Class 1 (Week 0): Introduction, overview, SPE review

To do for next time
- Read K&K ch. 2 and turn in study questions in class on Tuesday—but note there are more coming on Thursday
- Check out course web page, especially feature links
- Make sure you get a PTE number from me—I just need your student ID number

Overview: Big picture: what are we trying to do? Little picture: review of SPE rule mechanics.

0. Items of business
- Introduce ourselves; student info sheets; ungraded warmup problem
- Syllabus and administrative questions

Part A: Intro & overview

1. What is our job as phonologists? There are various answers...
- To describe phonologies (bullets from Goldsmith 1995)
  - What are the legal/possible words of the language?
    - phone inventory (set of legal sounds)
    - phonotactics (set of legal sound sequences)
  - What alternations occur (changes that sounds undergo when placed in different contexts)?
  - Which phonetic differences are contrastive?
- To explain why phonologies are the way they are by constructing…
  - a theory of what people's knowledge of linguistic sound patterns is and how they learn, store, and use that knowledge
  - plus a theory of how linguistic sound patterns change over time, which ought to follow from the above

This will be our focus

2. Chomskyan basics[1]
- Let a **grammar** consist of (at least)[2]
  - a function that labels any utterance as **grammatical** or **ungrammatical**.
  - a function that assigns truth conditions to any utterance

  might be implemented as alexicon and 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.[3]
- These functions should be accompanied by algorithms for calculating them.

---

[1] Chomsky 1965 pp. 25-27 plus other Chomsky works, simplified and filtered through my own views.
[2] We probably want the grammar to do much more. E.g., given an utterance, return a gradient "goodness score" rather than a binary judgment.And of course 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 do we know *a priori* that it's necessary?)
[3] Chomsky actually breaks this into a **linguistic theory**, which defines the set of possible grammars, and a **strategy** for selecting a grammar given the learning data.

So...

- an <u>observationally adequate grammar</u> labels the utterances that a typical learner would encounter as grammatical (perhaps trivially, e.g. by listing them), and assigns the right truth conditions to them.[4]
- a <u>descriptively adequate grammar</u> captures the “significant” (psychologically real?) generalizations
- the real prize, an <u>explanatorily adequate theory</u>, will, given typical learning data, return an descriptively adequate grammar

But how do we figure out what the significant/psychologically real generalizations are?????

3. Example: English noun plurals

cat	k^{h}æt	k^{h}æts	*pea*	p^{h}i	p^{h}**iz**
sack	sæk	sæks	*cow*	k^{h}aʊ	k^{h}aʊ**z**
dog	dɑg	dɑgz	*man*	mæn	m**ɛ**n
grub	gɹʌb	gɹʌb**z**	*foot*	fʊt	f**i**t
dish	dɪʃ	dɪʃ**ɨz**	*wife*	waɪf	waɪ**vz**
fudge	fʌd͡ʒ	fʌd͡ʒ**ɨz**	*whiff*	wɪf	wɪf**s**
			...		

Examples of observationally adequate grammars for English noun plurals

I. (just list every word you know)

k^{h}æt	k^{h}æts	p^{h}i	p^{h}**iz**
sæk	sæks	k^{h}aʊ	k^{h}aʊ**z**
dɑg	dɑgz	mæn	m**ɛ**n
gɹʌb	gɹʌb**z**	fʊt	f**i**t
dɪʃ	dɪʃ**ɨz**	waɪf	waɪ**vz**
fʌd͡ʒ	fʌd͡ʒ**ɨz**	wɪf	wɪf**s**

I.e., the grammar’s judgment function accepts utterances containing these items in positions where a plural is required (*I like cats*) and assigns appropriate truth-conditions (*I like cats* is true iff I like members of the <u>cat</u> set—it has nothing to do with whether I like members of the <u>dog</u> set).

