Class 2 & 3: SPE in depth

0. AMP 2025 report

> ***Last time we talked about the big picture. Let's wrap that up, then get into the nuts and bolts of how one theory works and what it predicts.***
> Material that's new to you: you can focus on the mechanics
> Material that's familiar to you: you can focus on the theoretical predictions

Continuing from last time…

1. Review: why is it hard to develop a descriptively adequate grammar in phonology?

- If a speaker already knows a word, like *feet* or *glasses*, it's uninformative to us!
  - Known words don't tell us anything about what generalizations the speaker has learned—they may have simply memorized those words

2. New: Why is it hard to develop an explanatorily adequate theory?

- Suppose we could magically achieve descriptive adequacy for all real languages.
  - That is, we know, for every language, what the "significant" (psychologically real) generalizations are
  - But 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

- In the English example…
  - Suppose we're convinced by the wug test that English speakers' grammar includes the rule "use the [ɨ**z**] form of the plural after sibilants" (*glasses*).

  → Exposed to the English data, learners choose a grammar with that rule

  - But we still know nothing about the learnability of hypothetical "use the [ɨ**z**] form of the plural after **non-**sibilants".
    - If the data had somehow reflected this rule instead, would children be able to learn it just as well?

- 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 generalizations preferred to others?
  - Are there hard limits on learnability?

- Again, this won't be our focus this quarter, but some interesting things you could read:

- Becker, Ketrez & Nevins 2011 and Becker, Nevins & Levine 2012 tackle this problem in a very interesting way, by comparing potential generalizations that exist within the same language—Turkish and English, here.



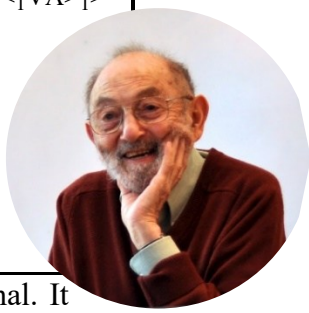

- Bowers 2012 argues that a sudden, one-generation change in Odawa happened because the data changed into something that children couldn't learn.

SPE (Chomsky & Halle 1968) rule notation review
SPE = *The Sound Pattern of English*

3. An example: SPE's main stress rule (p. 240)—let's just admire it for a minute

$$\mathrm{V} \rightarrow [1\ \text{stress}] \,/\, \Big[X \_\_ \mathrm{C}_0 \left(\begin{bmatrix} -\text{tense} \\ \gamma\text{stress} \\ \mathrm{V} \end{bmatrix} \mathrm{C}_0^1 \left(\begin{bmatrix} \alpha\text{voc} \\ \alpha\text{cons} \\ -\text{ant} \end{bmatrix} \right) \right)$$

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



Conditions: $\beta = \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}$
$\gamma \leq 2$ [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 never encounter a rule this complicated!!!
- Let's step through the crucial elements of rule notation.

4. Features

- You can think of a feature as a function that takes a phone and emits a value like + or –
- E.g., *voice* assigns + to [b, d, m, o, a] and – to [p, t, s, h]
- Features usually have some phonetic definition, e.g…
  - *voice* (spoken languages) means "vocal folds are vibrating"
  - *spread* (sign languages) means "fingers are abducted, away from middle finger" (Ormel et al. 2017)
- So what does [+voice] or [–spread] mean? As we'll see, it depends on where it appears in a rule!

5. A → B / X __ Y

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

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

- A is the **affected segment/focus/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.

6. *Something we'll skip, but for your reference*: 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.

7. Left side of the arrow, "A"

A can be a feature matrix or Ø.

➢ <u>If *A* is a feature matrix</u>, like $\begin{bmatrix}+\text{syllabic}\\-\text{low}\end{bmatrix}$, then the rule looks for any segment that is **non-distinct** from that matrix.

- Two feature matrices are **distinct** iff there is some feature F whose value is different in the two matrices.

  ? Which of the following are distinct from $\begin{bmatrix}+\text{syll}\\-\text{low}\end{bmatrix}$?

  A: $\begin{bmatrix}+\text{syll}\\-\text{low}\\+\text{round}\\+\text{back}\end{bmatrix}$ **B**: $\begin{bmatrix}-\text{low}\\-\text{round}\end{bmatrix}$ **C**: $\begin{bmatrix}-\text{syll}\\-\text{low}\\+\text{high}\end{bmatrix}$

  - 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 phone, we use a phonetic symbol instead:

  u → [–high] / __ C#

  - This is a good idea for readability, but in order to determine how long the rule is (e.g., if you think learners prefer short rules), you'd have to expand the IPA symbol into a feature matrix.

