1 Language Science

This course is about phonology, but it is valuable to begin with a broader sense of both the science of language and cognitive science to see how phonology fits into this enterprise.

The scientific study of language is an important part of understanding human cognition. As Chomsky (1965) explains, there are three fundamental goals of cognitive science.

1. One goal is to characterize the knowledge that humans possess in various cognitive domains such as perception, categorization, language, and reasoning. A theory which characterizes this knowledge is called a theory of human *competence*.

2. A second goal is to discover the methods by which knowledge is acquired, including a characterization of innate restrictions on the form and acquisition of knowledge. A theory achieving this goal is a theory of *learning*.

3. A third goal is to understand how knowledge is applied in particular behaviors, and what limitations prevent performance from being coextensive with competence. Achieving this goal entails construction of a theory of *performance*.

Chapter 1 of Chomsky (1965) explains what these goals mean for the science of language. Following Chomsky and Halle 1968, p. 3, it is useful to "think of a language as a set of sentences, each with an ideal phonetic form and an associated intrinsic semantic interpretation. The grammar of the language is the system of rules that specifies this sound-meaning correspondence."

Consequently, the *grammars* of languages constitute a theory of competence because they encapsulate the knowledge that humans have regarding their language (construed as a set of phonetic forms paired with semantic interpretations). Chomsky (1965, pp. 8–9) elaborates further.

> *. . . by a generative grammar I mean simply a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences. Obviously, every speaker of language has mastered and internalized a generative grammar that expresses his knowledge of his language. This is not to say that he is aware of the rules of the grammar or even that he can become aware of them, or that his statements about his intuitive knowledge of the language are necessarily accurate . . . . Similarly, a theory of visual perception would attempt to account for what a person actually sees and the mechanisms that determine this rather than his statements about what he sees and why, though these statements may provide useful, in fact, compelling evidence for such a theory.*

Where do these grammars come from? To some extent, they must be learned because humans learn the language spoken in the community they are raised in. The learning theory for language science must explain how humans come to acquire a particular grammar based on their experience. Chomsky (1965, p. 7) considers the learning theory to be a centerpiece of language science.

> *. . . the main task of linguistic theory must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning.*

A correct theory of learning would provide a deep explanation for theories of competence. It would provide a single explanation of how humans acquire the various specific "system[s] of rules" which relates pronounced forms with semantic interpretations.

Theories of performance are another important part of language science. These theories seek to explain systematic facts about language use: its production, comprehension, and the behavior we observe every day.

With respect to performance, Chomsky (1965, p. 10) writes

> *There seems to be little reason to question the traditional view that investigation of performance will proceed only so far as understanding of underlying competence permits. . . . To my knowledge, the only concrete results that have been achieved and the only clear suggestions that have been put forth concerning the theory of performance, outside of phonetics, have come from studies of performance models that incorporate generative grammars of specific kinds . . . .*

Chomsky is firmly stating that any correct theory of performance will depend on a correct theory of competence. There can be little doubt that this position is correct. A theory of performance is ultimately a theory of how many different factors interact to produce linguistic behavior. The grammar is but one of these factors. Chomsky and Halle (1968, p. 3) explain the distinction between competence and performance this way.

> Competence . . . is not to be confused with performance. Performance, that is, what the speaker-hearer actually does, is based not only on his knowledge of the language, but on many other factors as well-factors such as memory restrictions, inattention, distraction, non-linguistic knowledge and beliefs, and so on. We may, if we like, think of the study of competence as the study of the potential performance of an idealized speaker-hearer who is unaffected by such grammatically irrelevant factors.

The distinction between competence and performance is a fundamental one in language science. It isolates the grammatical factors from non-grammatical ones.

A useful analogy is to physics and motion. Performance is what we observe. When a projectile is launched, there are many factors that influence its movement. Is it raining, is it windy? The shape and size of the projectile will also matter. However, these factors are distinct from fundamental factors such as its mass and the force of gravity. One of Newton's central contributions was that if you wish to understand how objects move on earth or in space, it is important to understand how point masses move in vacuums. The point mass in a vacuum is clearly an idealization, but with it, the laws of physics come into plain

view. The competence/performance distinction likewise seeks to isolate the fundamental factors (grammatical competence) from ones that affects everyday behavior (performance). *Factorization* is a critical component of any scientific investigation.

Chomsky's words about performance from page 10 of *Aspects*, however, should not be taken to mean that research on the theory of performance cannot begin until a theory of competence is completed. On the contrary, different theories of performance which encode different grammatical and linguistic theories can be a fruitful way to determine what the right theory of competence is. In fact, in the past twenty years psycholinguistic studies have provided valuable insights into the nature of grammars and learning.

