Introduction to Phonology

École d'automne de linguistique, ENS

Class coordinates
Time: 14:30-15:50 (Session 4), Sept. 24, 25, 26, 27 (Monday to Thursday)
Place: Salle des Résistants (45 rue d'Ulm, 1er étage, couloir A-B)

Instructor coordinates
Name: Kie Zuraw [ˈkʰaɪ ˈzɝˌɔ]
Affiliation: UCLA (University of California, Los Angeles) Department of Linguistics
E-mail: kie@ucla.edu
Web page: www.linguistics.ucla.edu/people/zuraw

Course description
What do we know about our language's sound pattern, and how do we know it? This course will begin with a quick overview of characteristics of sound patterns that linguists have noticed (alternations and phonotactics), and of the approach to explanatory adequacy that will be adopted here. We will then look at research that has sought to determine what phonological generalizations speakers extract from the learning data, and follow the consequences of these findings for achieving a descriptively adequate grammatical framework (that is, a framework that can express speakers' implicit phonological knowledge): basic rule notation, features, and constraint interaction. Next we will consider why determining what speakers know is so difficult, and review a range of methods that have been tried. Finally, we will examine some recent work that moves towards explanatory adequacy—what kind of learner can, on exposure to typical learning data, choose a grammar similar to the one that human learners choose?

Prerequisites: None!

Course outline

Day 1: 24 September	sound patterns conceptual framework
Day 2: 25 September	descriptive adequacy: methods and consequences
Day 3: 26 September	explanatory adequacy: methods
Day 4: 27 September	explanatory adequacy: theoretical developments

Suggestions for further reading are included at the ends of the first two handouts

Language: In accordance with EALing policy, I'll lecture in English, but feel free to make comments or pose questions in French, to ask me to try to express something into French if it's not clear in English, to talk to me after class in French...

Day 1:   Sound patterns; Conceptual framework

1.1   Introduction

1.   The goal

In phonology, we're trying to understand how humans learn, store, and use the **sound patterns** of their languages.
These questions are connected to other questions: what is the range of sound patterns in the world's languages and why? how do languages' sound patterns change over time?

We'll begin by illustrating what a sound pattern might be. But first, a little exercise...

2.   A metaphor: handwriting [idea apparently originates with Andy Wedel]

See blackboard for data—conclusions of the exercise

- "A" and "B" represent distinct categories
- Each letter has a set of major variants: majuscule/minuscule, printed/cursive
- Which major variant is used depends on linguistic context (e.g., beginning of sentence, beginning of proper name take majuscule) and social context (filling out a government form with printed letters vs. writing a personal letter in cursive)
- Each major variant has infinitely many minor variants, conditioned by surrounding letters, speed/carefulness of writing, individual handwriting habits, health/mood of the writer, and random fluctuations.
- The difference between "major" and "minor" is not sharp.
- The "same" letter may tend to have different realizations in different countries/regions

3.   Sounds have a similar property!

/p/ and /b/ represent distinct categories—called **phonemes**—in many languages, including French and English

- Each phoneme has some major variants, called **allophones**
  - English: /p/ → [p], [p^h]; /b/ → [b], [b̥]
- Which allophone to use can depend on linguistic context
  - English: /p/ → [p^h] at the beginning of a word or beginning of a stressed syllable (*[p^h]otáto*, *a[p^h]ártment*), and otherwise [p] (*s[p]útter*, *ú[p]er*)[1] (Acute accent ´ indicates main stress)
- Which allophone to use can also depend on extralinguistic context.
  - London English: /t/ → [t] or [ʔ] at the beginning of an unstressed, non-word-initial syllable: *bu[t]er* or *bu[ʔ]er*. The choice can depend on social context.
- Each major allophone also has infinitely many minor variants.
  - conditioned by surrounding sounds: English [p] (and [p^h]) can be realized with the lower lip touching the upper teeth instead of the upper lip if followed by an [f], as in *Zipfian*.
  - conditioned by speech rate/carefulness: In rapid speech, English [p] (and [p^h]) may have lips closed for less time or not at all.
  - conditioned by individual speech habits: Some people are more likely than others to close their lips all the way for [p]

---

[1] With certain complications that we will ignore here. Ask some English speakers how they'd pronounce words like *escapee* (Albright 2006).

    - due to health/mood of the speaker: “scratchy” voice from inflamed larynx, loudness affected by mood...
    - due to random fluctuations: how long lips are closed, how long and loud the puff of air (^{h}) is in [p^h]
- The difference between major and minor variants is not sharp: should the “labiodental” /p/ that can occur in *Zipfian* be considered a major allophone of /p/?
- The “same” phoneme can have different typical realizations in different languages, dialects
    - Word-initially, French /b/ has more vocal-fold vibration than English /b/.
    - Compare French /i/ to English /i/.

Phonemes are very different from letters, however, in that we are rarely taught explicitly how to produce the phonemes of our native language, or what their allophones are!

4. Some conventions

- <puis> Angled brackets <> indicate spelling
- /pyi/ Slashes // indicate phonemes, in standard phonetic notation (IPA—similar to system used in many European dictionaries) or in the transcription system of the data source. If a symbol is unfamiliar and I don’t explain it, ask me! The sequence of phonemes making up a word is called its **underlying representation** (***forme sous-jacente***).
- [pɥi] Square brackets [] indicate a phonetic transcription, anywhere from major allophones ([pɥi]) to fine details ([pɥ̥i]...). May indicate how the word is usually pronounced, or may transcribe how it was pronounced on a particular occasion. Such transcriptions are known as **surface representations**.

1.2 What is a sound pattern: alternations

5. Example from Mbabaram[2]

(Dixon 1991; language from Australia with one speaker left at the time of Dixon’s research)

pɨr	‘emu’		
aba	‘body’		
alba	‘camp’		
nap	‘who’		
palán	‘moon’		
púmba	‘ashes’		
ɲɨ́p	‘what’	ɲɨ́b-ug	‘for what reason’
mbɛrp	‘wild dingo’	mbɛrb-ul	‘wild dingo-erg.’
tulbu	‘matches’		
tum	‘hard’		
kùludún	‘dove’		
adil	‘ring-tail possum’		
arɔk	‘magpie’	arɔg-uŋg^{ə}	‘magpie-erg.’
kuŋgak	‘kookaburra’	kuŋgag-ul	‘kookaburra-erg.’
kaɽúk	‘bandicoot’	kaɽúg-uŋ	‘bandicoot’s’

[2] The data are simplified! See the original for a fuller description of voicing

❒ Do you think [p] and [b] represent different phonemes or are allophones of the same phoneme in Mbabaram? If they're allophones of the same phoneme, in what contexts does each allophone appear?

❒ How about [t] and [d]? [k] and [g]?

6. Making our observation explicit

We can write **rules** to describe a phoneme turning into some allophone in some environment.

To express XAY → XBY, we extract the redundant X and Y and write **A → B / X __ Y**

❒ Write rules for what we observed in Mbabaram.

The sequence XAY is called the **structural description** of the rule, and the change from A to B is the **structural change**. A is sometimes called the rule's **target** (***cible***), and **X__Y** is called A's **environment** or **context**.

A change like [ɲɨ́p] to [ɲɨ́b] is called an **alternation** (***alternance***): the same morpheme is realized two different ways, depending on context. We an also say that in Mbabaram, [p] **alternates with** [b].

7. Grouping sounds

If you have studied phonetics, you may notice that [b], [d], and [g] have something in common that [p], [t], and [k] lack: vocal-fold vibration.

This **feature** (***trait***) is usually called [voice] (*voisement*, *sonore/sourd*). [b,d,g] are said to be [+voice], and [p,t,k] are [–voice].

We can rewrite our rules with the feature:

$$\begin{Bmatrix} \text{p} \\ \text{t} \\ \text{k} \end{Bmatrix} \rightarrow \text{[+voice] / [+voice] \_\_ [+voice]} \qquad \text{(where } X \text{ means "any consonant or vowel")}$$

A set of sounds in some language that can be defined by the values of one or more features is called a **natural class**. In Mbabaram, [p,t,k] form a natural class (the [–voice] sounds), as do [b,d,g] (the $\begin{bmatrix} \text{+voice} \\ \text{–sonorant} \end{bmatrix}$ sounds—the feature [sonorant] distinguishes [b,d,g] from the language's other voiced sounds.[3]

So we can make our rule even more general: [–voice]→ [+voice] / [+voice] __ [+voice]

The above is actually redundant (why?)—we could just say X→ [+voice] / [+voice] __ [+voice]

or even (at least for the data we have) X→ [+voice] / X __ X

❒ Think of hypothetical Mbabaram words that would be consistent with the above data, but would tell us that X→ [+voice] / [+voice] __ [+voice] is the right rule, not X→ [+voice] / X __ X

[3] Actually, there are several other [–sonorant] sounds...

1.3 What is a sound pattern: phonotactics

8. Restrictions on sequences

Even in the absence of alternations, the distributions of sounds can be systematic.

For example—unlike French—English forbids /ps/ at the beginning of a word: <psychology> = [saɪk^{h}ɑləd͡ʒi].

Adding a prefix does not help: *anthropo[s]ychic*, *megalo[s]ychic*, *para[s]ychology*[4]. There is no alternation; the sequence simply does not seem to occur in underlying forms.

Rather than try to write a rule (*p* → Ø / #__s?), many phonologists prefer to simply state the restriction as a **constraint**: *#ps, where "#" means word beginning or end. (The predominant framework today, Optimality Theory [Prince & Smolensky 1993/2004] replaces all rules with constraints.)

The description of which sounds may occur next to each other is called the language's **phonotactics** (***phonotaxe***).

Phonologists also use the term more generally to refer to a language's sound pattern apart from alternations and relationships among words (e.g., *cat-cats*)—that is, a language's phonotactics may simply describe which surface representations are legal (in the absence of any further information about the word/utterance): this includes the inventory of legal sounds, the legal positions for stress...

1.4 What is a sound pattern: contrast/information

9. Contrast

Although I won't have much to say about it, **contrast** is important in many phonological theories.

[p] and [b] are **contrastive** in French, because they can be used to differentiate words: *pas* [pa] vs. *bas* [ba]. This is evidence that [p] and [b] represent different phonemes.

pas and *bas* form a **minimal pair**—distinct words whose form differs only in one sound.