- Do you want a quick review of the features [high/low/front/back/round?
  - **A**: yes, I want a quick review **B**: no, I don't need a review

- 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*?
  Take a minute to write down your answer, then I'll ask for **raised hands**

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

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

Ø → i / C __ C#

- Why don't we use the empty matrix [ ] instead of Ø? Take 2 minutes to think about this alone, then I'll give you 1 minute to discuss it with a partner. If you're stuck, try applying the rule [ ] → i / C__C# to the word /potek/ and the word /bamk/.

8. An unsolved issue: underspecified targets (if we have time)

- Imagine a rule like $\begin{bmatrix}\text{+coronal}\\ \text{–voice}\end{bmatrix} \rightarrow$ Ø / __ #
  - And imagine we've decided that sonorants in the language in question are underlyingly underspecified for [voice] (meaning they have no value for this feature—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].

- How should the rule apply to /bil/ according to our definitions?
  A: produce [bi]
  B: produce [bil]

- Does this seem right? I think there's a case to be made either way. Answer with thumbs up or down and I'll call on one or more people to make their case.

- 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 non-distinctness 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.

9. After all that, we're finally ready for… the right side of the arrow, "B"

- *B* also can be a feature matrix or Ø. But, it is totally different from *A*—**it does not pick out a set of units!!** Instead, it prescribes a **set of changes** to apply.

➢ If *B* is a feature matrix, 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*.

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

  - If I were the boss of all phonology, there would be a whole different notation, like

    $\begin{bmatrix}+\text{syll}\\-\text{low}\end{bmatrix} \rightarrow \| +\text{high} \|$ (I use this in my undergraduate courses)

➢ If *B* is Ø, then the segment that *A* matched is deleted.

C → **Ø** / C__# (why not [ ]? Take a minute to think about it)

- 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 / __ #

  - 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 phonetic symbol for B abbreviates the features needed to pick it out of the language's phoneme inventory: Ø → i / C __ C#

10. Redundancy

- The claimed principle that shorter rules are preferred by learners over longer rules (which we'll get to later) means that unnecessary features should be eliminated from *A* and *B*.

  ? What is suboptimal about each of the following rules? Take a minute to jot something down

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

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

  $\begin{bmatrix}\text{+nas}\\ \text{+voice}\end{bmatrix} \rightarrow$ [+anterior]

(assume the phoneme inventory of English for this last rule)

11. Right side of the slash (context), "X__Y"

- *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). Phonetic symbols also work the same way

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

  # is $\begin{bmatrix}\text{–seg}\\ \text{–FB}\\ \text{+WB}\end{bmatrix}$ + 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, i.e. the boundary between syntactic words.

- 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#$]_N$

- By convention, this can be abbreviated as / __ VC$]_N$

? Discuss: What is the theoretical claim that Chomsky & Halle are making by imposing this only-at-the-edges condition?

12. Non-distinctness of strings

Here's how we extend the definition of non-distinctness from pairs of units to pairs of strings:

- *X* (or *Y*) matches (is non-distinct 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 $1 \leq i \leq n$.

13. + is special

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

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

- But—this is the special part—extra +s in the form are always OK: V → Ø / __VC does apply to *iba+uns*,

  - because "VVC" matches any of {VVC, V+VC, VV+C, V+V+C}.
    A ***B*** ***C*** ***D***

  ? Which version of the rule is matching here?

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

14. Basic rule application

- A rule applies to a form if the form contains a string that is non-distinct from *XAY*.

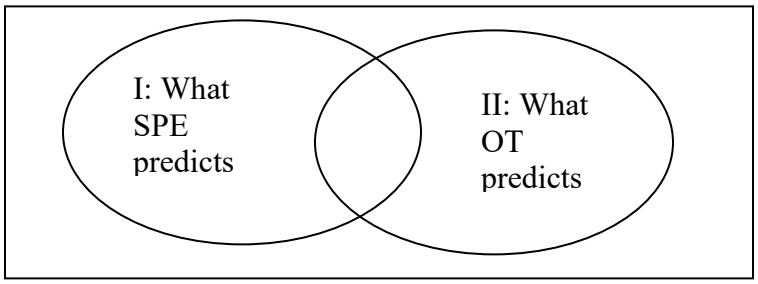
  - What if *X* or *Y* doesn't appear in the rule (A → B / __ Y or A → B / X __)?

