial vs. word final, onset vs. coda, cluster vs. simplex, etc.) and metrical status (e.g. in a stressed syllable vs. in an unstressed syllable), the contrast between /t/ and /d/ is more complex. We adopt the idea of trading relations (Repp 1982) between different acoustic parameters that facilitate the preservation of the consistent laryngeal contrast between /t/ and /d/. These more complex mappings between acoustic parameters and phonological features demonstrate that trading relations provide a useful tool in approaching the goal of invariance in phonological features.

The utility of trading relations is further supported by an examination of the relationship between vowel length and syllable shape in English. There are multiple levels of phonetic length in a vowel based on the rhyme shape of the syllable that the vowel occurs in. Open syllables with no codas are longest, then syllables closed with a sonorant, then syllables closed with a voiced obstruent and finally vowels in syllables closed with a voiceless obstruent are shortest. We connect these phonetic observations with the nature of laryngeal specification and phonological features in English.

8.0 Motor-control systems

The phonetics module is a grammatical interface module producing representations that must be further converted to representations interpretable by the motor control system. We demonstrate the utility and importance of the gestural score approach proposed by Browman & Goldstein (1989, 1992) by showing how they allow phonological invariance in the presence of phonetic variation. We will also consider more recent models of motor control, like DIVA (Directions Into Velocities of Articulators), cf. Guenther (2006) and Guenther et al. (2006).

Analyzing fast speech effects demonstrates how the acoustic signal may not overtly support a phonological contrast but the gestural score does contain the contrast. By manipulating the speed of the running of the gestural score, fast speech effects can be produced (Abbs 1986, Fujimura 1987). Simply put, if the overall time of a score does not

allow for all gestures to be executed fully, then overlap and/or undershoot of gestures will produce the acoustic effect of deletion.

We exemplify this type of effect by discussing the analysis of 'deaspiration' of English voiceless stops in unstressed syllables. These stops are not acoustically 'aspirated' but they are acoustically distinct from the corresponding laryngeally unmarked stops. For the unaspirated stops, there is still a [spread] gesture from the Glottal Width dimension which supports the phonological laryngeal contrast, but the different timing of the gesture creates the 'aspirated' vs. 'non-aspirated' acoustic contrast.

9.0 Lexical access

Formal phonology — in the specific sense of hypermodular phonology — is a theory of the representations that support lexical access (Poeppel, Idsardi & van Wassenhove 2007). This claim may seem problematic when speaker specific priming effects are considered. Some researchers suggest from these fine-grained acoustic priming effects that there is no role or utility for formal phonology (e.g., Port & Leary 2005). This type of argumentation is another case of 'false parsimony' because it does not recognize modular, representational and computational aspects of phonology in general.

We introduce the reader to the key findings on how lexical access can be manipulated along different phonological and acoustic parameters. There is, in particular, a conflict in the findings on how lexical access can be affected. One line of research clearly demonstrates that fine details of acoustics can affect lexical access reaction times (Kouider & Dupoux 2009). The other line of research clearly demonstrates that the stored representations must be abstract in nature to account for variation (McQueen, Cutler & Norris 2006, Mirman, McClelland & Holt 2006). A third line of research demonstrates that abstract non-acoustically grounded structures such as syllables and templates can be primed (Boudelaa & Marslen-Wilson 2001, 2004).

A model of phonology which has at least the modules presented in chapters 4-8, we argue, is capable of reconciling the conflicts in the literature on lexical access. We map the different findings about the nature of lexical access to different modules which allows a coherent picture of the lexical access literature from a phonological perspective to emerge.

10.0 Sociolinguistics

Sociolinguistics can be defined for present purposes as the study of language variation that has at most a slight impedance on comprehension and which signals distinctions based on (a) perceived kinship, commensality and a shared belief system, (b) register choices within a discourse setting, or (c) effects of the larger linguistic milieu in which a speaker is a member. It is another area of research that presents variation which at first glance suggests that formal phonology is inadequate to understand phonology in general.

Similar to chapter 9, we demonstrate how sociophonetic variation reinforces HMP as it has been explicated in earlier chapters.

The formal phonology of a given speech variety (dialect or language) identifies the parameters which can be used to mark group membership such as social class. The example of r-dropping in TO TEAR (above) already illustrates interaction between structural and social factors. The classic pattern of rhotic loss is in codas, but the generalization extends saliently to intervocalic position in some communities, so that both *tear* and *tearin*' are r-less. We will further demonstrate this idea by presenting the current situation of variation in low vowels and VOT times of stops in American English.