French [p]/[b]	represent different phonemes	are contrastive	are allowed in a similar range of contexts
Mbabaram [p]/[b]	are allophones of the same phoneme (/p/[5])	are predictable	are in **complementary distribution**: there is no context in which both sounds can occur

10. Information

We could also think about contrast in terms of informativeness.

In Mbabaram, the difference between [p] and [b] is not informative—if you wrote them all as <p>, you could still identify all the words (indeed, spelling systems tend to follow this principle: distinguish the phonemes, but not the allophones).

But informativeness could also be gradient:

---

[4] Except maybe in words borrowed already-prefixed from Greek, as in *a[ps]ychic*

[5] You could also devise an analysis inwhich the phoneme is /b/ and there is a rule to make it [–voice] in certain environments—try it!

- In English, according to the 125000-word Carnegie-Mellon Pronouncing Dictionary (www.speech.cs.cmu.edu/cgi-bin/cmudict), there are 3418 word pairs that differ solely in that one has [p] where the other has [b] (*ample/amble*, *beeper/Bieber*, etc.). Thus, if we collapsed the distinction, we'd create about 3418 new homophone pairs.
- Although the dictionary doesn't transcribe ^{h}, it is likely that there are no word pairs distinguished solely by [p] vs. [p^h] (there might be some close ones in my speech at least, such as *tra[p^h]eze* vs. *Ya[p]ese*). So collapsing the distinction creates about 0 new homophone pairs—it is not an informative distinction.
- There are two sounds that can be spelled <th> in English, [θ] and [ð]: [ð]ese, [θ]ick. There are just 61 word pairs distinguished solely by [θ] vs. [ð]: *thy/thigh*, *teeth/teethe*, etc. And many of these are actually variant pronunciations of the same word (*without*). So this difference is not very informative. On the other hand, it is not completely predictable unless you know the word in question. So these are generally regarded as separate phonemes (contrastive), even though the contrast is not very informative.

1.5 Relationship to other aspects of grammar

1.5.1 Morphology

11. Phonology is affected by morphology (word structure)

Korean (data from Kang 2007, though I'm using a different transcription system, and Yahoo! Korea dictionary [kr.dic.yahoo.com/])

/t/			/t^h/		
받으니	pa**d**-ɨni	'receive-therefore'	같으니	ka**t^h**-ɨni	'same-therefore'
받아	pa**d**-a	'receive-imperative'	같아	ka**t^h**-a	'same-imperative'
턱받이	t^{h}ɔk-p'a**d͡ʒ-i**	'bib' (lit. chin-receive)	같이	ka**t͡ʃ^h-i**	'together'

❐ Write a rule for what happens to /t/ and /t^h/ before /i/. (We need another rule too, to account for /t/ becoming [d].) Assume that [d͡ʒ],[t͡ʃ^h] are [–anterior] and [t,d,t^h] are [+anterior].

But the rule doesn't apply inside a morpheme:

어디	ɔ**di**	'where'	티	**t^{h}i**	'spot'

Lots of phonological rules are like this: they apply only when their structural description is created by a morphological operation.

An easy fix in this case is to add the symbol "+" (morpheme boundary) to the rule—where should we add it?

12. Morphology can also be affected by phonology

In English, the suffix *–en* that turns adjectives into verbs can apply only to monosyllabic stems that end in obstruents (but not **wetten* or **smoothen*, for some reason):

thicken	*thinnen	shorten	*tallen
blacken	*bluen	deepen	*shallowen
whiten	*greenen	darken	
redden	*orangen	lighten	
	*yellowen	brighten	

	*grayen	madden	
	*brownen	moisten	*dryen
	*purplen	quicken	*slowen
sweeten	*souren	fasten	
sharpen	*dullen	loosen	
fatten	*slimmen	tighten	
sicken	*wellen	worsen	*betteren
soften		smarten (up)	*dumben
harden		weaken	*strongen (cf. strengthen)
roughen		deaden	
widen	*narrowen	liven (up)	
broaden		stiffen	
	*calmen	slacken	

See Raffelsiefen (1996, 1998) for an overview of English suffixes that place phonological restrictions on their stems.

1.5.2 Syntax

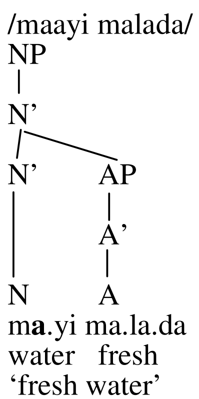
13. Phonology can be conditioned by syntactic structure

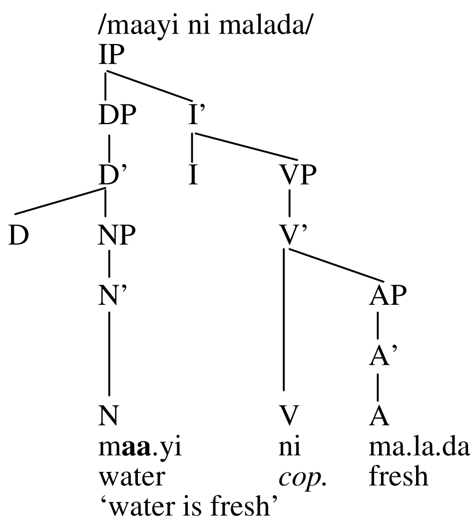
Chimwiini (Kisseberth 2000; Chimwiini is a dialect of Swahili formerly with about 40,000 speakers in Somalia; most have emigrated to Kenya)

Long vowels are allowed only in the penultimate and antepenultimate syllables of a **phonological phrase**. (In a sequence ...*EDCBA*, *B* is the penultimate and *C* is the antepenultimate.)

Under Kisseberth's analysis, in Chimwiini the end of an XP (DP, NP, AP, VP...) ends a phonological phrase (but the beginning of an XP is irrelevant).

- Why is the vowel of 'water'—underlyingly /maayi/—short in the first tree but long in the second?





14. But word order, at least in narrow cases, can also be affected by phonology
Latin (Embick & Noyer 2001): the clitic *–que* 'and', attaches after first word of second conjunct:

[bonī puerī] [bonae–**que** puellae]
good boys good–and girls
'good boys and good girls' (p. 575)

But when the second conjunct begins with a preposition, its syllable count matters:

circum–**que** ea loca	in rēbus–**que**
around–and those places	in things-and
'and around those places'	'and in things'
contrā–**que** lēgem	dē prōvinciā–**que**
against-and law	from province-and
'and against the law'	'and from the province' (p. 576)

So phonological properties—here, monosyllabic vs. longer prepositions, assuming that's the defining difference between the two sets of prepositions—can influencing word order.

1.6 Summary

- A language's sound pattern includes alternations, phonotactic restrictions, and relationships of contrastiveness or noncontrastiveness among sounds.
- A sound pattern can be characterized in terms of individual consonants and vowels and in terms of those sounds' features. (Also, although we didn't see examples, in terms of features that extend over longer spans than a single consonant or vowel, like stress or tone.)
- Some aspects of the sound pattern require information from morphology and syntax; morphology and syntax may also be affected by sound.

1.7 Framework

We've now had an overview of some of the phenomena that can occur in a language's sound pattern. So let's get back to the main job: understanding how humans learn, store, and use those sound patterns.

15. Preliminaries
Chomsky lays out a useful framework for investigating this question for language in general (see Chomsky 1965 pp. 25-27—but what it is below is an amalgam of various works, slightly simplified and coloured by my own views).

Let a **grammar** consist of (at least)[6]

- a function that labels any utterance as **grammatical** or **ungrammatical**. We can call such labellings **grammaticality judgements.**
- a function that assigns truth conditions to any utterance

[6] We probably want the grammar to do much more. It could, given an utterance, return a gradient "goodness score" rather than a simple binary judgement. Given one utterance and some instruction, it could return some other utterance (e.g., *cat* + PLURAL = *cats*). And there's a lot more to meaning than truth conditions! (Chomsky also requires a grammar to assign a structural description to an utterance, but I wonder if this is begging the question: the structural description can be used to explain more observable properties of a sentence like its truth-conditions, but we don't know *a priori* that it's necessary.)

The grammar might be implemented as a lexicon and a list of rules, or a set of constraints, or something else.
Let a **linguistic theory** be a function that, given a (finite) set of utterances (the **learning data**), produces a grammar.[7]

These functions should be accompanied by algorithms for calculating them.

Let's use a concrete example, English noun plurals.

cat	kʰæt	kʰæt**s**
sack	sæk	sæk**s**
dog	dɑg	dɑg**z**
grub	gɹʌb	gɹʌb**z**
dish	dɪʃ	dɪʃ**əz**
fudge	fʌd͡ʒ	fʌd͡ʒ**əz**
pea	pʰi	pʰi**z**
cow	kʰaʊ	kʰaʊ**z**
man	mæn	m**ɛ**n
foot	fʊt	f**i**t
wife	waɪf	waɪ**vz**
whiff	wɪf	wɪf**s**
...		

16. Observational adequacy
A grammar that accepts all the forms that a typical speaker would have been exposed to and assigns the right truth conditions to them (we don't care what the grammar says about other forms), is an **observationally adequate grammar**.

Note that there are infinitely many observationally adequate grammars for any (finite) set of learning data (why?).

Examples of observationally adequate grammars for English noun plurals

I. (just list every word you know)

kʰæt	kʰæt**s**	pʰi	pʰi**z**
sæk	sæk**s**	kʰaʊ	kʰaʊ**z**
dɑg	dɑg**z**	mæn	m**ɛ**n
gɹʌb	gɹʌb**z**	fʊt	f**i**t
dɪʃ	dɪʃ**əz**	waɪf	waɪ**vz**
fʌd͡ʒ	fʌd͡ʒ**əz**	wɪf	wɪf**s**

[7] Chomsky's definition of a linguistic theory is weaker: it need only define the set of possible grammars, independent of learning data. This allows Chomsky to define the term **descriptively adequate theory**, which is a theory that includes, as possible grammars, a descriptively adequate grammar for every language—but does not necessarily return that grammar given learning data for that language.