15. 5-minute practice break

- 1 min: On an index card, write down a rule—doesn't have to be a real rule from a language
- 2 min: Now exchange cards with a neighbor
  - Paraphrase their rule in prose
  - Devise one imaginary underlying form that *would* undergo the rule and one that wouldn't
- 2 min: Discuss the rule with the person who wrote it

16. Summary of SPE review

- We've started going 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 aspects of its implementation vague.
  - 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?

- Why structure the course this way?
  - Doing theory comparison is important and useful!
    - How else will we find a better theory?
- But it also feels like we're covering a bit of history of the field—is that a goal of this course?
  - Not explicitly, though intellectual history is of course an interesting field…
  - Rather, it's so that you can see what it looks like when a theory is perceived to have problems, and solutions are proposed to those problems[1]
    - This should help you feel more equipped to spot and solve theoretical problems yourself!
    - Otherwise it can be hard to know where to start on one's journey from theory-consumer to theory-contributor

[1] Thanks to Amanda Grein for raising this point (p.c.)—I've incorporated her insights here

Next: Expansion conventions

- 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.

17. "Expansion conventions"

- Devices like parentheses, curly brackets ("braces"), and angle brackets are used to collapse related rules into a single *rule schema*
  - With the idea being that if learners prefer brevity, they will save "space"[2] by collapsing rules when possible

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

18. 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]

19. 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#

? Would you ever want to use parentheses in a feature matrix? Consider both A and B in A → B

[2] But does it really matter whether your phonological grammar takes up, I don't know, 10,000 bytes or 12,000 bytes when your memory has a capacity of 2,500,000,000,000,000 bytes? Maybe processing speed is more important, since we seem to have little trouble storing our language's basic phonology but a lot of trouble avoiding speech errors.

20. 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 sub-rule and don't try any of the other rules from the same schema
  - If not, move on to the next sub-rule and proceed in the same fashion.
- In other words, you never apply two rules of the same schema to a single word.

  ? What's the result of applying V → Ø / __(V)C#, 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! It turns out to be interesting.)

21. Braces, a.k.a. curly brackets

➢ Used to indicate multiple possibilities

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

i → Ø / __V
o → Ø / __V

○ 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!
  - ? 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).

22. 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"
  - Most commonly used as C_0

$C \rightarrow \emptyset / \_\_C_0\#$ = …
C → Ø/ __CCCC#
C → Ø/ __CCC#
C → Ø/ __CC#
C → Ø/ __C#
C → Ø/ __#

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

? What would the schema change /tabskt/ to?

23. Parentheses with star

- (…)* 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 __
...

- Parentheses can be omitted if the scope of * is just one symbol: / __C*#

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

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

? How would C → Ø/ __C*# apply to /tabskt/?

24. Angled brackets (skip if not enough time)

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

- A schema with two pairs of 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

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

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

- You can also have more than one pair of pairs. Then you have to use subscripts to show which ones go together (I have almost never seen this in real life):

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#

25. Transformational rules—these greatly increase the computational power of the theory

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

- In SPE, these were given in two parts:

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

Structural change: 1 2 → $\begin{bmatrix}1\\ -\text{lo}\\ +\text{long}\\ \alpha\text{round}\\ \alpha\text{back}\end{bmatrix}, \begin{bmatrix}2\\ \text{Ø}\end{bmatrix}$

? What does this rule do?
? 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:

$$\begin{matrix}\begin{bmatrix}+\text{syll}\\+\text{low}\end{bmatrix} & \begin{bmatrix}+\text{syll}\\+\text{hi}\\ \alpha\text{round}\end{bmatrix} & \rightarrow & \begin{bmatrix}1\\ -\text{lo}\\ +\text{long}\\ \alpha\text{round}\\ \alpha\text{back}\end{bmatrix}\\ 1 & 2 & & \end{matrix}$$

? What's wrong with just saying this:

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

- It's not too important to be fluent in this notation.

26. In closing

- Write down as many notation devices as you can remember from today:

References

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

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

Ormel, Ellen, Onno Crasborn, Gerrit Kootstra & Anne de Meijer. 2017. Coarticulation of Handshape in Sign Language of the Netherlands: A Corpus Study. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1). 10. https://doi.org/10.5334/labphon.45.