Great dialectal variation is observed in the number of phonetic targets of low/back vowels in English, /a, ɔ/, often exemplified by the pair 'cot' and 'caught' (Labov et al. 2006). Phonological merger allows for the phonetic target of the low back vowel to be exploited by speakers sociolinguistically in various ways. For example, not all merged dialects have the same phonetic target for this vowel (Labov 1994).

'Ash raising' is another example of sociolinguistic marking via acoustics. In the Upper Midwest, the raised /æ/ acts as a geographical marker. The actual phonetics of this vowel raising pattern may involve adjustments by the lips, pharynx or larynx, and not an actual manipulation of vowel height. Rounding, ATR and Larynx Height (Purnell 2008, Stevens 1998) are not contrastive in the vowel system in English and thus are available gestures for sociolinguistic control.

A final example is drawn from the social variation in obstruent voicing in English. As already indicated, the laryngeal phonology of English identifies the 'voiceless' series as being marked with Glottal Width. This allows the unmarked category of 'voiced' sounds to vary in their actual phonetics. This holds true in that there are dialects of English where the phonetics of voiced stops in word initial position are 'prevoiced' and other dialects of English have voiced stops in word initial position as 'short lag'. The different phonological representations again allow for sociolinguistic variation.

11.0 Sound change

Language transmission occurs as infants construct an internal grammar from the external language they are exposed to. This means that the perspective on variation from any of the sources we have discussed in the previous chapters is distinct for the competent adult language user vs. an infant engaged in the process of language acquisition. While much of the particular shape of synchronic sound patterns reflects historical inheritance (as argued most forcefully by Blevins 2004), our understanding of sound change can be greatly advanced by a modular approach, including a more robust abstract phonological component.

Variation provides ways in which language can change as new generations of learners analyze the variation in new ways, such as extending existing phonological generalizations. In our earlier example of r-dropping, learners and speakers first

vocalized rhotics in codas, while later generations expanded the generalization to intervocalic position. However, things are often far less straightforward. For instance, the historical origin of /æ/ raising may lie in lip rounding and it may have spread thanks to an emerging association with social identity, but it has now come to affect the shape of the vowel system in English. Children acquiring language in these communities are producing vowel chains where the raised ash vowel is now affecting the mid front vowels.

Another example where existing variation within English provides a path to language change is the variation in VOT for voiced stops. As already noted, English, like most other Germanic languages, possesses a laryngeal system contrasting Glottal Width (GW) with unmarked obstruents, GW vs. Ø. Even within the Germanic family, there are documented cases of language change where a GW system has evolved into a voicing system (i.e. Glottal Tension, GT). Dutch and Yiddish have both undergone such changes, and we discuss Pennsylvania Dutch as a language that is presently at the tipping point between a GW and GT laryngeal system which may be the crucial developmental stage between the GW system of English/German and the GT system of Dutch. Again, the formal system allows the change to be clearly identified and questioned.

12.0 Conclusion: Linguistics as Cognitive Science

The final chapter will begin with a sketch of the big picture of hypermodularity, tying together the key threads of the preceding chapters in terms of the overall model presented at the outset, in Chapter 1 and Figure 1.

From there, this chapter situates the general endeavor of HMP as the study of sound patterns as a sub-discipline of cognitive science. Cognitive science is the investigation of knowledge and the use of knowledge in a cognitive system. There are two main aspects of this grounding that must be understood.

The first is that HMP is a science that investigates human nature. This situation is why it is difficult to pigeonhole linguistics into one of the classical divisions of natural sciences, social sciences and humanities. HMP benefits from many different methodologies from these different areas and does not eschew contributions from any of these areas. Only by being aware of different results and research questions in different areas can formal phonology benefit from them. Furthermore, the researcher must evaluate the quality and relevance of any of these external findings.

The second aspect of modular grounding is to recognize the importance of the general scientific method. All contemporary and past knowledge about any aspect of phonology can be demonstrated to be incorrect through falsification. In recognizing this fact, the question is how to move forward once this point is recognized. Phonological theorizing, we maintain, has matured to the point that we can put false parsimony behind us. The HMP approach presented in this book argues that the question of which module should account for a particular phonological observation should be the first one asked. If the

wrong module is chosen for the source of explanation then a false negative will be produced. Only by pairing the right potential source of explanation with a given phenomenon will we better understand phonology in general. Moreover, only by considering a modular approach to phonology which allows for many different potential sources of explanation can we even ask the question of whether we have the correct source of explanation.