2 Phonological Knowledge

The aspect of language that phonology addresses is the knowledge speakers have regarding the pronunciation of words and phrases in their language. Systematic facts about the pronunciation of words and phrases in a language are often referred to as *sound pattern* of a language.

It is a striking fact that natural languages have sound patterns. I will give examples of sound patterns in a moment in an effort to convince you that there are systematic aspects to your pronunciation, and that every human language exhibits sound patterns. But for now, assume that there is something systematic about the way you pronounce words and phrases in your language. What then are the goals of phonology? It is precisely to find an explanation for this fact.

Not surprisingly, there are three dimensions this explanation, which match the goals expressed above. Understanding the sound patterns of languages means answering the following three questions.

1. (Competence) How can we characterize the knowledge speakers have about the sound patterns of their language?
2. (Learning) How do children come to acquire this knowledge about their language?
3. (Performance) How do humans produce and perceive speech?

In service to the above questions, the study of phonology tries to answer the following:

- What are the right linguistic generalizations regarding sound patterns? This can also be expressed as follows.
  - What are the sound patterns of the world's languages?
  - What are the principles underlying these patterns?
  - What are possible sound patterns?
- How could anything learn a sound pattern from the kinds of input

- How do language phonologies change over time?

In this course, we mainly address the questions in 2 above. Answers to these questions help us better understand the competence humans bring to bear on the sound patterns of their language.

With these questions in mind, I would like to now present a brief survey of different kinds of phonological knowledge. Speaking generally, there are three kinds of knowledge regarding the pronunciation of words and phrases in natural language: *phonotactic* knowledge, knowledge of *processes*, and knowledge of *contrasts*. These are all examples of sound patterns that humans have internalized. A fourth type of sound pattern comes from *typology*: are there recurring tendencies or patterns *across* languages? In this section, a brief introduction to each type of sound pattern is presented.

2.1 Phonotactics

Phonotactic patterns refer to the possible words in a language (Chomsky and Halle, 1965; Halle, 1978).

English speakers can coin new words. However, speakers are much more willing to coin new words with words on the left, and not with the words on the right. It is striking

flump	flunp
blick	bnick
bist	bizt
slem	srem

Table 1: Possible and impossible English words.

that many English native speakers agree with the division in 1 even though they have zero experience with *all* of those words. This is an example systematic, uniform linguistic behavior that requires an explanation.

Assuming agreement among the native English speakers, how did they learn to discriminate words they never heard before in the same way? The answer of course is that there is a system of rules or constraints—a grammar—that speakers have internalized that allow them to classify logically possible words into well- and ill-formed groups. This is knowledge that English speakers acquired and learned since for example words like [srem] are perfectly acceptable in other languages. But it is also knowledge which English speakers were never explicitly taught.

With respect to the above case, it would appear that this system of rules is sensitive to the sub-parts of words. All the sub-parts of the words on the left in 1 are well-formed, but that is not the case for the words on the right. English words cannot end with [np] or [zt] sequences, nor can they begin with [bn] or [sr] sequences.

This last sentence is an example of a *linguistic generalization*, which is a hypothesis about the character of our linguistic knowledge and competence. This hypothesis makes

at least two predictions. First, existing English words should not begin or end with the aforementioned sequences. We can examine existing English words to see whether this prediction is true. Second, it predicts that English speakers should avoid coining new words with the aforementioned sequences. This can be is testable in principle with a behavioral experiment. Native speakers can also conduct this behavioral experiment on themselves with a little bit of introspection.

Here is another example of a sound pattern. Below are actual and hypothetical words below from Navajo (Sapir and Hojier, 1967).

Possible Navajo Words		Impossible Navajo Words
ʃiːteːʒ	'we (dual) are lying'	ʃiːteːz
dasdoːlis	'he (4th) has his foot raised'	dasdoːliʃ
sokos	(hypothetical)	sokoʃ
ʃokoʃ	(hypothetical)	ʃokos
kiːteːp	(hypothetical)	
piːteːk	(hypothetical)	

Note [ʃ] is like *sh* in *shoe* and [ʒ] is like *ge* in *beige*. Can you determine what grammar speakers of Navajo have internalized that allow them to distinguish between these two groups of words?

It clearly has to do with the sounds [s,ʃ,z,ʒ], which are examples of sibilant sounds. In Navajo, words can contain either [s,z] sounds or [ʃ,ʒ] sounds, but not both. As the examples indicate, it is not necessary that the sibilants be adjacent to each other. In fact, they can be separated by many other speech sounds as evidenced by the word [dasdoːlis] 'he (4th) has his foot raised.'

