Class 2: Expansion conventions

To do for next time
- Nothing to do for Thursday, but if you want to get ahead you can start on the next reading.

Overview: We've seen how the basic rule formalism works. Today we'll consider the mechanics and implications of notation like αvoice, ( ), { }, < >, *, C_0.

Expansion conventions

- Devices like parentheses, curly brackets ("braces"), and angle brackets are used to collapse related rules into a single *rule schema* (whose length is shorter = cost is lower).

- Rather than adjusting the definition of nondistinctness, SPE (Chomsky & Halle, 1968) gives *expansion conventions* to turn those schemata into lists of rules that can then be applied using the simple definition of nondistinctness.

1. Lowercase Greek letters

➢ Variables that stand for +, –, or whatever values the theory says some feature can take (could be 1,2,3 for some features—can you think of any good candidates?).

C → [αvoice] / [αvoice] __ [αvoice] expands into

C → [+voice] / [+voice] __ [+voice]
C → [–voice] / [–voice] __ [–voice]

2. Parentheses

➢ Used to indicate optionality.

- For example, the rule schema V → Ø / __(V)C# is expanded into these two rules (in that order—but we'll come back to that another day):

V → Ø / __VC#
V → Ø / __C#

o Would you ever want to use parentheses in a <u>feature matrix</u>? *plickers: A—yes, B—no*

3. Disjunctive ordering

- The rules that a schema expands into are *disjunctively ordered*.
- Informally:
  - First you try the first sub-rule
  - If its structural description is met, you apply that first rule and don't try any of the other rules from the same schema
  - If not, move on to the next rule and proceed in the same fashion.
- In other words, you never apply two rules of the same schema to a single word.

o How does the rule above, V → Ø / __(V)C#, apply to /bauk/?

(This is a bit too crude, because it doesn't give the right result for cases where different rules of a schema apply to different parts of a word—in those cases, we want multiple rules of the schema to apply to the same word, just in different places. We'll come back to that another day too.)

4. Braces, a.k.a. curly brackets

➢ Used to indicate multiple possibilities

For example, the rule schema $\left\{\begin{matrix} i \\ o \end{matrix}\right\}$ → Ø / __V is expanded into these two rules (in this order):

i → Ø / __V
o → Ø / __V

o Can you imagine a way to translate parentheses into braces? Try it with V → Ø / __V(C)#

- Rules from the same curly-bracket schema apply ***conjunctively*** (apply the first one, then the second, etc.)
  - Thanks to Patrick Jones for de-confusing me on this!
  - o SPE gives an example where you do actually need to apply multiple sub-rules (p. 341)—can you devise an input for the rule above where conjunctive and disjunctive order would produce different results?

- Some phonologists think that curly brackets are so powerful that the theory shouldn't allow them—that resorting to them is an admission of failure (either of the analyst or of the theory).

5. Super- and subscripts

- X_n^m means from *n* to *m* Xs
  - C_n : "*n* or more Cs" (most common is C_0)
  - V^m : "up to *m* Vs"
  - C_n^m : "anywhere from *n* to *m* Cs"

$C \rightarrow \emptyset / \_\_C_0\#$ = …
$C \rightarrow \emptyset / \_\_CCCC\#$
$C \rightarrow \emptyset / \_\_CCC\#$
$C \rightarrow \emptyset / \_\_CC\#$
$C \rightarrow \emptyset / \_\_C\#$
$C \rightarrow \emptyset / \_\_\#$

- The tricky thing is that the "..." is at the top of the list.
  - That is, we apply the *longest* rule whose structural description matches.

o How would the schema above apply to /tabskt/?

6. **Parentheses with star** (But see discussion in Week 4 Anderson reading—he argues for changing this a bit)

- (…)* means that the material in parentheses can occur zero or more times.

V → [+stress] / #C(VCVC)*__ expands to

V → [+stress] / #C__
V → [+stress] / #CVCVC__
V → [+stress] / # CVCVCVCVC __
...

- **With ()*, disjunctive ordering does *not* apply.**
  - Every version of the rule that can apply does apply—simultaneously.

o How would the stress rule above apply to /badupidome/?

o How would $C \rightarrow \emptyset / \_\_C^*\#$ apply to /tabskt/?