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

dɑg	dɑgz	k^{h}aʊ	k^{h}aʊ**z**
gɹʌb	gɹʌb**z**	mæn	m**ɛ**n
dɪʃ	dɪʃ**ɨz**	fʊt	f**i**t
fʌd͡ʒ	fʌd͡ʒ**ɨz**	waɪf	waɪ**vz**
p^{h}i	p^{h}**iz**	...	...

[4] That’s the best I can do for a definition—observational adequacy isn’t discussed that much

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

k^{h}æt	k^{h}æts	mæn	mɛn
sæk	sæks	fʊt	fit
dɪʃ	dɪʃ**ɨz**	waɪf	waɪ**vz**
fʌd͡ʒ	fʌd͡ʒ**ɨz**	wɪf	wɪf**s**
		...	...

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

mæn	mɛn
fʊt	fit
waɪf	waɪ**vz**
...	...

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

mæn	mɛn
fʊt	fit
wɪf	wɪf**s**
...	...

Which generalizations are real? How about a wug test.

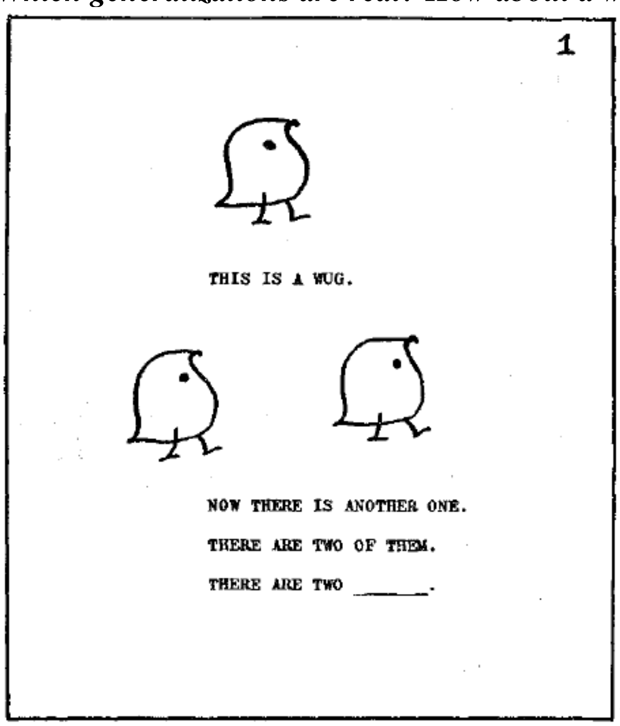


Figure 1. The plural allomorph in /-z/.

(Berko 1958, p. 154)

Berko 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**ɨz**	nɪz	nɪz**ɨz**
kæʒ	kæʒ**ɨz**	kɹa	kɹa**z**
toɹ	toɹ**z**	tæs	tæs**ɨz**

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

hif hif**s**, hiv**z**

4. Why is it hard to develop a descriptively adequate grammar 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 a challenge.
  - Contrast this with syntax, where it's easier to construct sentences that—presumably—the speaker has not encountered before.
- We often can't be sure that these novel situations really test what we want them to test.

- In 200A, we'll mostly ignore this problem and proceed as though generalizations that we notice in the data are real to speakers.
- In 201A there will probably be a bit more emphasis on methods for determining which generalizations are real.

5. Why is it hard to develop an explanatorily adequate theory?

Suppose we could magically achieve description adequacy for all real languages.

- That only tells us which generalizations people have extracted for existing sets of data
- We don't know what people would do if faced with a language with different generalizations
- To build our linguistic theory, we need to know which generalizations people can extract or tend to extract from all kinds of learning data, not just attested learning data.
  - ▪ Are some preferred to others?
  - ▪ Are there hard limits on learnability?

For example

- Suppose we're convinced by the wug test that English speakers' grammar includes "use the [ɨz] form of the plural after sibilants".

→ Exposed to the English data, learners prefer grammar with that generalization to one without

But we still know nothing about the learnability of "use the [ɨz] form of the plural after **non-**sibilants". (Becker, Ketrez, & Nevins 2011 and Becker, Nevins, & Levine 2012 tackle this problem in a very interesting way, by comparing potential generalizations that exist within a language.)