The above examples have established one kind of sound pattern: phonotactic knowledge. There are rules and constraints which govern the possible words in languages. Speakers can coin new words, but they cannot coin any arbitrarily sequence of speech sounds as a new word. Speakers distinguish logically possible words with which they have had no prior experience. This is the expected behavior of individuals who have internalized a productive and generative system of rules and constraints.

The above examples made a binary distinction between "possible" and "impossible" words. Some question whether this is just a convenient abstraction (Albright and Hayes, 2003). It has been argued that there are multiple levels of well-formedness. For example, it has been argued that in English [kɪp] is more well-formed than θwiːks, which is more well-formed than bzɑrʃk. Whatever the details may be, the central fact of phonotactic knowledge is that speaker-hearers of the same language community uniformly distinguish logically possible words in the same way, at least more or less. What is the nature of this knowledge—what are the rules and constraints that govern this system—and how do children come to learn it?

2.2 Processes

Another aspect of phonological knowledge comes from phonological *processes*. Evidence for phonological processes comes from *morphological alternations*. Morphemes are the smallest sequences of speech sounds with a particular meaning. Therefore, unlike phonotactic patterns, the semantic meanings or morphemes play an important role in understanding the evidence for phonological processes.

A morphological alternation is the observation that the same morpheme is pronounced differently in different contexts. The English plural provides a familiar example.

	singular	plural
cat	k^hæt	k^hæt**s**
sack	sæk	sæk**s**
dog	dɑg	dɑg**z**
grub	gɹʌb	gɹʌb**z**
dish	dɪʃ	dɪʃ**əz**
fudge	fʌd͡ʒ	fʌd͡ʒ**əz**
pea	p^hi	p^hi**z**
cow	k^haʊ	k^haʊ**z**
man	mæn	m**ɛ**n
foot	fʊt	f**i**t
wife	waɪf	waɪ**vz**
whiff	wɪf	wɪf**s**
...		

Ignoring irregular forms like *men* and *feet*, it is clear that plural morpheme has three forms *-s, -z, -əz*. These do not appear to be assigned arbitrarily to nouns. One way to see this is to conduct the following experiment. Which pronunciation goes with which of the following made-up words?

1. lɛp
2. rɑg
3. nɪz

As with the phonotactic patterns above, the fact that English speakers answer this question uniformly is strong evidence that they have internalized a system of rules and/or constraints.

There are several ways these facts can be analyzed. Below I provide two analyses to illustrate.

Analysis 1

1. If the word end with a sibilant, use [-əz] for words

2. Else if the word ends with a voiceless consonant, use [-s].
3. Else use [-z].

Analysis 2

1. Add the suffix [-z] to the noun. Consider the resulting word.
2. If there are two adjacent sibilants, *epenthesize* (insert) a schwa [ə] between them.
3. Then if there is a voiceless consonant before the [-z] *change* the [-z] to [s].

Both analyses can account for the facts of regular plural formation in English. In the first one, which I will refer to as the *morphological analysis*, a choice is made among variants based on the pronunciation of the noun. In the second one, which I will refer to as the *phonological analysis*, the plural morpheme is fixed as [-z] and the resulting word may undergo *transformations* if the proper conditions are met. These transformations are insertion of schwa (called schwa-epenthesis) or devoicing of the [z]. Such transformations are called *phonological processes*.

Generative phonologists argue, as I will do in the next chapter, that in general the phonological analysis (Analysis 2) is the better scientific explanation than the morphological explanation (Analysis 1) for the various pronunciations of the plural morpheme. On the basis of those arguments, we will conclude that morphological alternations constitute evidence for phonological transformations. Indeed, these arguments are so important they provide what I call the Fundamental Principle of Phonology:

- The best explanation of morpheme's alternations is to posit a single abstract lexical representation of the morpheme and one or more phonological processes which transform this abstract lexical representation to its concrete surface pronunciations.

From this principle (and really the arguments for it), the rest of the field of phonology follows. But we are getting ahead of ourselves. This fundamental principle will be discussed, defended, and argued for in the next chapter.

Let's provide another example, this time from Georgian (Aronson, 1982). Consider the form of the adjectival suffix below:

phizik-uri	'physical'
kimi-uri	'chemical'
akti-uri	'active'
phrang-uli	'French'
german-uli	'German'
reakti-uli	'reactive'
real-uri	'real'
terminal-uri	'terminal'

What form of the suffix do you think would go on to a hypothetical words like

[pir]	'anxiety, pain'	[bar]	'fence'
[p^{h}ir]	'Turn on!'	[b^{h}ar]	'burden'
[tal]	'lake'	[dar]	'a kind of tree'
[t^{h}al]	'plate'	[d^{h}ar]	'edge'
[kal]	'time, death'	[gol]	'circle, charcoal'
[k^{h}al]	'kind, skin'	[g^{h}ol]	'Mix! Stir!'