7. Angled brackets

➢ Like parentheses, but when the optional information is in more than one place.

- A schema with angle brackets expands into two rules: the rule with the information in the angle brackets and the rule without that information.

C → Ø / V<C>__<C>V (silly example) expands to

C → Ø / VC__CV
C → Ø / V__V

o Expand the following schema and apply it to *putod*, *luged*, and *fesil*.

$$\begin{bmatrix}\text{+syll}\\ \text{<+back>}\end{bmatrix} \rightarrow \text{[–hi] / \_\_ C} < \begin{bmatrix}\text{+syll}\\ \text{+back}\\ \text{–hi}\end{bmatrix} \text{ C >\#}$$

- You can also subscript angle brackets to show which ones go together:

C → Ø / V<$_1$C>$_1$__<$_2$s>$_2$<$_1$C>$_1$V<$_2$h>$_2$# (even sillier rule) expands to

C → Ø / VC__sCVh#
C → Ø / V__sVh#
C → Ø / VC__CV#
C → Ø / V__V#

8. Transformational rules

➢ Useful for metathesis, coalescence…anything where more than one segment is affected at once.

- In SPE, these were given in two parts:

Structural description: $\underset{1}{\begin{bmatrix}\text{+syll}\\ \text{+low}\end{bmatrix}}, \underset{2}{\begin{bmatrix}\text{+syll}\\ \text{+hi}\\ \text{αround}\end{bmatrix}}$

Structural change: 1 2 → $\begin{bmatrix}1\\ \text{–lo}\\ \text{+long}\\ \text{αround}\\ \text{αback}\end{bmatrix}, \begin{bmatrix}2\\ \text{Ø}\end{bmatrix}$

o What does this rule do?
o It may seem arbitrary to say that 1 changes and 2 deletes rather than the reverse. Try writing the rule the other way too.

- It's common to use a simplified notation instead that collapses the structural description and structural change:

$$\begin{bmatrix}+\text{syll}\\+\text{low}\end{bmatrix}_1 \quad \begin{bmatrix}+\text{syll}\\+\text{hi}\\ \alpha\text{round}\end{bmatrix}_2 \rightarrow \begin{bmatrix}1\\-\text{lo}\\+\text{long}\\ \alpha\text{round}\\ \alpha\text{back}\end{bmatrix}$$

  - What's wrong with just saying this:

$$\begin{bmatrix}+\text{syll}\\+\text{low}\end{bmatrix} \quad \begin{bmatrix}+\text{syll}\\+\text{hi}\\ \alpha\text{round}\end{bmatrix} \rightarrow \begin{bmatrix}-\text{lo}\\+\text{long}\\ \alpha\text{round}\\ \alpha\text{back}\end{bmatrix}$$

  - Formulate a transformational rule for the metathesis seen in the warm-up problem last time.

  - In the directions for the warm-up problem, I said that you could use a rule of the form AB → BA, but that this was not quite adequate. Rewrite your transformational rule from above as a AB → BA rule. What's wrong with it?

9. How does the learner choose a grammar?

- SPE proposed that if more than one grammar can generate the observed linguistic data, the learner must have some ***evaluation metric*** for choosing one.
- The evaluation metric tentatively proposed in SPE is brevity: learner chooses the grammar with the fewest symbols. (What about ties?)
- If that's right, and if we've got the notation right too, then you can tell which grammar, out of some set of candidate grammars, the learner would choose.
- More plausibly, we want to find independent evidence as to which grammar is right, and then make sure our theory explains how/why the learner chose that one—this is a lot harder!

10. ***Example:*** **French elision/liaison** (SPE p. 353 ff.)