I.e., the grammar's judgement function accepts utterances containing those items in positions where a plural is required (*I like cats*); its truth-condition-assigning function assigns the appropriate truth-conditions to utterances containing the items in the right column (*I like cats* is true iff I like members of the cat group—it has nothing to do with whether I like members of the dog group).

II. Add –s to everything, except for these exceptions:

dɑg	dɑg**z**	kʰaʊ	kʰaʊ**z**
gɹʌb	gɹʌb**z**	mæn	m**ɛ**n
dɪʃ	dɪʃ**əz**	fʊt	f**i**t
fʌd͡ʒ	fʌd͡ʒ**əz**	waɪf	waɪ**vz**
pʰi	pʰi**z**	...	...

III. Add –z to everything, except for these exceptions:

kʰæt	kʰæt**s**
sæk	sæk**s**
dɪʃ	dɪʃ**əz**
fʌd͡ʒ	fʌd͡ʒ**əz**
mæn	m**ɛ**n
fʊt	f**i**t
waɪf	waɪ**vz**
wɪf	wɪf**s**
...	...

*IV. Add –***əz** *after "sibilant" sounds, –s after non-sibilant [–voice] sounds, and –z otherwise, except for these exceptions:*

mæn	m**ɛ**n
fʊt	f**i**t
waɪf	waɪ**vz**
...	...

*IV. Change final /f/ to [v], and then add –***əz** *after "sibilant" sounds, –s after non-sibilant [–voice] sounds, and –z otherwise, except for these exceptions:*

mæn	m**ɛ**n
fʊt	f**i**t
wɪf	wɪf**s**
...	...

17. Descriptive adequacy

A grammar that not only is observationally adequate, but also gives the same treatment to novel utterances that a real speaker of the target language gives is a **descriptively adequate grammar**.

The idea is that a descriptively adequate grammar captures the generalisations that real learners extract from the learning data—I think it makes the most conceptual sense to operationalise this in terms of novel utterances.

In a famous early study of children, Berko (1958) found that English-speaking adults (all highly educated, in her sample) consistently give the following plurals when presented with invented words (pp. 155-158):

wʌg	wʌg**z**	lʌn	lʌn**z**
gʌt͡ʃ	gʌt͡ʃ**əz**	nɪz	nɪz**əz**
kæʒ	kæʒ**əz**	kɹa	kɹa**z**
toɹ	toɹ**z**	tæs	tæs**əz**

❒ Which of the grammars above could be descriptively adequate, given these data?
❒ The adults disagreed about this word—what might we conclude?

heaf	hif**s**, hi**vz**

18. Explanatory adequacy
A theory that, when given a typical set of learning data, returns a grammar that is descriptively adequate, is an **explanatorily adequate theory**.

Obviously, developing an explanatorily adequate theory is a huge challenge!

19. We will adopt the above framework in this course.
First, we will see some techniques for discovering which generalizations speakers have learned—that is, for discovering which, out of all the infinitely many observationally adequate grammars, are the "real" grammars.

Along with these techniques, we will see some consequences of findings on descriptive adequacy: what must grammars be made of in order to express these generalizations?

Next, we'll see work that, without presenting a learning algorithm, tries to discovers what types of grammars are preferred by learners, as a step towards achieving explanatory adequacy.

Finally, we will see recent work that aspires to explanatory adequacy by developing learning algorithms designed to produce descriptively adequate grammars.

1.8 To learn more

- Terminology
    - For a pretty comprehensive English/French glossary of linguistics, see www.sil.org/LINGUISTICS/glossary_fe/ or www.sil.org/LINGUISTICS/glossary_fe/?lang=fr (en français)
- Phonetic symbols
    - The standard symbol set is the one developed by the International Phonetic Association (IPA). At www.arts.gla.ac.uk/ipa/ipa.html you can find a chart of all the symbols and diacritics (www.arts.gla.ac.uk/ipa/IPA_chart_(C)2005.pdf). Peter Ladefoged created a version at http://phonetics.ucla.edu/course/chapter1/chapter1.html that allows you to click and hear the sounds.
    - Other symbol sets are used, however. Pullum & Ladusaw (1986) is a guide to all the symbols the authors were able to gather.
    -

- Phonetics in general
    - The place to start is Ladefoged 2006 (multimedia materials available at http://phonetics.ucla.edu/course/contents.html).
    - After that, if you want to learn more about acoustic phonetics, try Ladefoged 1996.
- Features
    - There is no one set of features that all linguists agree on. But to get a rough consensus view that is internally consistent over a broad range of sounds, see Bruce Hayes's feature spreadsheet: http://www.linguistics.ucla.edu/people/hayes/120a/index.htm#features
    - You can also practice, if you have a Windows machine, with free software called FeaturePad that I wrote a long time ago: http://www.linguistics.ucla.edu/people/hayes/120a/FeaturePad.htm
- Phonology in general
    - My favourite introductions are Kenstowicz & Kisseberth 1979 for the basics and lots of exercises, Gussenhoven & Jacobs 1998 for a quicker and broader overview, and Kenstowicz 1994 for in-depth looks at various pre-Optimality-Theory issues.
    - Dell 1973 is a good introduction in French, and also has lots of French examples. (Exists in English and Japanese translations too.)
    - Goldsmith 1996—a collection of survey articles by various authors—is very useful for reading about specific topics; a new edition is coming out soon.
    - The foundational non-textbook work is Chomsky & Halle 1968 ("SPE").
    - I'll give you reading suggestions for Optimality Theory next time.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Dell, François (1973). *Les règles et les sons: introduction à la phonologie générative*. Paris: Hermann.

Goldsmith, John (ed.) (1996). *The handbook of phonological theory*. Cambridge, MA: Blackwell.

Gussenhoven, Carlos & Haike Jacobs (1998). *Understanding phonology*. London: Arnold.

Kenstowicz, Michael (1994). *Phonology in generative grammar*. Oxford: Blackwell.

Kenstowicz, Michael & Charles Kisseberth (1979). *Generative phonology*. San Diego: Academic Press.

Ladefoged, Peter (2006). *A course in phonetics*. Boston: Thomson/Wadsworth. 5th edition.

Ladefoged, Peter (1996). *Elements of acoustic phonetics*. Chicago: University of Chicago Press. 2nd edition.

Pullum, Geoffrey & William Ladusaw (1986). *Phonetic symbol guide*. Chicago: University of Chicago Press.

And of course, you may want to read some of the works cited today...

1.9 References

Albright, Adam (2006). Why eatees are not E.T.'s: Blocking of aspiration by output-output constraints. 14th Manchester Phonology Meeting, May 25-27.

Berko, Jean (1958). The child's learning of English morphology. *Word* 14: 150-177.

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

Dixon, R.M.W. (1991). Mbabaram. In R.M.W. Dixon & Barry J. Blake (eds.) *The handbook of Australian Languages. Vol 4: The aboriginal language of Melbourne and other grammatical sketches*. Oxford: Oxford University Press. Pp. 349-402.

Embick, David & Rolf Noyer (2001). Movement operations after syntax. *Linguistic Inquiry* 32: 555-595.

Kang, Yoonjung (2007). Frequency effects and regularization in Korean nouns. Talk given at the Workshop on variation, gradience, and frequency in phonology, Stanford University.

(Handout downloadable from www.stanford.edu/dept/linguistics/linginst/nsf-workshop/workshop-july-2007.html)
Kisseberth, Charles (2000). The phonology-syntax interface: Chimwiini revisited. Ms.
Raffelsiefen, Renate (1996). Gaps in word formation. In Ursula Kleinhenz (ed.) *Interfaces in phonology*. Berlin: Akademie Verlag. Pp. 194-209.
Raffelsiefen, Renate (1999). Phonological constraints on English word formation. In Geert Booij & Jap van Marle (eds.) *Yearbook of morphology 1998*. Dordrecht: Kluwer. Pp. 225-287.

Day 2: Descriptive adequacy: methods and consequences

1. Recall from last time

An **observationally adequate grammar** accepts the utterances that a typical speaker has been exposed to.
A **descriptively adequate grammar** behaves the same as a typical speaker when confronted with novel utterances.
An **explanatorily adequate theory** generates a descriptively adequate grammar, given a set of typical learning data.

2. Why is it hard to develop a descriptively adequate theory in phonology?

Words that the speaker already knows are uninformative! (They don't tell us anything about what generalizations the speaker has learned—she may have simply memorized that word.)

Constructing novel phonological situations to put speakers in is difficult. Contrast this with syntax, where it's easy to construct sentences that—presumably—the speaker has not encountered before.

2.1 The "wug" test: have subjects perform operations on invented words

3. Named after one of Berko's (1958) stimuli

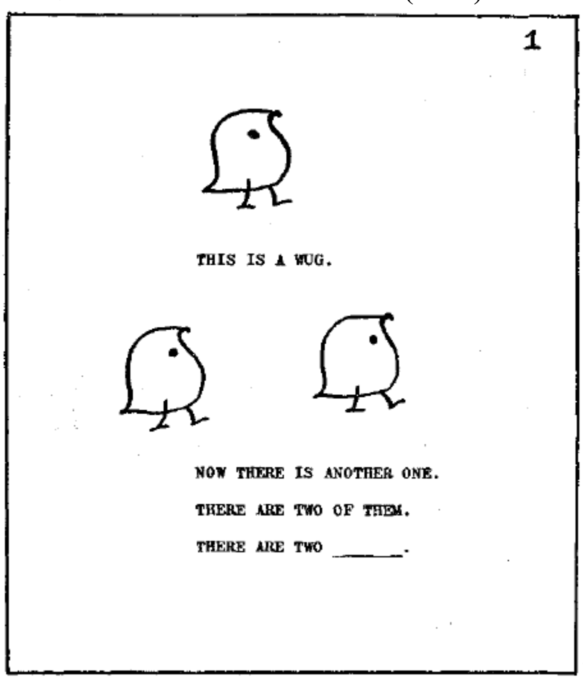


Figure 1. The plural allomorph in /-z/.

(p. 154)

4. Consequences of *wug*-test results

Recall Berko's results from last time—adults can do the task, applying appropriate forms of the plural suffix. (She looked at other morphology too.)