Table 2: Minimal pairs for aspiration as contrastive in Nepali.

1. misato
2. pita
3. badurpi
4. ramuto
5. ralako

As with the example from the English plural, it is possible to state two analyses. One selects the correct pronunciation of the morpheme based on qualities of the root. The other fixes the phonetic form of the morpheme, affixes it to the root, and then subjects the resulting word to a series of phonological processes, or transformations.

Also, it is useful to ask ourselves: How did English and Georgian speakers learn these patterns? What possible purpose could such patterns serve?

2.3 Contrasts

A third kind of sound pattern has to do with what is called *contrast*. As with phonological processes, meaning plays a role here too.

Speech sounds are *contrastive* if they can be used to signal different meanings. For instance consider the words shown below from Nepali.

Here there are pairs of words which establish that aspiration [^{h}] is contrastive in Nepali. In each pair, the words are identical except for the presence or absence of aspiration in the first consonant. Therefore, in order to know what meaning is being conveyed, it is necessary to know whether aspiration is a present or absent property of these sounds. This is why aspiration is said to be contrastive in Nepali.

The relations that exist between contrastive units of speech are an important part of the study of contrast. In Nepali it is clear that the relation between [p] and [p^{h}] is the same as the relation between [g] and [g^{h}]. However, it is not always straightforward to determine the system of contrast present in a language. For instance what are the relations that are present in a language with a three-way contrast between the vowels [i,u,a]? As we will see, there is evidence that part of the knowledge that speakers have of the system of contrast of their language includes these relations. This evidence partly comes from the fact that languages with the same speech sounds, even the same contrastive speech sounds, organize

the relations between these sounds *differently*, as was famously pointed out by Sapir (Sapir, 1925).

Furthermore, we will see, some languages may allow certain speech sounds to be contrastive in all positions in a word and others may allow certain speech sounds to be contrastive in only some positions. The lexicon of some languages may show that many words make use of particular contrastive features, but other contrastive features are only present in a small portion of the lexicon. Marginal and limited cases like these are of interest because it raises the question: why should a language make use of a contrast in only a limited number of circumstances?

2.4 Cross-linguistic patterns

A fourth type of sound pattern refers to the patterns that emerge when multiple languages are considered. Typological generalizations are especially interesting because they exist despite the considerable variation observed across languages. They therefore can lead to hypotheses about the *linguistic universals*. These in turn help distinguish the *humanly possible* phonological grammars from the *logically possible* ones.

Here are some examples of typological generalizations that have been presented in the phonological literature.

- If a language has phonemes {b,d,g}, it tends to have {p,t,k} (Maddieson, 1984).
- The word-initial triple consonant clusters a language allows is a subset of the double consonant clusters it allows (Greenberg, 1978).

  Let me explain with an example from English. In this language, every initial triple consonant cluster (like *str*) is decomposable into clusters of length two; in this case, *st* and *tr*. This generalization does not claim the converse. In fact, not all legal initial clusters of length two can be combined to form legal initial triple clusters. While English words can begin with both though *st* and *tw*, the general consensus appears to be they cannot begin with *stw* (as in **stwɛk*).

- If a language palatalizes consonants before [e], they also do before [i] but not vice versa (and people generalize this way in experimental conditions) (Wilson, 2006). This was understood to be exceptionless, until recently (Oxford, 2014).

2.5 Summary

Phonological patterns are the sound patterns of language. Phonotactic patterns, phonological processes, and systems of contrast are facets of a phonological system that govern the knowledge speakers have about the pronunciation of words and phrases in their language. When these phonological systems are studied en masses, typological sound patterns emerge as well. Each kind of sound pattern provides some insight into the nature of the underlying

system, and consequently into the nature of how the language faculties of human minds work.

As we will see, different phonological theories intertwine systems of contrast, phonotactic patterns, and phonological processes to different degrees and in different ways. In this course, we will develop an awareness of the empirical facts and will concentrate on developing principles for phonological analysis and for evaluating theories.

3 The Fundamental Principle of Phonology Theory

Phonology is a theory that explains the systematic variation in the pronunciation of meaningful units of speech, which is found in every natural language. The theory states that each meaningful unit of speech—these are called *morphemes*—typically has a single mental representation and that its various pronunciations follow from context-dependent, language-specific laws. Therefore the major issues within phonology are the nature of the representations and the nature of the laws.