- By the logic above, a theoretical innovation is held, in SPE, to be a good one if it allows more concise descriptions of attested/common phenomena than of unattested/uncommon phenomena.

		obstruent- or nasal-initial	liquid-initial	vowel-initial	glide-initial
		/**g**arson/ 'boy'	/**l**ivr/ 'book'	/**ɛ**nfant/ 'child'	/**w**azo/ 'bird'
obstruent- or nasal-final	/pətit/ 'small'	pəti_ garsõ	pəti_ livr	pətit ãfã	pətit wazo
liquid-final	/ʃɛr/ 'dear'	ʃɛr garsõ	ʃɛr livr	ʃɛr ãfã	ʃɛr wazo
vowel-final	/lə/ 'the'	lə garsõ	lə livr	l_ ãfã	l_ wazo
glide-final	/parej/ 'similar'	parej garsõ	parej livr	parej ãfã	parej wazo

For the sake of reconstructing the argument, use the archaic feature [vocalic] and the still-current feature [consonantal]:

	vocalic	*consonantal*
obstruents	–	+
nasals	–	+
liquids	+	+
glides	–	–
vowels	+	–

- Propose rules to account for the C- and V- deletions, without using Greek-letter variables.

- Revise the rules, using Greek-letter variables

- Do Greek-letter variables allow us to compress these two rules:

$\begin{bmatrix} +\text{voc} \\ -\text{back} \end{bmatrix} \rightarrow \emptyset$ / __ # [–cons] "nonback vowels and liquids delete before vowels and glides"

$\begin{bmatrix} -\text{high} \\ +\text{cons} \end{bmatrix} \rightarrow \emptyset$ / __ # [+nasal] "nonhigh consonants and glides delete before nasals"

- According to SPE's logic, how should the typology guide us in deciding whether to allow the same Greek-letter variable to apply to different features within a rule?

11. This is very different from the reasoning you read in Kenstowicz & Kisseberth 1979

- SPE: divide up lexicon (idiosyncratic properties) and rules (systematic properties) so as to make the whole thing as short as possible
- K&K: get external evidence to justify the rules; what you're left with and can't justify as systematic is idiosyncratic

What counts as external evidence?

- Productivity—putting the speaker in a new situation
  - forming new words through productive morphology and syntax
  - wug tests (natural and laboratory)
  - L2
  - speech errors
- Plausibility
  - phonetic naturalness
  - typological attestedness

o Discuss: as some of you pointed out, typological attestedness can be problematic/circular, if we are talking about the typology of e.g., phonemic inventories, or anything else that depends on an analysis.

- FYI there's a huge controversy lurking with phonetic naturalness too: almost everyone agrees that a rule is more likely to arise if phonetically natural (e.g., Papago palatalization before high vowels), but does that mean that learner are likely to learn that rule (vs. de-palatalization when not before a high vowel)?

12. *If we have time:* SPE reasoning above in #9 relies on assumptions about linguistic typology

- Assume a rule is cross-linguistically common only if it's favored by learners—i.e., learners tend to mislearn, in the direction of a more-favored grammar.
- Assume that learners favor short/simple/whatever rules.
- Therefore, rules that are cross-linguistically common should tend to be short.
- Therefore, our theory of rules, which determines what type of notation length is calculated on, should make common rules shorter than uncommon ones.
- Therefore, a theoretical innovation is good if it makes common rules shorter than uncommon ones.

=> We're not really using "short" (or "simple") in any fixed sense. Rather, we're tailoring the notation to make the rules that we think learners favor appear short. [And of course, that first assumption is questionable...]

This leads us into slippery territory in deciding whether shortness is the right criterion:

- Are learners innately endowed with a certain notation, which they use to calculate grammar length? (i.e., shortness really is the evaluation criterion)
- Or is it the case that learners employ some other evaluation metric entirely, but we've created a system of notation that makes goodness according to the real evaluation metric translate into shortness in our notation?

Something for you to think about, though no answers will be forthcoming: We've seen how to evaluate a particular description or even a theoretical innovation, given a framework like SPE.

- But how do you evaluate the framework itself—in particular, how can we evaluate a principle such as "if more than one grammar can generate the observed linguistic data, the learner chooses the grammar with the fewest symbols"?

Next time: What if the grammar contains more than rule? We'll see the SPE approach to rule interaction, extrinsic ordering (what until now you've probably known as just "ordering").

References

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. Harper & Row.

Kenstowicz, M., & Kisseberth, C. (1979). *Generative Phonology: Description and Theory*. New York: Academic Press.