Part B: SPE (Chomsky & Halle 1968) rule notation review

6. An example: SPE's main stress rule (p. 240)

$$V \rightarrow [1\ \text{stress}] \ \big/ \ \Big[\ X\_\_C_0 \left(\begin{bmatrix} -\text{tense} \\ \gamma\text{stress} \\ V \end{bmatrix} C_0^1 \left(\begin{bmatrix} \alpha\text{voc} \\ \alpha\text{cons} \\ -\text{ant} \end{bmatrix} \right) \right)$$

$$\big/ \ \_\_ \ \Big\langle \left(\begin{Bmatrix} (fik)At \\ [+D]C_0 \end{Bmatrix} \right) \begin{Bmatrix} <_1 +C_0 >_1 \begin{bmatrix} -\text{stress} \\ -\text{tense} \\ -\text{cons} \end{bmatrix} [+\text{cons}]_0 \\ <_1 \begin{bmatrix} -\text{seg} \\ <_2 -\text{FB} >_2 \end{bmatrix} >_1 C_0\ [\beta\text{stress}]\ C_0 <_2 V_0 C_0 >_2 \end{Bmatrix} \Big\rangle \ \Big]_{<NSP <_1 VA >_1>}$$

<u>Conditions</u>: $\beta = \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}$

$\gamma \leq 2$ [in another version, says γ is 2 or weaker]

X contains no internal #

- Not much is said in SPE about these "conditions", except that they are truth-functional. It makes a big difference to the theory's computational power what restrictions we place on them.
- Don't worry—you'll almost never encounter a rule this complicated.

7. A → B / X __ Y

Example: $\begin{bmatrix} +\text{syll} \\ -\text{low} \end{bmatrix} \rightarrow [+\text{high}]$ / __ CC#

- ➢ means "*XAY* is rewritten as *XBY*", or, to put it another way, "*A* is rewritten as *B* when preceded by *X* and followed by *Y*".

- A is the affected segment, focus, or target of the rule.
- B is the structural change that the rule requires
- X__Y is the context for the rule
- XAY is the structural description

We'll use *A*, *B*, *X*, and *Y* to stand for these positions throughout this handout.

8. *Something we'll skip*: A → B / X __ Y / P__Q

- ➢ Means "*PXAYQ* is rewritten as *PXBYQ*".
- I.e., A → B / PX __ YQ.
  - Except that ordering for "expansion conventions" (which we haven't discussed yet) is affected—see SPE pp. 72-77.

9. Left side of the arrow

A can be a feature matrix or Ø.

➢ If *A* is a feature matrix, like $\begin{bmatrix}+\text{syll}\\-\text{low}\end{bmatrix}$, then the rule looks for any segment that is **nondistinct** from that matrix.

- Two feature matrices are **distinct** iff there is some feature F whose value is different in the two matrices.
  - o Which of the following are distinct from $\begin{bmatrix}+\text{syll}\\-\text{low}\end{bmatrix}$? $\begin{bmatrix}+\text{syll}\\-\text{low}\\+\text{round}\\+\text{back}\end{bmatrix}$, $\begin{bmatrix}-\text{low}\\-\text{round}\end{bmatrix}$, $\begin{bmatrix}-\text{syll}\\-\text{low}\\+\text{high}\end{bmatrix}$
  - This means that if *A* doesn't mention some feature F, it "doesn't care" about it—that part of the rule matches segments that are +F, or –F, or even fail to have a value for F.

- Sometimes, if *A* is meant to pick out a single sound, we use an IPA (or other transcription system) symbol instead:

  u → [–high] / __ (C)#

  - This is a good idea for readability, but in order to determine how long the rule is (for purposes of applying a length-based evaluation metric), you'd have to expand the IPA symbol into a feature matrix.