==> A grammar can contain more than just a list of memorized forms.
This supports the use rules like what we saw last time (A → B / X__Y)

❒ Let's write some rules for the basic realizations of the English (regular) plural, using the data below—some of the "wug" words come from Berko's study, others are hypothetical ones that I made up (but that are supported by my own intuitions)

	singular	*plural*	*similar "wug" word sg.*	*"wug" pl.*
cat	kʰæt	kʰæt**s**	blut	blut**s**
sack	sæk	sæk**s**	plɪk	plɪk**s**
dog	dɑg	dɑg**z**	wʌg	wʌg**z**
grub	gɹʌb	gɹʌb**z**	sæb	sæb**z**
fuss	fʌs	fʌs**əz**	tæs	tæs**əz**
watch	wɑt͡ʃ	wɑt͡ʃ**əz**	gʌt͡ʃ	gʌt͡ʃ**əz**
pea	pʰi	pʰi**z**	kɹa	kɹa**z**
cow	kʰaʊ	kʰaʊ**z**	splu	splu**z**

2.2 Novel words from other languages

5. Loan adaptation as natural *wug*-test

What do speakers do with words imported from other languages (**loan adaptation)**, or when learning other languages (**second-language phonology**, or **L2 phonology**)?

This constitutes a sort of "natural *wug* test"—the context is less controlled (who did they first hear the word from? do they know the spelling in the original language? how well do they speak the foreign language? are there established conventions for borrowing words from this language?).

6. Background: native Russian words

Kenstowicz & Kisseberth 1979, p. 46

dative sg.	*nominative pl.*	*nominative sg.*	
xlebu	xleba	xlep	*'bread'*
gribu	griby	grip	*'mushroom'*
grobu	groby	grop	*'coffin'*
čerepu	čerepa	čerep	*'kull'*
xolopu	xolopy	xolop	*'bondman'*
trupu	trupy	trup	*'corpuse'*
sadu	sady	sat	*'garden'*
prudu	prudy	prut	*'pond'*
cvetu	cveta	cvet	*'color'*
zakatu	zakaty	zakat	*'sunset'*
razu	razy	ras	*'time'*
zakazu	zakazy	zakas	*'order'*
lesu	lesa	les	*'forest'*
usu	usy	us	*'whisker'*
storožu	storoža	storoš	*'guard'*
dušu	dušy	duš	*'shower'*
rogu	roga	rog	*'horm'*
porogu	porogy	porog	*'threshold'*
raku	raky	rak	*'crayfish'*
poroku	poroky	porok	*'vice'*

❒ Write a rule for the alternation that you see in the stems (don't worry about the two different nom. pl. suffixes).

7. A couple of remarks

By the way, not all Russian consonants undergo this rule: the consonants illustrated here are all **obstruents**, sounds articulated with a total stoppage of airflow out the vocal tract (/p,b,t,d,k,g/) or turbulent airflow because the air is forced through a narrow opening in the mouth (/s,z/, "š" (IPA /ʃ/), "ž" (IPA /ʒ/).

This example brings up a new point about phonemes and allophones: two different phonemes can become **neutralized** in some contexts: in Russian, /p/ is realized typically as [p], but /b/ also has an allophone [p].

8. Productivity in loans

Now, K&K report that words borrowed into Russian behave the same way (p. 53):

dative	*nominative*	
garažu	garaš	'garage'
gazu	gas	'gauze'
klubu	klup	'club'

Moreover, final devoicing can be seen in a typical Russian accent when speaking English (p. 53).

9. New sounds

What's interesting about foreign words is that we can ask what would happen with sounds that don't exist in Russian.

Here is the inventory of obstruents in Russan (K&K p. 337):

stops (no airflow)	p	b	t	d		k g
fricatives (turbulent airflow)	f	v	s	z	š (IPA ʃ) ž (IPA ʒ)	x
affricates (stop released into fricative)			c (IPA t͡s)		č (IPA t͡ʃ)	

Missing are /dz/ (voiced equivalent of /c/), /ǰ/ (IPA d͡ʒ, voiced equivalent of /č/), and /ɣ/ (voiced equivalent of /x/). So what do Russian learners typically do with these sounds?

K&K report (p. 337), for speakers who have already mastered /ǰ/ in other environments,

badge	ba[č]
judge	[ǰ]u[č]

Findings like this—not that much research into this question exists, actually!—support the **features** that we saw last time.

10. Features

We would like to say that Russian speakers have learned not {b,d,g,v,z,ž} → [–voice] / __ #
but something like {obstruents} → [–voice] / __ #

(or, more standardly, [–sonorant] → [–voice] / __ # ; [sonorant] is the feature that distinguishes the obstruents from the **sonorants**, which are just all sounds besides obstruents.)

11. Another example of loan adaptation (Shibatani 1973)

Japanese doesn't allow words to end in consonants, except /ɴ/. More generally, Japanese requires any consonant (C) that has its own "place of articulation" (see below) to be followed by a vowel.

In native words, this is never a problem: only verb roots end in Cs, and they always take suffixes.[8]

	suffix begins with V anyway		*root-final C becomes V, /ɴ/, or part of long C*	*depending on analysis: suffix begins with V, or suffix's initial C deletes*	
UR	*pres. polite*	*negative*	*past*	*present*	
/mat/	matʃ-imasu	mat-anai	mat-ta	mats-u	'wait'
/kak/	kak-imasu	kak-anai	kai-ta	kak-u	'write'
/aruk/	aruk-imasu	aruk-anai	arui-ta	aruk-u	'walk'
/job/	job-imasu	job-anai	jon-da	job-u	'call'
/asob/	asob-imasu	asob-anai	ason-da	asob-u	'play'
/isog/	isog-imasu	isog-anai	isoi-da	isog-u	'hurry'
/hanas/	hanaʃ-imasu	hanas-anai	hanaʃ-ita	hanas-u	'speak'
/nom/	nom-imasu	nom-anai	non-da	nom-u	'drink'
/kaer/	kaer-imasu	kaer-anai	kaet-ta	kaer-u	'return'
/gambar/	gambar-imasu	gambar-anai	gambat-ta	gambar-u	'hang in there'
/tabe/	tabe-masu	tabe-nai	tabe-ta	tabe-ru	'eat'
/mise/	mise-masu	mise-nai	mise-ta	mise-ru	'show'
/mi/	mi-masu	mi-nai	mi-ta	mi-ru	'see'
/deki/	deki-masu	deki-nai	deki-ta	deki-ru	'can'

But in loanwords, offending consonants are treated consistently:

peɴ	'pen'
doresu	'dress'
sukuriputo	'script'

12. Interpretation

Shibatani argues that Japanese speakers must have learned,[9] on the basis of the words of their language, that non-/ɴ/ consonants are illegal unless followed by a vowel.[10]

He further argues that what speakers had reason to learn was not a rule for inserting vowels (because they would have had no evidence for such a rule) but simply a prohibition on these illegal sequences:

$$*\begin{bmatrix}+\text{consonantal}\\ \begin{Bmatrix}-\text{nasal}\\ +\text{place}\end{Bmatrix}\end{bmatrix}\begin{Bmatrix}+\text{consonantal}\\ \#\end{Bmatrix}$$

= "don't have a consonant that either is non-nasal or has place [this is to exempt /ɴ/] followed by a consonant or the end of the word".

---

[8] This is probably a simplification. There have been many different proposals about Japanese verbs—what their underlying forms are and what rules apply. But I think it is close to what Shibatani has in mind.

[9] Or retained, under some theories. See, e.g., Stampe 1979.

[10] He's actually making a subtler point: that Japanese speakers must have learned a surface constraint rather than a rule or a constraint on underlying forms.

Place is a cover term for any constriction occurring in the mouth. /N/ tends to get realized with the place of the following consonant or, before a pause, nasalization of the preceding vowel or a uvular nasal—thus, it has no place of its own and is [–place].

The **curly brackets** { } mean "or".

How the speech community agreed that this prohibition should be obeyed by inserting vowels rather than through, say, deleting consonants, is not resolved here.

2.3 Replacing rules with constraints

13. Constraints

In the 1970s, more and more cases built up in which constraints (forbidden or avoided surface structures) seemed to play a larger role than rules (operations). This led to increasing use of constraints throughout the 1980s, but with no real framework for how they should work.

- How does the grammar indicate which rule to apply to correct a constraint violation? This is especially an issue if the grammar contains multiple rules that could do the job.
- What if two constraints conflict?
- Relatedly, why are some constraints always obeyed and others obeyed only sometimes?
- Can the grammar look ahead to the result of applying a rule to see if it would violate a constraint?

14. Optimality Theory

Optimality Theory (Prince & Smolensky 1993/2004) presented a framework for constraint interaction that eliminated rules entirely (or at least, made their role trivial).

- All possible rules (delete, insert, replace, maybe others too) are applied to the underlying representation (also called **input**) to create a set of (infinitely many) candidate surface forms, also called **output candidates**
  - /skript/ → {skript, sukuriputo, ki, ...}
- A list of constraints chooses the best candidate(s)—we can use tricks like manipulating regular expressions to deal with the infinite size of the candidate set.
- The highest-ranked constraint picks out the candidates that do best on it, then passes those to the next constraint (etc.).
- Some of the constraints are **markedness constraints** (like the constraint above), which penalize certain surface structures. Others are **faithfulness constraints**, requiring the underlying form (input) and surface form (output) to be similar in various ways.
- A **tableau** is a common device for illustrating the competition among candidates—obviously, only a finite subset of the candidates can be listed.

/skript/	$*\begin{bmatrix}\text{+consonantal}\\ \begin{Bmatrix}\text{–nasal}\\ \text{+place}\end{Bmatrix}\end{bmatrix}\begin{Bmatrix}\text{+consonantal}\\ \#\end{Bmatrix}$	DON'T DELETECS	DON'T CHANGECS	DON'T ADDVS
☞ *a* [skript]	*!***			
b [sukuriputo]				****
c [ki]		*!***		
d [kiN]		*(!)**	*(!)	
e [skuriputo]	*!			***
f [skripto]	*!**			*

15. Reading a tableau

- A pointing finger (☞), arrow, or something similar indicates the winning candidate—this is the actual output (surface form).
- Highest-ranked constraints are on left.
- The number of asterisks (*) in a cell indicates how many times that candidate violates that constraint.
- An exclamation mark goes after an asterisk that takes a candidate out of the competition.
- A cell is shaded to indicate that it doesn't matter—how well or badly that candidate does on that constraint won't affect the outcome (because the candidate has already been ruled out or the winner has already been chosen).
- The names of constraints that I've used above aren't standard.
- The dashed line above mean we can't determine the ranking of the top three constraints—for all the data that we have, the result is the same regardless of the ranking.
- Not everyone does this, but I put some exclamation marks in parentheses: we don't know which violation takes candidate *d* out of the competition, because we don't know the ranking of the constraints.