The idea that morphemes typically have a single mental representation (despite their potentially many different pronunciations) and the idea that there are mental laws which govern aspects of speech is controversial. After all, while we all witness speech acts every day, no one has ever witnessed a mental representation or law.

Even though phonological constructs cannot be observed directly, I will endeavor to show that there is plenty of evidence for them. The evidence for phonology is indirect evidence, but it is evidence. The use of indirect evidence to establish the existence of otherwise invisible objects is not a methodology unique to phonology. Astronomers, for instance, have recently deduced the existence of extrasolar planets, not by direct observation, but instead by detecting their effects on the motion of their home stars. The reasoning is that the motion of those stars is best understood if a planet-sized object exists in its near vicinity, even if this planet-sized object is not directly observable.

Much evidence for phonology comes from the fact that many languages have multiple morphemes which exhibit the *same* kind of variation in their pronunciation. If different morphological paradigms have the same character and there are no underlying laws binding them together then this fact is unexpected and unexplained. This is the sort of evidence that will be presented here.

3.1 Morphological versus phonological analyses

The figures below summarize two distinct views of the role phonology plays in the architecture of human grammar:

A. Syntax → Morphology → Phonetics

B. Syntax → Morphology → Phonology → Phonetics

The architecture in (A) appears sufficient for all situations. There are distinct morphemes with distinct meanings and pronunciations. These are concatenated according the word formation rules. No further modifications are necessary and so the result can be passed immediately to the phonetics component. In a model of speech production, the phonetics component would then construct a motor plan to be executed by the neuromuscular system. Why then is there any need for an architecture like (B)?

The key evidence for phonology, and thus for an architecture like (B), comes from the fact that many languages have multiple morphemes which exhibit the *same* kind of variation in their pronunciation. Evidence also comes from the fact that certain patterns of variation are *not* attested. It may seem counterintuitive but it is true: when seeking explanations, it is important to not only explain what is there, but also to explain what is not there. The Copernican theory was successful in part because it could explain why certain logically possible motion of the heavenly bodies was not present (in contrast to the Ptolemaic theory which could put another epicycle on another epicycle).

Consider the data below from Lithuanian (Odden, 2005, p. 114).

[ateiti]	'to arrive'
[atimti]	'to take away'
[atneʃti]	'to bring'
[atleisti]	'to forgive'
[atlikti]	'to complete'
[atkoːpti]	'to rise'
[atpraʃiːti]	'to ask'
[atkurti]	'to reestablish'

The above forms show a verbal prefix *at-* concatenated to various verbal roots, which are concatenated with a verbal suffix *-ti*. We will leave aside the verbal suffix. When the same verbal prefix is concatenated to words which begin with voiced obstruents, however, it is not pronounced *at*; instead, it is pronounced *ad-*. (The data shown below is also from (Odden, 2005, p. 114).

[adbekti]	'to run up'
[adgauti]	'to get back'
[adbukti]	'to become blunt'
[adgimti]	'to be born again'

How can these facts be explained? There are at least two possible morphological analyses.

1. There are two classes of words in Lithuanian, call them X and Y. If the verb is an X-verb select *at-* otherwise (it is a Y-verb and so) select *ad-*.

2. With respect to the verbal prefix which means μ_1, there are two pronunciations: *at-*,*ad-*. If the root begins with a voiced obstruent, select *ad-*, otherwise select *at-*.

It is worthwhile to work out derivations of *ateiti, atkurti* and *adbekti* for each morphological analysis.

The problem with the first analysis should be obvious. Under this account, there is no explanation for the fact [ad-] always occurs with words beginning with voiced obstruents. It makes no predictions regarding what a native speaker of Lithuanian would do when faced with newly coined verbal stems such as *bel* or *pel*. In terms of its generative capacity, anything goes.

The second analysis, on the other hand, is much more plausible. It correctly captures the generalization that the verbal prefix is pronounced *ad-* before roots beginning with a voiced obstruent and pronounced *at-* elsewhere. It also makes predictions about how speakers would behave with newly coined words, which for now we will assume to be correct. So given such an analysis, why is a phonological analysis an improvement?

Here is a phonological analysis of the Lithuanian data above.

1. The *abstract, underlying* form of the prefix is *at-*. There is a phonological *process* which voices voiceless obstruents before voiced obstruents.

Again, it is worthwhile to work out derivations of *ateiti, atkurti* and *adbekti* for the phonological analysis.

The phonological analysis only posits a single pronunciation for this verbal prefix. However, its pronunciation changes according to the phonological rule. This analysis also provides an explanation as to why Lithuanian speakers pronounce this verbal prefix *ad-* before roots beginning with a voiced obstruent and pronounced *at-* elsewhere. It also makes predictions regarding newly coined words. For the moment, let's assume these predictions are right. In other words, the phonological analysis so far makes the *same* predictions as the second morphological analysis above.