  - o What's the smallest feature matrix that "u" could abbreviate if the language's vowel inventory is *i*, *a*, *u*? If it's *i*, *a*, *u*, *o*? If it's *i*, *y*, *a*, *u*, *o*?

- Sometimes we also use C to abbreviate [–syll] or V to abbreviate [+syll].
  - Again, this is good for readability.
  - Be careful when *you* read, though, because some authors, following SPE, use C and V to abbreviate {[–voc], [+cons]} and [+voc, –cons].

➢ If *A* is Ø, you've got an insertion rule (the idea is that insertion changes "nothing" into something):

  Ø → i / C __ C#

o Why don't we use the empty matrix [] instead of Ø?

10. An unsolved issue: underspecified targets

- Imagine a rule like $\begin{bmatrix}+\text{coronal}\\-\text{voice}\end{bmatrix}$ → Ø / __ #
  - And imagine we've decided that sonorants in the language in question are underlyingly underspecified for [voice] (some later rule will fill in their voicing values).
  - E.g., feature matrix for /n/ doesn't contain any kind of [voice], either [+voice] or [–voice].

o How should the rule apply to /bil/ according to our definitions? Does this seem right?

- There's an inconclusive discussion on pp. 382-389 of SPE about whether we should...
  - change the definition of when a rule is applicable so that nondistinctness isn't enough
  - or impose a condition that segments always have to be specified for all the features that a rule's structural description mentions, by the time the rule applies
  - or impose conditions on lexical entries that will rule out some of these cases

In practice, this won't come up much. If it does, you'll need to decide how the rule should apply and be explicit about your decision.

11. Right side of the arrow

B also can be a feature matrix or Ø.

➢ <u>If *B* is a feature matrix</u>, then any of the affected segment's features that are mentioned in *B* are changed to the value given in *B*. *All other features are left unchanged.*

- o What does $\begin{bmatrix}\text{+syll}\\ \text{–low}\end{bmatrix} \rightarrow$ [+high] do to [o]? To [u]?

➢ <u>If *B* is Ø</u>, then the segment that *A* matched is deleted.

C → **Ø** / C__# (why not []?)

- Again, we sometimes use an IPA symbol as an abbreviation for all the feature changes necessary to change anything that could match *A* into the desired *B*:

  $\begin{bmatrix}\text{+syll}\\ \text{–low}\end{bmatrix} \rightarrow$ i / __ #

  - o What does the "i" above abbreviate if the language's vowel inventory is *i*, *a*, *u*? If it's *i*, *a*, *u*, *o*? If it's *i*, *y*, *a*, *u*, *o*?

  - If *A* is Ø, then the IPA symbol for B abbreviates the features needed to pick it out of the language's phoneme inventory: Ø → i / C __ C#

12. Redundancy

➢ The claimed principle that shorter rules are preferred by learners over longer rules (see below) means that unnecessary features should be eliminated from *A* and *B*.

o What is suboptimal about each of the following rules?

$\begin{bmatrix}\text{+syll}\\ \text{–round}\end{bmatrix} \rightarrow$ [+round]

$\begin{bmatrix}\text{+nas}\\ \text{+voice}\end{bmatrix} \rightarrow$ [+anterior] (assume the phoneme inventory of English)

13. Right side of the slash (context)

X and *Y* are strings made up of

- feature matrices
- IPA symbols, which abbreviate feature matrices
- the boundary types # and +, which in SPE also abbreviate feature matrices
- at their outside edges, category boundaries

➢ Feature matrices in *X* and *Y* match segments in the same way that *A* does (i.e., they match a segment if not distinct from it). IPA symbols also work the same way

➢ Boundaries, # (word boundary) and + (morpheme boundary), are treated in SPE as feature matrices that happen to be [–segmental]:

$$\# \text{ is } \begin{bmatrix} -\text{seg} \\ -\text{FB} \\ +\text{WB} \end{bmatrix} \qquad + \text{ is } \begin{bmatrix} -\text{seg} \\ +\text{FB} \\ -\text{WB} \end{bmatrix}$$

([FB] is "formative (roughly, morpheme) boundary" and [WB] is "word boundary").