If you want to read phonological literature, you'll need familiarity with (at least!) both rules and OT.

16. Another example with OT
Let's try Russian final devoicing together.

2.4 Learn more

- Issues in interpreting loan adaptation
  - See Peperkamp (in press) and the papers cited therein.
- Optimality Theory (OT)
  - The original proposal is in Prince & Smolensky 1993/2004.
  - Kager 1999 is a textbook on OT.
  - Of course, OT is not the only way that constraints could interact! See the references in the first chapter of Prince & Smolensky for alternatives. See especially harmonic grammar, which you can read about in Smolensky & Legendre 2006.
  - How can we deal, algorithmically, with the infinite candidate set? Most of the work in this area uses finite-state automata. A good reading list would be Eisner 1997 (generation), Eisner 2002 (comprehension), Riggle 2004 (various new algorithms, including a way to make the candidate set finite), Albro 2005 (application of computationally implemented OT to the phonology of a whole language [Malagasy]).
  - For useful software, visit Jason Riggle's Language Modeling Lab (clml.uchicago.edu) and Bruce Hayes's OTSoft page (www.linguistics.ucla.edu/people/hayes/otsoft)—OTSoft is more user-friendly but less flexible.

Albro, Daniel (2005). Studies in computational Optimality Theory, with special reference to the phonological system of Malagasy. UCLA dissertation. (available at www.linguistics.ucla.edu/people/grads/albro/diss.pdf)

Eisner, Jason. 1997. Efficient generation in primitive Optimality Theory. Proceedings of the ACL 35.

Eisner, Jason. 2000. Directional constraint evaluation in Optimality Theory. In COLING 2000.

Kager, René (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Peperkamp, Sharon (in press) A psycholinguistic theory of loanword adaptations. *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*, held in 2004. (available at www.ehess.fr/centres/lscp/persons/peperkamp/BLS30.pdf)

Prince, Alan, and Paul Smolensky (1993/2004). *Optimality Theory: constraint interaction in generative grammar.* Originally a ms., Rutgers University and University of Colorado, Boulder. Later published by MIT Press. (available at roa.rutgers.edu/view.php3?id=845)
Riggle, Jason (2004). Generation, recognition, and learning in finite-state Optimality Theory. UCLA dissertation. (available at hum.uchicago.edu/~jriggle/riggleDiss.html)
Smolensky, Paul & Géraldine Legendre (2006). *The Harmonic Mind: From Neural Computation To Optimality-Theoretic Grammar.* Vol. 1: Cognitive Architecture; Vol. 2: Linguistic and Philosophical Implications. MIT Press.

2.5 References

Berko, Jean (1958). The child's learning of English morphology. *Word* 14: 150-177.
Kenstowicz, Michael & Charles Kisseberth (1979). *Generative phonology*. San Diego: Academic Press.
Prince, Alan, and Paul Smolensky (1993/2004). *Optimality Theory: constraint interaction in generative grammar.* Originally a ms., Rutgers University and University of Colorado, Boulder. Later published by MIT Press.
Shibatani, Masayoshi (1973). The role of surface phonetic constraints in generative phonology. *Language* 49: 87-106.
Stampe, David. (1979). *A dissertation on natural phonology.* New York: Garland.

Day 3: Methods for investigating explanatory adequacy

1. Review

An **observationally adequate grammar** accepts the utterances that a typical speaker has been exposed to.
A **descriptively adequate grammar** behaves the same as a typical speaker when confronted with novel utterances.
An **explanatorily adequate theory** generates a descriptively adequate grammar, given a set of typical learning data.

Last time we saw some methods that have been used to try to discover what generalizations speakers have extracted from their learning data, and thus to achieve descriptive adequacy.

2. Moving towards explanatory adequacy

We saw evidence that English speakers know generalizations like "use the [əz] form of the plural after sibilants". So, our grammar for English should include that somehow.

This also tells us something about the theory of learning: learners must be capable of extracting that regularity—for the English data, they prefer a grammar with "use the [əz] form of the plural after sibilant" to a grammar without it.

But it doesn't tell us anything about whether learners prefer a generalization like "use the [əz] form of the plural after sibilants" to a generalization like "use the [əz] form of the plural after *non*-sibilants", because the second generalization is not available in the English data.

How can we discover what generalizations are preferred (if any!)?

3. Typology?

Chomsky & Halle 1968 ("SPE") proceed according to this logic:

- Assume that languages change because members of one generation learn a slightly different grammar from the grammar that generated the data they were exposed to.
- Further assume that these changes involve learners constructing a more-preferred grammar than what would be strictly consistent with the learning data.
- Therefore, if a certain phonological phenomenon is predominant cross-linguistically, it must be because learners prefer it (and therefore have introduced it into many languages).
- Thus, we can tell what learners prefer by inspecting cross-linguistic tendencies.

For example, if there are lots of languages that have both voiced and voiceless obstruents at the ends of words, and lots of languages that devoice obstruents at the ends of words, but no (or almost no) languages that voice obstruents at the ends of words, then final devoicing must be preferred over final voicing.

4. Problems with this view

See Blevins 2004 and references therein—Ohala (1992 and many others) is also an important author on this point.
Blevins's main points (for our purposes) are

- There are examples of languages that run counter to the cross-linguistic tendency in some way. This means that the tendency reflects at most a preference among learners, not the impossibility of the counter-typological pattern (we would not want a theory that completely rules out word-final voicing if some languages have it—see Yu 2004 for a language that, though a series of diachronic changes, ended up with final voicing).

- Cross-linguistic tendencies can often be explained without recourse to learner preferences. This doesn't mean we can conclude that learners don't have preferences, just that typology alone isn't evidence for such preferences.

The clearest cases of explaining cross-linguistic tendencies without recourse to learner preferences come from patterns that plausibly result from learners simply mis-hearing sounds.[11]

Suppose that the acoustic cues to obstruent voicing are weaker in final position than elsewhere because of the lack of a following vowel or sonorant. (This is not the reason usually cited for final devoicing, but it's plausible: an obstruent's voicing affects the pitch of the following sound, and "voicing" is often really lack of delay in voicing the following sound.)

Speaker produces /ab/ → [ab] 'cat', but learner hears [ap]
Speaker produces /ab+o/ → [abo] 'cats', and learner hears [abo]

=> Learner has (mis)learned that there is an alternation *ap ~ ab-o*.

But nothing in this scenario requires the learning function to prefer final devoicing.
The learner could be just as capable of learning an alternation [ip+i], [ib] (final voicing)—but because there is no reason to mis-hear things in this way, learners would be very unlikely to innovate a final-voicing pattern. Thus, we see little final voicing cross-linguistically.

5. Investigating learner preferences

It's a computational necessity that the learner have *some* preferences/biases, even if they are very boring. For example, the learner might be hard-wired to only memorize lists of word-pairs (*cat-cats*, *dog-dogs*, etc.). Or, the learner might require a certain number of examples before extracting a generalization.

But does the learner have any "substantive" biases—biases that treat sounds not just as algebraic symbols but as having physical content—for instance preferring...

- sounds that are easy to produce
- contrasts to be realized in environments where they're easy to hear
- sound sequences that are easy to plan

?

For the remainder of today, we'll review research into this question.

3.1 Processing of native-language rules (Zhang & Lai submitted)

6. Mandarin tone sandhi

Chinese languages often display a phenomenon known as **tone sandhi**: when two syllables are put together into a word, their tones change:

[11] Blevins discusses two other mechanisms, but it is less clear to me how they can account for cross-linguistic tendencies—rather than random drift—without speaker or learner preferences.

(2) Mandarin tone sandhi:

a. 213 → 35 / ___ 213

xaw213-tɕju213 → xaw35-tɕju213	'good wine'
tʂan213-lan213 → tʂan35-lan213	'exhibit'

b. 213 → 21 / ___ {55, 35, 51}

xaw213-ʂu55 → xaw21-ʂu55	'good book'
xaw213-ɻən35 → xaw21-ɻən35	'good person'
xaw213-kʰan55 → xaw21-kʰan51	'good-looking'

(Zhang & Lai p. 80)

The numbers represent pitches on a scale from 1 (lowest) to 5 (highest).

7. Naturalness

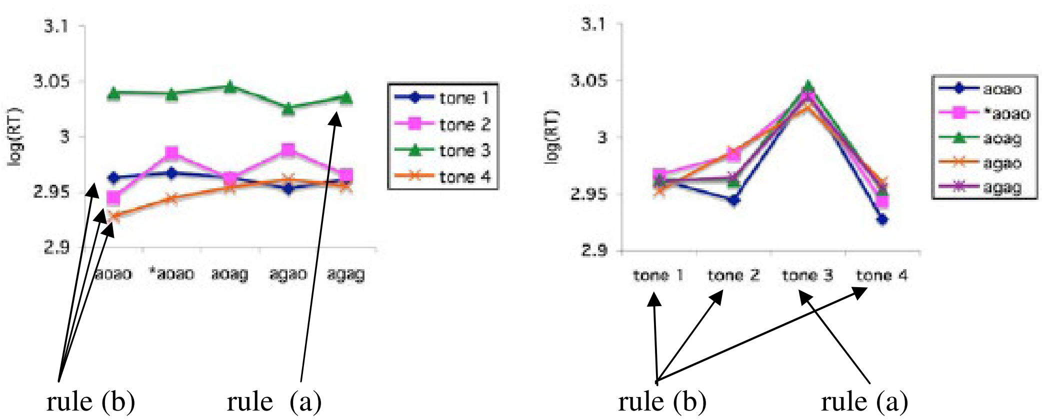
There are various reasons to think that rule (b) should be "better" than rule (a):

- Both rules simplify a complex contour, so that it is easier to realize in a shorter time (being nonfinal makes the first word shorter)—see Zhang 2002, 2004.
- But (a) also involves raising of pitch, which increases articulatory demands in a short time.
- (b), on the other hand, involves straightforward simplification of the original tone
- (Zhang & Lai discuss other reasons that are too advanced to get into here)

Mandarin speakers use both rules very frequently—but is (b) nevertheless "easier" than (a)?