The difference between the mental representations of the verbal prefix are illustrated in Figure 1. The phonological analysis posits only one mental representation for the pronunciation of the prefix, whereas the morphological analysis posits two. I will refer to lexical items with one and more than one mental representations of their pronunciation as *uniforma* and *multiforma* morphemes, respectively. I use these terms to refer explicitly to the mental representation of *pronunciation*—whether a morpheme has only one or more than one meaning is irrelevant to whether it is uniforma or multiforma.

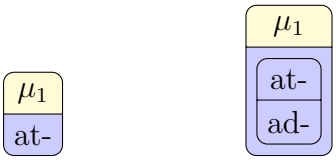

Figure 1: Phonological hypothesis posits uniforma morphemes (left) and the morphological analysis posits multiforma ones (right).

How does the size of the phonological analysis compare with the second morphological analysis above? While it simplifies the lexicon (reducing two pronounced forms in the lexical

item to one) it complicates the grammar by adding a phonological rule. Indeed it adds another stage to the derivation of pronounced forms that did not exist before. Is this extra cost worth the savings?

Generative phonologists believe it is. Most importantly, this theory is more *restrictive* because it makes more predictions. It predicts that words in Lithuanian should prohibit voiceless-voiced obstruent clusters. If words are required to be subject to the phonological module before they are pronounced, and it contains this rule, then this theory predicts voiceless-voiced obstruent clusters should not be present in Lithuanian. Furthermore, it predicts that if there are *other* morphemes whose concatenation results in voiceless-voiced obstruent clusters that the phonological process *will apply* and that this morpheme will exhibit the *same* variation in pronunciation as the /at-/ morpheme.

The following data from Lithuanian of another verbal prefix bears directly on these predictions.

[apeiti]	'to circumvent'
[apieʃkoːti]	'to search everywhere'
[apakti]	'to become blind'
[apmoːkiːti]	'to train'
[aptemdiːti]	'to obscure'
[apʃaukti]	'to proclaim'
[abgauti]	'to deceive'
[abʒʲureti]	'to have a look at'
[abʒelti]	'to become overgrown'
[abdauʒiːti]	'to damage'
[abdraskiːti]	'to tear'

How does the second morphological analysis explain the systematic variation in the verbal forms in Lithuanian? As shown in Figure 2, the phonological analysis posits another uniforma morpheme and may keep the same rule. The morphological analysis will have to posit another multiforma morpheme, and will have to posit another statement which explains how the pronunciation is selected. This rule would look something like this

- With respect to the verbal prefix which means μ_2, there are two pronunciations: [ap-,ab-]. If the root begins with a voiced obstruent, select *ad-*, otherwise select *at-*.

In other words, the morphological analysis appears to be missing an important generalization because the two verbal prefixes treated independently but the variation they exhibit is of the same kind.

The theory the morphological analysis is embedded in appears to allow the following kind of variation in a language. Like Lithuanian, there would be a verbal prefix with meaning μ_1 which is pronounced [ad-] before roots with voiced obstruents and pronounced [at-] elsewhere. However, unlike Lithuanian, there would be a verbal prefix with meaning μ_2 which

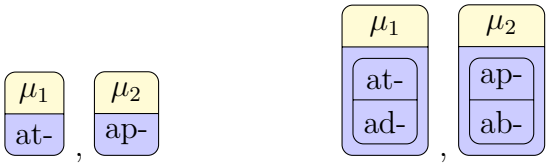

Figure 2: Phonological hypothesis posits uniforma morphemes (left) and the morphological analysis posits multiforma ones (right).

is pronounced [ap-] everywhere. This is precisely the kind of variation that the phonological analysis says is not permitted in natural language.

When we say that one analysis fails because it "misses an important generalization," that is another way of saying that the analysis is not sufficiently restrictive. The scientific principle which begets phonology is the one that says the strongest theories are not only ones that are consistent with the observations but are also the ones which make the most falsifiable predictions. In science, we want the stronger, more falsifiable theories. That is how progress is made.

Trying to improve the morphological analysis. One way to improve the morphological analysis could be as follows. Instead of providing individual statements for each morpheme which explain under what circumstances a particular pronunciation is selected, these statements could refer to groups of morphemes and be based on their phonological properties. In the case of Lithuanian, for example, the selection rule may be something like the following.

- If a multiforma prefix has one pronunciation P ending in a voiced obstruent and the root begins with a voiced obstruent then select P as the pronunciation, otherwise select the other pronunciation.