References

Abbs, James H. 1986. "Invariance and variability in speech production: a distinction between linguistic intent and its neuromotor implementation." Joseph S. Perkell and Dennis H. Klatt (eds.) *Invariance and variability in speech processes*. Hillsdale, NJ:Lawrence Erlbaum Associates, Publishers, pp. 202-218.

Avery, Peter & William Idsardi. 2001. "Laryngeal dimensions, completion and enhancement". *Distinctive Feature Theory*, ed. Tracy Alan Hall, pp. 41-70. Berlin: Mouton de Gruyter.

Baugh, John. 1988. *Black Street Speech: Its history, structure and survival*. Austin, TX: University of Texas Press.

Blevins, James. P. 2006. "Word-based morphology." *Journal of Linguistics* 42.531-573.

Blevins, Juliette. 2004. *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Boudelaa, Sami & William D. Marslen-Wilson 2001. "Morphological units in the Arabic mental lexicon." *Cognition* 81.65-92.

Boudelaa, Sami & William D. Marslen-Wilson 2004. "Abstract morphemes and lexical representation: The CV-skeleton in Arabic." *Cognition* 92.271-303.

Browman, Catherine P. & Louis Goldstein. 1989. "Articulatory gestures as phonological units". *Phonology* 6.201-251.

Browman, Catherine P. & Louis Goldstein. 1992. "Articulatory phonology: An overview." *Phonetica* 49.155-180.

Bybee, Joan. 2001. *Phonology and Language Use.* Cambridge: Cambridge University Press.

Campbell-Kibler, Kathryn. 2006. *Listener Perceptions of Sociolinguistic Variables: The case of* (ING). Ph.D. dissertation, Stanford University.

Carruthers, Peter. 2008. Précis of *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. *Mind & Language* 23.257-262.

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

Coltheart, Max. 1999. "Modularity and cognition." *Trends in Cognitive Science* 3.115-120.

Fodor, Jerry A. 1983. *The modularity of mind: An essay on faculty psychology.*

Cambridge: MIT Press.
Fujimura, Osamu. 1986. "Relative invariance of articulatory movements: an iceberg model". Joseph S. Perkell & Dennis H. Klatt (eds.) *Invariance and variability in speech processes*, pp. 226-233 Hillsdale, NJ: Lawrence Erlbaum Associates.
Gilbert, Scott, Opitz, John & Rudolf Raff. 1996. Resynthesizing evolutionary and developmental biology. *Developmental Biology* 173.357-372.
Guenther, Frank H. 2006. "Cortical interactions underlying the production of speech sounds." *Journal of Communication Disorders* 39.350-365.
Guenther, F.H., Ghosh, S.S., & Tourville, J.A. 2006. "Neural modeling and imaging of the cortical interactions underlying syllable production." *Brain and Language* 96.280-301.
Gussenhoven, Carlos & Haike Jacobs. 2005. *Understanding Phonology.* London: Arnold. 2nd edition.
Halle, Morris & Alec Marantz. 1993. "Distributed morphology and the pieces of inflection." Kenneth Hale & Samuel J. Keyser (eds.), *The view from Building 20*, pp. 111–176. Cambridge, MA: MIT Press.
Halle, Morris & Andrew Nevins. 2009. "Rule application in phonology." Eric Raimy & Charles Cairns. (eds.), *Contemporary views on architecture and representations in phonology*, pp. 355-382. Cambridge, MA:MIT Press.
Hayes, Bruce. 2008. *Introductory Phonology*. Oxford: Blackwell.
Hayes, Bruce, Robert Kirchner & Donca Steriade. 2004. *Phonetically-based Phonology*. Cambridge: Cambridge University Press.
Iverson, Gregory K. & Joseph C. Salmons. 1995. "Aspiration and laryngeal representation in Germanic." *Phonology* 12.369-396.
Iverson, Gregory K. & Joseph C. Salmons. 2003. "Laryngeal Enhancement in Early Germanic." *Phonology* 20.43-72.
Iverson, Gregory K. & Joseph C. Salmons. Forthcoming. "Final Devoicing and Laryngeal Neutralization." *Companion to Phonology*, ed. by Marc van Oostendorp, Colin Ewen, Beth Hume & Keren Rice. Oxford: Blackwell.
Jackendoff, Ray. 1997. *The Architecture of the Language Faculty.* Cambridge, MA: MIT Press.
Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.
Kiparsky, Paul. 2006. *Paradigms and Opacity*. Chicago: University of Chicago Press.
Klingenberg, Christian. 2008. Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution, and Systematics* 39.115-132.
Kouider, Sid & Emmanuel Dupoux. 2009. "Episodic accessibility and morphological processing: Evidence from long-term auditory priming." *Acta Psychologica* 130.38-47.
Labov, William. 1994. *Principles of Linguistic Change.* Vol. 1: *Internal factors*. Oxford: Blackwell.