8. Experimental results

Zhang & Lai presented Mandarin speakers with a variety of real and "wug" combinations. Subjects hear the two syllables and had to pronounce them as a single word.



(Zhang & Lai p. 96)

Subjects responded more slowly (higher values on vertical axis) when applying rule (a), for all types of words (real and "wug"). (There are other interesting results concerning how the words were produced—see the paper.)

Conclusion: Mandarin speakers have learned both rules, but have more difficulty using the "unnatural" one.

3.2 Literary invention: imperfect rhyme (Kawahara 2007)

9. Perceptual similarity

Cross-linguistically widespread phonological phenomena often seem to privilege perceptual similarity (Steriade 2001):

- A contrast is likely to be abandoned in a context where the two phonemes sound too similar (potential example: contrast between voiced and voiceless consonants likely to be abandoned in word-final position)
- Changes needed to satisfy a markedness constraint tend to be perceptually minimal: if there's a constraint $*\begin{Bmatrix}+\text{voice}\\ -\text{sonorant}\end{Bmatrix}\#$, it could be satisfied in many ways, but seems always to be satisfied by devoicing:

/ab/	$*\begin{Bmatrix}+\text{voice}\\ -\text{sonorant}\end{Bmatrix}\#$	DON'T DELETECs	DON'T CHANGE VALUEOF [nasal]	DON'T ADDVs	DON'T CHANGE VALUEOF [voice]
a [ab]	*!				
☞ *b* [ap]					*
c [am]			*!		
d [abə]				*!	
e [a]		*!			

(I'm assuming that [ap] sounds more like [ab] than do [am], [abə], and [a].

- ...similarity plays other roles too, that we won't consider here.

10. Knowledge of perceptual similarity?

To test the claim that these are active learner preferences, rather than results of differential mishearing, we need to establish, at least, that speakers have knowledge of perceptual similarity. (Following work by Steriade 2003 on Romanian imperfect rhymes)

Kawahara gives evidence from imperfect rhymes in Japanese rap lyrics:

(2) *Mastermind* (DJ HASEBE feat. MUMMY-D & ZEEBRA)

a. kettobase **kettobase**
kick it kick it
'Kick it, kick it'

b. kettobashita kashi de **gettomanee**
funky lyrics with get money
'With funky lyrics, get money'

(Kawahara p. 115)

The claim is that ([k],[g]), ([b],[m]), ([s],[n]) are relatively similar.

Overall, sounds that belong to more natural classes together occur more often in rhymes:

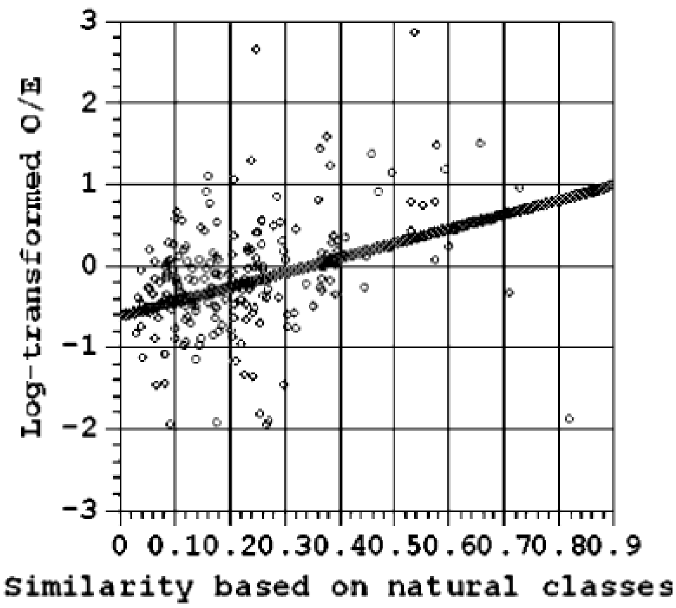


Fig. 2 The correlation between rhymability and similarity calculated based on the number of shared natural classes. The line represents a linear regression line (Kawahara p. 121)

In particular, sounds that have the same value for [palatal], [continuant], [voice], and [nasal] are especially likely to rhyme.
(There are further interesting results about context.)

Literary invention is attractive to study because more- and less-similar rhymes are all legal within the genre but rap-lyric composers (and other poets) seem drawn to rhymes that are more similar—this preference is presumably not something they have learned explicitly.

3.3 Literary invention: alliteration (Minkova 2001, 2003)

11. Cluster splittability
There is diverse evidence that languages treat initial *sp*, *st*, *sk* differently from other word-initial consonant sequences like *pl*, *kr*.

Sp, *st*, *sk* behave as "less splittable"—see Fleischhacker 2002a,b for various types of evidence.

Most famous type of evidence: loanword adaptation and second-language phonology

Farsi: ***e**sparta* 'Sparta' vs. *p**e**lutus* 'Plutus' (Fleischhacker 2002a)

Is there a real preference for grammars that don't split *s{p,t,k}*, or is it just a matter of mis-hearing or mis-articulation?

12. Alliteration
Minkova gives evidence that speakers have knowledge of this splittability from **alliteration**, a poetic device common in Middle English (and many other traditions):

CONTIGUITY in OE (*sp-*, *st-*, *sk-*)

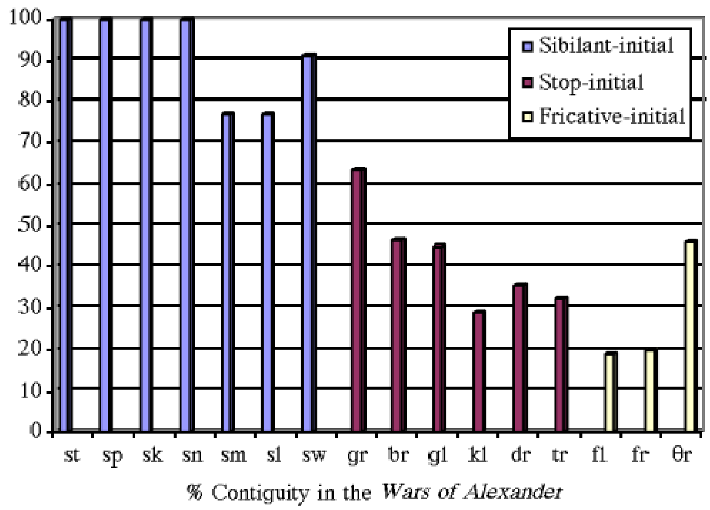
scaðan scirhame / to scipe foron[1]	*Beo* 1895	
stopon styrnmode, / stercedferhðe[2]	*Judith* 227	
and πæt spere sprengde, / πæt hit sprang ongean[3]	*Maldon* 137	(Minkova 2001 p. 1)

When words that start with 2 or more consonants alliterate, sometimes the whole consonant sequence is repeated (*sc...sc...*), but sometimes poets allow C_1C_2 to alliterate with just C_1 (*sl...s...*):

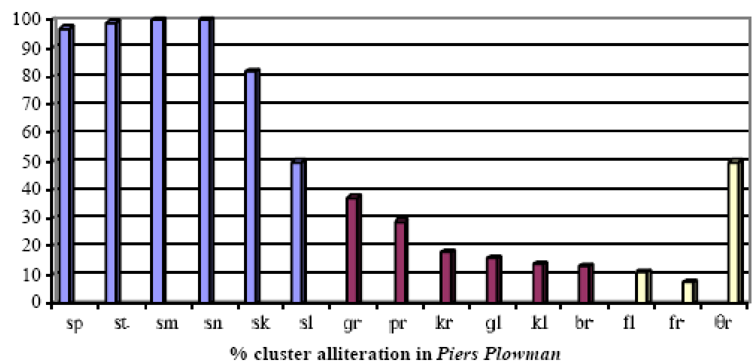
∂urh sliπne niπ / sawle bescufan[4]	*Beo* 184
druncen 7 dolhwund. / Næs ∂a dead πa gyt[5]	*Judith* 107
πe πær baldlicost / on πa bricge stop[6]	*Maldon* 78

(Minkova 2001 p. 1)

13. Cluster differences in alliteration
Minkova examines how often different clusters are kept together, and finds that *s{p,t,k}* are kept together more than other sequences.



(Minkova 2001 p. 3)



(Minkova 2001 p. 6)

As with Japanese imperfect rhyme, this is interesting because it results from poets' exercising their freedom within a tradition that allows both kinds of alliteration.

3.4 Artificial language learning

14. Artificial language learning
This is a burgeoning field: experimenters expose subjects (infant, child, or adult) to the output of some invented grammar, and then try to discover what the subjects have learned.
The studies that are relevant here are those that compare two different grammars to see which is easier for subjects to learn.

Some of this research is conducted locally at the LSCP (www.ehess.fr/centres/lscp/index.html)—Peperkamp & Dupoux (in press); Peperkamp, Skoruppa & Dupoux (2006).