- There are some complications about #: in SPE, it's not exactly equivalent to the place where you'd write a space in ordinary writing.
- SPE also proposes a third boundary type, =, which has the features $\begin{bmatrix} -\text{seg} \\ -\text{FB} \\ -\text{WB} \end{bmatrix}$ and is more or less the boundary between nonproductive or nontransparent affixes and stems (e.g., English *per=mit*). We won't use this one much.

- The term 'unit' is used in SPE to refer to all feature matrices, including true segments and boundaries.

➢ Category boundaries (labeled brackets) like $]_{\text{Noun}}$ and $_{\text{Verb}}[$ can also be used, but only at the edges of *X*__*Y* (and if both edges have labeled brackets, the labels have to match):

/ __ VC#$]_{\text{N}}$

- By convention, this can be abbreviated as / __ VC$]_{\text{N}}$.

Here's how we extend the definition of nondistinctness from pairs of units to pairs of strings:

- *X* (or *Y*) matches (is nondistinct from) some substring *M* of a form iff *X* and *M* have the same number of units *n*, and the i^{th} unit of *X* matches (is not distinct from) the i^{th} unit of *M* for all $1 \leq i \leq n$.

14. + is special

➢ If + is included in *X* and *Y*, then it is required

- V → Ø / __+VC does not apply to *ibauk*, because +V does not match any substring of *ibau*.

- But—this is the special part—extra +s in the form are always OK: V → Ø / __VC does apply to *iba+uns*,
  - because "__VC" matches any of { __VC, __+VC, __V+C, __+V+C}.

o Which version of the rule is matching here?

doesn't work this way; it works like any other feature matrix.

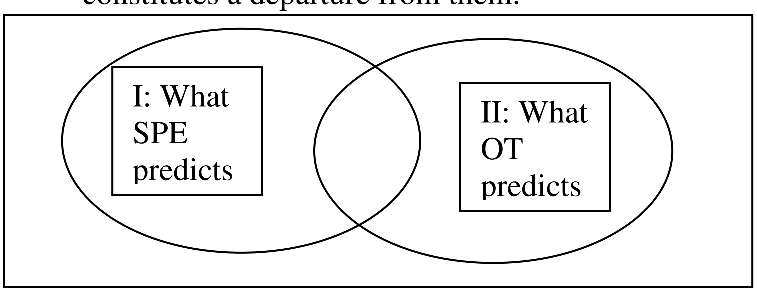
15. Basic rule application

➢ A rule applies to a form if the form contains a string that is nondistinct from *XAY*.

o What if *X* or *Y* doesn't appear in the rule (A → B / __ Y or A → B / X __)?

16. Wrap-up of today

- We've gone into excruciating detail about how a seemingly simple theory works—why?
  - In the past, you've probably been taught a theory of convenience that worked well for the course material.
    - It may have cobbled together elements of various proposals and left various questions open.
  - Here we're going to try to be very explicit about what are our 2 base theories and what constitutes a departure from them.



Are there real cases that are in I but not II?
In II but not I?

- **Next time**: You may recall seeing symbols like ( ) { } < > * ^{0} and others in rules, and treating them as convenient abbreviatory conventions. We'll review these symbols and see how SPE takes them seriously as theoretical claims.

References

Becker, Michael, Nihan Ketrez & Andrew Nevins. 2011. The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1). 84–125. (15 March, 2012).

Becker, Michael, Andrew Nevins & Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88(2). 231–268. doi:10.1353/lan.2012.0049 (1 June, 2013).

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*.. Cambridge, Mass.: MIT Press.

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

Goldsmith, John A. 1995. Phonological Theory.. In John A Goldsmith (ed.), *The Handbook of Phonological Theory*, 1–23. Cambridge, Mass., and Oxford, UK: Blackwell.