There are two things to say about this. First, this theory takes a step closer to the phonological analysis. It does so because it explains the behavior of multiple morphemes with a single generalization.

Second, it still makes weaker predictions. In particular, the phonological analysis predicts that voiceless-voiced obstruent sequences should not be present even in monomorphemic words. *Every* word is subject to the phonological module, which contains the process which voices obstruents voiceless obstruents before voiced ones. This refined morphological analysis only predicts voiceless-voiced obstruent sequences are banned across morpheme boundaries. So again, the phonological analysis is making stronger predictions. To the extent these are supported by the available evidence, the phonological analysis is the better one.

Third, and perhaps most importantly, this improvement does not appear to address lack of restrictiveness of the morphological analysis. It still appears possible for under this type of theory for there to be a language with two distinct verbal prefixes, with one pronounced *ap-* everywhere and the other pronounced *ad-* before verbs beginning with voiced obstruents and *at-* elsewhere.

3.2 The fundamental principle of phonological analysis

We thus state the fundamental principle of phonological analysis of phonological analysis as follows.

The systematic variation of pronunciation in morphological paradigms is best explained by positing

1. a *single* abstract lexical representation ('the underlying form') for each morpheme, and
2. phonological processes which transform these abstract representations to surface phonetic representations.

Consequently, there is a phonological module between the the output of the morphological component of the grammar and the phonetics component. The output of morphology is the input to the phonological component. The output of the phonological component is the input to the phonetics module.

B. Syntax → Morphology → Phonology → Phonetics

Consequently, phonology, as a field itself is primarily concerned with three questions.

1. **What is the nature of the abstract, lexical ('underlying') representations?**
2. **What is the nature of the more concrete ('surface') forms?**
3. **What is the nature of the transformation from underlying forms to surface forms?**

To emphasize the different levels of representation that are entailed by a phonological analysis, they are distinguished by different kinds of delimiters. Slashes are standardly used to represent sequences at the abstract, underlying level and braces are used at more concrete surface level. In the Lithuanian example above, one would say that the lexical representation of the verbal prefix meaning μ_1 is /at-/ and that it has two pronunciations at the surface [at-, ad-]. We will refer to representations at the abstract and more concrete levels as *underlying representations* and *surface representations*, respectively.

People have tried to derive the fundamental principle from other principles. For example, a principle on economy of description (Chomsky and Halle, 1968). It may very well derive from such a principle. However, trying to establish this depends on the descriptive formalism. So the devil, as always, is in the details.

3.3 Multiforma morphemes

In previous sections, we argued strongly against multiforma morphemes as a way to explain the systematic variation in the pronunciation of meaningful units of speech. However, this does not mean under no circumstances should multiforma morphemes ever be posited.

For instance, in Korean the nominative singular suffix has two pronunciations: [-ka] and -[i]. The former only occurs after nouns which end with a vowel and the latter only occurs after nouns that end with a consonant. So clearly a phonological condition governs the choice of the pronunciation.

A phonological analysis of the above generalization can come about in (at least) two ways. First, the underlying form could be posited as /ka/ along with a process that changes /ka/ to [i] when following a consonant-final stem. Second, the underlying form could be posited as /i/ along with a process that changes /i/ to [ka] when following a vowel-final stem. A third possibility might be to posit the underlying form to be /ki/ along with two processes: one which lowers the vowel to [a] after a vowel-initial stem and one which deletes [k] before consonant-final stems. Derivations illustrating these examples are shown below.

Analysis #1			Analysis #2			Analysis #3		
UR	/pat-ka/	/pa-ka/		/pat-i/	/pa-i/		/pat-ki/	/pa-ki/
ka⟶i	pati	—	i⟶ka	—	paka	Lowering	—	paka
						Deletion	pati	—
SR	pati	paka		pati	paka		pati	paka

Table 3: Three logically possible phonological analyses for the [ka/i] alternation in the Korean nominative singular.

The fundamental problem with each of these analyses is that the processes posited do not extend beyond the morphemes in question. There is no general fact about Korean that prohibits the sound sequence [ka] after consonants. Nor is it an accurate generalization to prohibit [i] (to the exclusion of other vowels) after vowels. Similarly, [i] does not lower to [a] generally after [Vk] sequences. So the general predictions that these analyses make are wrong.

Another problem that these analyses have is they do not fare well with respect to the naturalness criterion. While it is natural to prefer [ka] to [i] after vowel-final stems and [i] to [ka] after consonant-final stems, it is strange to change /i/ to [ka] after vowel-final stems and to change /ka/ to [i] after consonant-final ones. Of all the repairs that could be done, this one seems too far-fetched.