Labov, William, Sharon Ash, & Charles Boberg. 2006. *Atlas of North American English: Phonetics, phonology, and sound change*. Berlin: Mouton de Gruyter.

Lightner, Theodor. 1974. *Problems in the theory of phonology*: vol. 1, *Russian phonology and Turkish phonology*. Edmonton: Linguistic Research.

Marr, David. 1982. *Vision*. San Fransisco: W. H. Freeman & Co.

McCarthy, John J. 2002. *A Thematic Guide to Optimality Theory*. Cambridge, UK: Cambridge University Press.

McCarthy, John J. 2008. "The serial interaction of stress and syncope." *Natural Language & Linguistic Theory* 26.499-546.

McCarthy, John J. & Alan Prince. 1995. "Faithfulness and reduplicative identity." Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) *Papers in Optimality Theory*, pp. 249–384. Amherst : GLSA.

McQueen, J. M., Cutler, A., & Norris, D. 2006. "Phonological abstraction in the mental lexicon." *Cognitive Science* 30(6).1113-1126.

Mirman, D., McClelland, J. L., & Holt, L. L. 2006. "An interactive Hebbian account of lexically guided tuning of speech perception." *Psychonomic Bulletin & Review,* 13(6).958-965.

Odden, David. 2005. *Introducing Phonology*. Cambridge: Cambridge University Press.

Ohala, John. 2005. "Phonetic explanations for sound patterns: Implications for grammars of competence." W. J. Hardcastle and J. M. Beck (eds.) *A figure of speech: A festschrift for John Laver,* pp. 23-38. London:Erlbaum.

Pierrehumbert, Janet. 2001. "Exemplar dynamics: Word frequency, lenition, and contrast." Joan Bybee & Paul Hopper (eds.) *Frequency effects and the emergence of lexical structure*, pp. 137-157. Amsterdam: John Benjamins.

Poeppel, David, Idsardi, William & van Wassenhove, Virginie. 2007. "Speech perception at the interface of neurobiology and linguistics." *Proceedings of the Royal Society of London* doi:10.1098/rstb.2007.2160.

Port, Robert & Adam Leary. 2005. "Against formal phonology." *Language* 81(4).927-964.

Prince, Alan S. & Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden: Blackwell. (Originally 1993, Technical Report 2, Rutgers Center for Cognitive Science.)

Purnell, Thomas C. 2008. "Pre-velar raising and phonetic conditioning: Role of labial and anterior tongue gestures." *American Speech* 83.373-402.

Purnell, Thomas C., Salmons, Joseph C., Tepeli, Dilara, & Mercer, Jennifer. 2005. "Structured heterogeneity and change in laryngeal phonetics: Upper Midwestern final obstruents." *Journal of English Linguistics* 33.307-338.

Repp, Bruno H. 1982. "Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception." *Psychological Bulletin* 92(1).81-110.

Roca, Iggy & Johnson, Wyn. 1999. *A Course in Phonology*. Oxford: Blackwell.

Salmons, Joseph. Forthcoming. “Segmental phonological change.” *A Companion to Historical Linguistics*, ed. by Vit Bubenik & Silvia Luraghi. London & New York: Continuum.

Sledd, James H. 1966. “Breaking, umlaut and the Southern drawl.” *Language* 42.18-41.

Stevens, Kenneth N. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Weismer, Gary. 1979. “Sensitivity of Voice Onset Time (VOT) Measures to Certain Segmental Features in Speech Production.” *Journal of Phonetics* 7.197–204.

Yang, Charles. 2002. *Knowledge and learning in natural language.* Oxford University Press.