15. An example where the phonology is simple: Wilson 2003
Two grammars:

(1) a. Consonant harmony

dume	dumena	tuko	tukola	suto	sutola
binu	binuna	dige	digela	dabu	dabula

b. "Random" alternation

dume	dumela	tuko	tukona	suto	sutola
binu	binula	dige	digena	dabu	dabula

(Wilson p. 101)

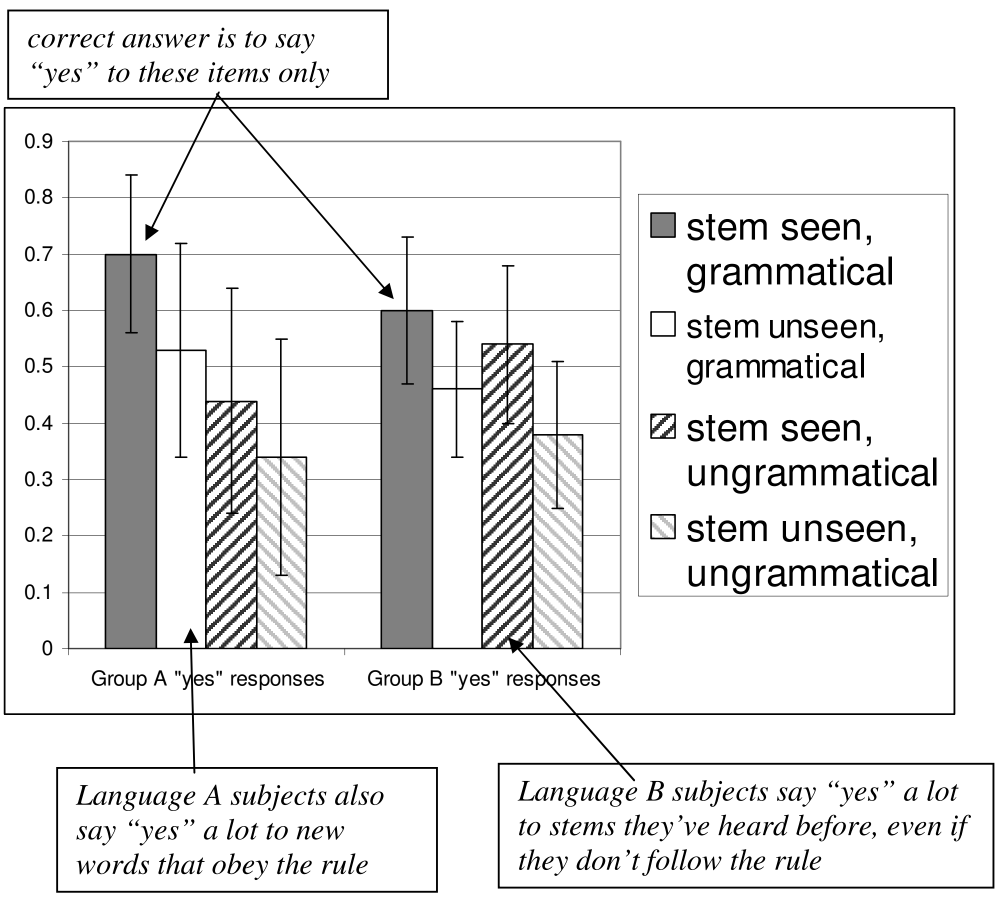
In Language A, the "suffix" appears as *–na* iff the preceding consonant is also [+nasal].
In Language B, the suffix appears as *–na* iff the preceding consonant is [velar] (*k* or *g*).

Language A is more typical of what we see cross-linguistically.

16. Experimental results
Wilson exposed some subjects to words from Language A and other subjects to words from Language B.

Subjects were then presented with new words and asked whether they had seen them in the previous portion of the experiment.

=> Subjects exposed to Language A—the "natural" one—were more likely to say "yes" to words that they hadn't actually seen before, but that followed the pattern.



=> Language A speakers have learned the rule better.

3.5 Conclusions

Though research into which grammars are "better" is still somewhat scarce, we're starting to get some evidence, from various sources, about

- what people know (independent of the direct evidence from their language)
- how people apply that knowledge so that some grammars are preferred over others.

3.6 References

Chomsky, Noam, and Morris Halle 1968. *The sound pattern of English*. Cambridge, MA: MIT Press.

Blevins, Juliette (2004). *Evolutionary phonology: the emergence of sound patterns.* Cambridge: Cambridge University Press.

Fleischhacker, Heidi 2002a. Cluster-dependent epenthesis asymmetries. UCLA Working Papers in Linguistics 7, Papers in Phonology 5.71-116.

Fleischhacker, Heidi 2002b. Onset transfer in reduplication. Los Angeles: UCLA ms.

Kawahara, Shigeto (2007) Half rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics (JEAL)* 16.2.

Minkova, Donka (2001). Testing CONTIGUITY in Middle English alliteration. Handout from talk presented at the 15th ICHL, Melbourne. (available at www.english.ucla.edu/faculty/minkova/Handout_Melbourne.pdf)

Minkova, Donka 2003. Alliteration and sound change in Early English verse. Cambridge Studies in Linguistics 101. Cambridge: Cambridge University Press.

Peperkamp, S. and Dupoux, E. (in press). Learning the mapping from surface to underlying representations in an artificial language. In J. Cole & J. Hualde (eds.) *Laboratory Phonology IX*.

Peperkamp, S., K. Skoruppa & E. Dupoux (2006) The role of phonetic naturalness in phonological rule acquisition. In: D. Bamman, T. Magnitskaia & C. Zaller (eds.) *Proceedings of the 30th Annual Boston University Conference on Language Development*. Somerville, MA : Cascadilla Press, 464-475.

Steriade, Donca 2001. The phonology of perceptibility effects: the P-map and its consequences for constraint organization. Cambridge, MA: MIT, ms.

Steriade, Donca 2003. Knowledge of perceptual similarity and its uses: evidence from half-rhymes. In Solé, Recasens & Romero, 363-366.

Wilson , Colin (2003). Experimental investigation of phonological naturalness. In *Proceedings of WCCFL 22*.

Yu, Alan (2004). Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80: 73-97

Zhang, Jie & Yuwen Lai (submitted). Testing the role of phonetic naturalness in Mandarin tone sandhi. Also in *Kansas Working Papers in Linguistics*, Vol. 28 (2006): 65-126. (available at www2.ku.edu/~ling/faculty/Dr_Zhang/wug-mandarin-KWPL-2006.pdf)

Day 4: Towards explanatory adequacy

1. Review

An **observationally adequate grammar** accepts the utterances that a typical speaker has been exposed to.
A **descriptively adequate grammar** behaves the same as a typical speaker when confronted with novel utterances.
An **explanatorily adequate theory** generates a descriptively adequate grammar, given a set of typical learning data.

- We've seen some methods that have been used to try to discover what generalizations speakers have extracted from their learning data, and thus to achieve descriptive adequacy.
- We've seen methods for trying to discover what makes one grammar better than another—out of all the observationally adequate grammars imaginable, why do learners pick the ones they do (the descriptively adequate grammars)?

To end, we'll look at some ideas of what a learning function might look like, so that it can display the same grammar preferences that real learners do.

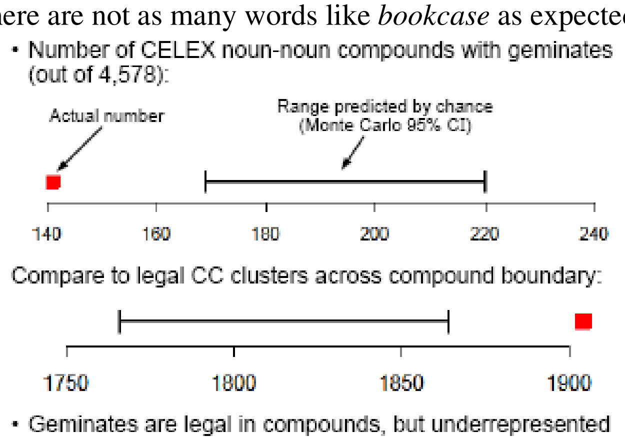
2. Chomsky & Halle 1968 ("SPE")

- SPE did not propose an explicit learning procedure.
- It did, however, propose that the learner should choose the "simplest" grammar consistent with the learning data (or actually, as we'll see, the learner should prefer simplicity sometimes even at the expense of accuracy).
- Thus, learners' choices (as reflected, it was hoped, in typology), would give evidence as to what counts as simple.
- This can render the notion "simplicity" circular, so maybe better to just say that the learner chooses the "best" grammar, and studying the learner's choices will tell us what counts as good.

4.1 Trading off goodness against accuracy: Martin 2007a

3. Facts to be accounted for

- English does not allow double consonants ("**geminates**") within a morpheme: there can be no minimal pair [hæpi]/[hæppi].
- English does allow double consonants in compounds and affixed words: *no[nn]egotiable*, *sou[ll]ess*, *boo[kk]ase*.
- Martin discovered, however, that geminates are less common than would be expected by chance—that is, there are not as many words like *bookcase* as expected:



(Martin 2007b)

Martin discovered similar compound underrepresentation for sibilant harmony in Navajo and vowel harmony in Turkish.

4. Martin's approach

It's easy to construct a learner that can learn these facts.
What Martin set out to do was construct a learner that, presented with no bias in compounds, will learn a bias anyway.

5. Martin's toy language—contains only two sounds

- The training data consists of biconsonantal clusters of [p] and [t], with an optional morpheme boundary:

Cluster	Structure	Number of examples
pt	monomorpheme	2000
tp	monomorpheme	2000
p+t	compound	1000
t+p	compound	1000
p+p	compound	1000
t+t	compound	1000

No bias in training data

- Tautomorphemic geminates [pp], [tt] do not occur in training data, but heteromorphemic geminates occur freely

(Martin 2007b)

6. Constraints available to learner

Structure-sensitive constraints:

***pp** no geminates within morpheme
***tp** no non-geminate clusters within morpheme
***p+p** no geminates across morpheme boundary
***t+p** no non-geminate clusters across morpheme boundary

Structure-blind constraints:

***p(+)p** no geminates
***t(+)p** no non-geminate clusters

(Martin 2007b)

7. Grammar uses Maximum Entropy with a smoothing term

(See Goldwater & Johnson 2003, Hayes & Wilson accepted and references in both).

Each constraint has a numerical **weight**. A word's **acceptability score** is determined by which constraints it violates, and how heavily weighted those constraints are (it's a lot like OT).

	*pp weight = 4.01	*tp weight=0.13	*p(+)p weight=.03	*t(+)p weight=.00	*p+p weight=.00	*t+p weight=.00	score
a pp	*		*				$e^{-4.04}$=0.02
☞ *b* tp		*		*			$e^{-0.13}$=0.87
c p+p			*		*		$e^{-0.04}$=0.96
d t+p				*		*	$e^{-0.00}$=1.00

- *pp* gets a low score, as expected—because *pp has a big weight
- *tp* gets a high score, as expected—because *tp has a small weight
- *t+p* gets a high score, as expected
- but *p+p* gets a slightly lower score—because *p(+)p has a non-negligible weight

8. Learning

How does the learner assign the weights? It attempts to maximize the following equation, which trades off accuracy against "goodness"—goodness, here, is how much the weights deviate from zero:

$$\sum_{i=1}^{N} \log P(x_i) - \sum_{j=1}^{M} \frac{w_j^2}{2\sigma_j^2}$$

Probability of data — Smoothing term

The smoothing term penalizes high constraint weights. This is necessary to avoid overfitting the training data.