In other words, the best explanation in this case is to posit a multiforma morphemes and a word formation rule which is sensitive to the phonological make-up of the stem. This is known as **phonologically-conditioned allomorphy**.

The fact that phonologically-conditioned allomorphy exists (i.e. is the right characterization of the knowledge speakers have regarding the pronunciation of certain morphemes) may raise the question whether phonology should exist at all. After all, if it is possible to posit multiforma morphemes and phonologically conditioned word formation rules, why not do everything that way? The answer is the same as the one given before. In *general*, multiforma morphemes are not good explanations of the systematic variation in the pronunciation of morphemes, even if, as we have seen here, they may be in some instances. When

analyzing the pronunciation of morphemes in a language, the best scientific practice is to begin with a phonological analysis which makes the strongest predictions and to *retreat* to the phonologically-conditioned allomorphy if pushed.

4 Formal Theories of Phonology

Next we turn to different formalizations of phonological theory. Broadly speaking, there are two influential ones. They differ fundamentally in the way they answer the key questions in phonology, which are repeated here.

1. What is the nature of the abstract, lexical ('underlying') representations?
2. What is the nature of the more concrete ('surface') forms?
3. What is the nature of the transformation from underlying forms to surface forms?

These two theories are:

1. Ordered Rewrite Rules (Chomsky and Halle, 1968)
2. Optimality Theory (Prince and Smolensky, 1993, 2004)

I will usually refer to the former as SPE[1] and the latter as OT. However, it should be understood that the SPE and OT I describe are not exactly the theories as they are described in their original formulations. Instead they are variants, around which I believe some consensus has formed.

These theories have different *ontologies*. According to Wikipedia, the principles questions of ontology are include "What can be said to exist?", and "Into what categories, if any, can we sort existing things?" In phonology, the existing things we are interested in is our phonological *knowledge*—our knowledge of the lexical representations, the surface forms, and the mapping—and how this knowledge is manifest in our minds (what is psychologically real).

Broadly speaking, the difference between these theories can be stated as follows.

1. In SPE, the rules—which describe transformations from underlying forms to surface forms—are the primary psychological objects. These rules apply in some order. Additionally, there are statements on the well-formedness of abstract, lexical representations. The rules, their ordering, and these statements are language-specific. The surface phonotactic knowledge *emerges* as a consequence of these constraints on underlying representations and the rules themselves.

[1]This is the acronym for *Sound Pattern of English*, the book by Chomsky and Halle.

2. In OT, there are no rules, only *constraints*. In most exponents of the theory, these constraints come in two forms: some constraint the well-formedness of surface representations and others constraint directly and simply how underlying forms may relate to surface forms. Transformations between underlying and surface forms *emerge* via the interaction of these constraints. The constraints are universal to all languages, but their relative importance in any given language is language-specific.

These are not the only two theories that have been studied. Especially in the 1980s, prior to the advent of OT in the early 1990s, there were many theories which mixed rules and constraints. More recently, there is a new grammatical model called Harmonic Serialism (McCarthy, 2008), which also synthesizes aspects of these two earlier approaches, broadly construed. However, we will focus on the aforementioned two theories, and only mention these others as needed.

Finally, while the theories do differ, it is important to realize they all admit that the phonological component of a grammar includes underlying forms, surface forms, and a mapping between them. In other words, the common base for these two theories is precisely the content of the previous chapter.

References

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90:119–161.

Aronson, Howard. 1982. *Georgian, A Reading Grammar*. Slavica Publishers, Inc.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam, and Morris Halle. 1965. Some controversial questions in phonological theory. *Journal of Linguistics* 1:97–138.

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

Greenberg, Joseph. 1978. Initial and final consonant sequences. In *Universals of Human Language: Volume 2, Phonology*, edited by Joseph Greenberg, 243–279. Stanford University Press.

Halle, Morris. 1978. Knowledge unlearned and untaught: What speakers know about the sounds of their language. In *Linguistic Theory and Psychological Reality*. The MIT Press.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge, UK: Cambridge University Press.

McCarthy, John J. 2008. The gradual path to cluster simplification. *Phonology* 25:271–319.

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

Oxford, Will. 2014. Patterns of contrast in phonological change: Evidence from Algonquian vowel systems. *Language* To appear.

Prince, Alan, and Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Tech. Rep. 2, Rutgers University Center for Cognitive Science.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.

Sapir, Edward. 1925. Sound patterns in language. *Language* 1:37–51.

Sapir, Edward, and Harry Hojier. 1967. The phonology and morphology of the Navaho language. *University of California Publications* 50.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30:945–982.