(Martin 2007b)

Because the **smoothing term** uses w^2, the squared values of the weights, it's better to account for data like the absence of *pp* by spreading the "blame" over two constraints—*pp and *p(+)p—than by loading all the blame onto one constraint.

Thus, in the presence of both structure-blind constraints and a smoothing term in the learner, generalizations that are true of one type of word (here, monomorphemes) will "leak" onto other types of word (here, compounds).

9. Bias for leakage

Martin shows that given enough *p+p*, the learner can learn that geminate compounds are good (even, better than non-geminate compounds), but if monomorphemes lack geminates, there is always a bias against geminates in compounds.
(E.g., to learn that *p+p* and *t+p* are equally good, the learner must see nearly 60% *p+p*/40% *t+p*.)

Martin considers other ways that this "leakage" could occur, but what's interesting about this model for us is that it explicitly models how mislearning can result from a tradeoff between accuracy and goodness.

4.2 Alternation-learner, with substantive bias: Wilson 2006

10. Velar palatalization

Cross-linguistically, it's common for /k/ and /g/ to become [t͡ʃ] and [d͡ʒ] before [i], and to a lesser extent before [e] and other "front" vowels (these examples, from Guion 1996, are of diachronic sound change):

(1) k > tʃ / __ {j, ĭ, i, e, ɛ, ē}

Pre-Proto-Slavic	OCS	Gloss
*wilk-e	vl̥itʃe	'wolf' (voc.)
*pla:k-j-o:-m	platʃõ	'I cry'

(Guion ch. 2 p. 4)

(9) k, k', x > tʃ, tʃ', ʃ / __ i [5]

Proto-Salish	Cowlitz Salish	Gloss
*k'ilk	tʃ'ilk	'window'
*kitaq-	tʃæq-	'argue'
*túlxils-	túlʃils-	'hint'

(Guion ch. 2 p. 12)

11. Confusability

This is presumably because the "fronter" articulation of /k/ and /g/ before [i,e] creates a sound that is hard to distinguish from [t͡ʃ]/[d͡ʒ], as can be seen for [i] in this confusability table from Guion 1998:

Confusions of velars and palatoalveolars

Stimulus	Response [ki]	[t͡ʃi]	[gi]	[d͡ʒi]	[kɑ]	[t͡ʃɑ]	[gɑ]	[d͡ʒɑ]
[ki]	43	35	10	12				
[t͡ʃi]	10	85	0	5				
[gi]	4	4	71	21				
[d͡ʒi]	9	28	12	51				
[kɑ]					84	13	3	0
[t͡ʃɑ]					10	87	0	3
[gɑ]					4	0	87	9
[d͡ʒɑ]					2	23	10	65

Note. From "The Role of Perception in the Sound Change of Velar Palatalization," by S. G. Guion, 1998, Phonetica, 55, pp. 18–52. Copyright 1998 by S. Karger AG, Basel. Adapted with permission.

Wilson p. 949

(English-speaking subjects, stimuli masked by noise.)

12. The bias: [k,g] should be more confusable before [i] than [e], and more before [e] than [a]

Wilson devises a measure of similarity based mainly on "peak spectral frequency", fitted to Guion's confusion data that would predict intermediate status for [e]:

Maximum likelihood estimates of perceptual similarities in three vowel contexts

[ki]/[t͡ʃi]	[ke]/[t͡ʃe]	[kɑ]/[t͡ʃɑ]	[gi]/[d͡ʒi]	[ge]/[d͡ʒe]	[gɑ]/[d͡ʒɑ]
9.23^{-1}	*12 .68*$^{-1}$	88.72^{-1}	21.13^{-1}	*40.60*$^{-1}$	126.93^{-1}

Note. *i/j* denotes $b_j\eta_{ij}$. Values in italics are interpolated.

(Wilson p. 954)

13. Wilson's artificial-language-learning experiment

Subjects in the "High" group were taught palatalization only before [i]—Wilson predicts that they won't generalize to [e], and they didn't.

Subjects in the "Mid" group were taught palatalization only before [e]—as predicted, they generalize that the rule applies everywhere equally.

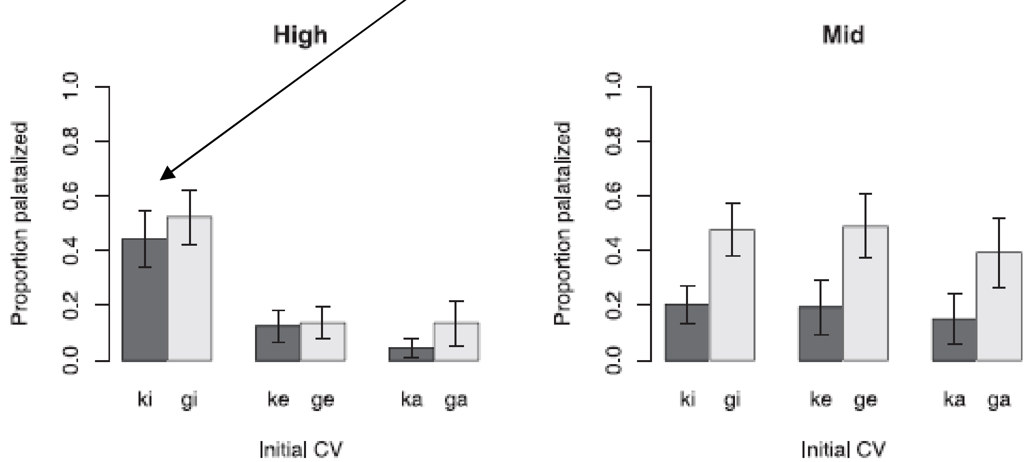


Fig. 2. Results of Experiment 1 by condition. *Note.* Error bars represent standard error of the mean.

(Wilson p. 966)

14. The learner

Wilson uses the similarity values derived above to assign to each of the markedness constraints below its own **prior** σ^2 (see equation from Martin above),[12] which determines how reluctant that constraint is to move from its default weight μ (0 in all cases here).

Markedness constraints on palatalization

	Prior Values			
	Biased		Unbiased	
Constraint	μ	σ^2	μ	σ^2
*ki	0.0	9.23^{-2}	0.0	10^{-2}
*ke	0.0	12.68^{-2}	0.0	10^{-2}
*kɑ	0.0	88.72^{-2}	0.0	10^{-2}
*k$V_{[-low]}$	0.0	12.68^{-2}	0.0	10^{-2}
*k$V_{[-high]}$	0.0	88.72^{-2}	0.0	10^{-2}
*kV	0.0	88.72^{-2}	0.0	10^{-2}
*gi	0.0	21.13^{-2}	0.0	10^{-2}
*ge	0.0	40.60^{-2}	0.0	10^{-2}
*gɑ	0.0	126.93^{-2}	0.0	10^{-2}
*g$V_{[-low]}$	0.0	40.60^{-2}	0.0	10^{-2}
*g$V_{[-high]}$	0.0	126.93^{-2}	0.0	10^{-2}
*gV	0.0	126.93^{-2}	0.0	10^{-2}

For example, it requires little data to increase the weight of *ki (big σ^2), but much more data to increase the weight of *kɑ.

There are also faithfulness constraints, one against changing /k/ and one against changing /g/.

The learner is similar to the one used in Martin—goodness of fit to the data is traded off against a smoothing term, which would prefer all weights to remain 0.

15. Results

For the High condition, where the subjects are essentially repeating back what they were taught, the learner does fine at matching the experimental results with or without biased priors .

But for the Mid condition, the learner matches the experimental results much more closely with biased priors:

Correlations (r) between observed and predicted rates of palatalization in Experiment 1

Condition	Model	All Items	Critical Items
High	Substantively biased	.910 (.83)	.870 (.76)
	Unbiased	.913 (.83)	.871 (.76)
Mid	Substantively biased	.859 (.74)	.758 (.58)
	Unbiased	.550 (.30)	.396 (.16)

Note. Values in parentheses are percentage variance explained (r^2).

Wilson p. 968

[12] "the prior σ of a Markedness constraint is equal to the perceptual similarity of the sounds in the greatest change that is motivated by the constraint" (p. 959)

4.3 Course summary

- We saw some examples of what a language's sound pattern can include—but this only scratches the surface! There's a lot of phonology out there...
- We saw the two most widespread means of describing these patterns: rules and OT constraints, though of course we couldn't cover all the notation in these approaches or all the differences between the two approaches.
- We saw some techniques for probing real speakers' grammars—what do we know (implicitly) about the sound pattern of our language?
- We also saw some techniques for probing (i) what speakers know independent of the specific sound pattern they've been exposed to and (ii) what kinds of grammars speakers prefer.
- Finally, we've seen some attempts to model how learning might proceed, balancing fitting a grammar to the data against the learner's biases and preferences.
- It's unclear what the relationship between the theory and the typology is! We're still a long way from having a complete, integrated theory of learning and language change—see Niyogi (2006) for some (not-phonology-specific) schematic models of how this might work.

4.4 References

Goldwater, Sharon and Mark Johnson 2003. Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Pp. 111-120. (www.cog.brown.edu/~mj/papers/GoldwaterJohnson03.pdf)

Guion, Susan (1996). Velar palatalization: coarticulation, perception, and sound change. University of Texas at Austin dissertation. (available at darkwing.uoregon.edu/~guion/diss.htm)

Hayes, Bruce & Colin Wilson (accepted). A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*. (www.linguistics.ucla.edu/people/hayes/#Research)

Martin, Andy (2007a). *The evolving lexicon*. UCLA dissertation. (www.linguistics.ucla.edu/faciliti/diss.htm)

Martin, Andy (2007b). Grammars leak: how categorical phonotactics can cause gradient phonotactics. Poster at the Workshop on Variation, Gradience and Frequency in Phonology, Stanford University. (www.linguistics.ucla.edu/people/grads/amartin/)

Niyogi, Partha (2004). *The computational nature of language learning and evolution*. MIT Press.

Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30: 945-982. (www.linguistics.ucla.edu/people/wilson/papers.html)