Class 20: Retrospective and prospective course wrap-up

To do
• Fijian due this Friday (Dec.2)
• Work on presentation (Dec. 6-7) and paper (due Dec. 9—if office closes early, you can slide it under my door)

Overview: Some summarizing, some stock-taking, some prospect, a little synthesis. “☼” means you’re likely to learn more about the topic in 201A.

1. Learnability
   • Review of the Chomskyan basics:
     ▪ an observationally adequate grammar labels the utterances that a typical learner would encounter as grammatical (perhaps trivially, e.g. by listing them)
     ▪ a descriptively adequate grammar captures the psychologically real generalizations—this could be operationalized as ‘treats novel utterances the same way real speakers do’
     ▪ the real prize, an explanatorily adequate theory, is a function that, given typical learning data, returns a descriptively adequate grammar
   • Achieving an explanatorily adequate theory is going to have to involve ☼learning algorithms.
     ▪ Interestingly, there was never a good learning algorithm that could induce an ordered list of rules from surface forms, or even from underlying-surface pairs.
     ▪ By contrast, there’s a big literature on learning algorithms in OT.
   • In OT, under the assumption of a finite, universal constraint set...
     ▪ ...and given input-output pairs, it’s easy: see Tesar & Smolensky 2000, Riggle 2004
     ▪ ...and given inputs and just the audible portion of the outputs (e.g., no foot boundaries): it’s harder. See Tesar 2000, Jarosz 2013.
     ▪ ...and given just outputs (with or without their inaudible parts): it’s a lot harder. See Tesar et al. 2003, Jarosz 2006, Jarosz 2015.
       ▪ A fair amount of phonotactic learning can be accomplished, which could later be used to learn alternations, though that second step remains largely unimplemented (see Hayes 2004).
   • There are also learning algorithms for ☼variable/probabilistic constraint rankings:
     ▪ Maximum Entropy OT: Goldwater & Johnson 2003
       You can try out these two (plus a couple of non-variable algorithms) by downloading software from Bruce Hayes’s webpage.
       You can try out by downloading OTSoft from Bruce Hayes’s webpage
What if the constraint set isn’t universal, and constraints have to be constructed by the learner?

- This is still fairly uncharted territory—see Heinz 2007, Hayes & Wilson 2006.

1.1 When multiple grammars are consistent with data, which one does learner select?

- This is the evaluation-metric problem that we’ve seen since the beginning of the course—solving it is part of developing an explanatorily adequate theory.

- The subset problem—say you are exposed to the following language:
  - tagu ‘goat’
  - tagune ‘goats’
  - taguba ‘my goat’
  - ale ‘mango’
  - alene ‘mangos’
  - aleba ‘my mango’
  - siri ‘corkscrew’
  - sirine ‘corkscrews’
  - siriba ‘my corkscrew’

  o In a rule framework, what grammar would you learn?

  o How do you think you would then react to the word sirab? Is this predicted by the grammar?

  o Same question for OT—what ranking would you learn for the constraints NoCODA, MAX-C, and DEP-V? What does this ranking predict for sirab?

- Some learning algorithms have addressed this question of how a learner knows that something they’ve never seen is forbidden, in the absence of helpful alternations (Prince & Tesar 2004, Hayes 2004).
  - The idea is, force markedness constraints to be ranked as high as is consistent with data.

1.2 Ranking bias within markedness or faithfulness constraints?

- Wilson 2006, drawing on Guion 1996: Cross-linguistically, velar palatalization (k→tʃ, ɡ→dʒ) before one front vowel implies palatalization before a higher front vowel—that is, we see languages ki, ke and tʃi, ke and tʃi, tʃe but not ki, tʃe.

  o If we simply have these three constraints, what’s the predicted typology: *ki, *ke, IDENT(place) (I’m leaving out *ka to keep things simple)
- One approach is to build more structure into the constraint inventory: *k[+hi], *k[–lo], IDENT(place).

- What typology do we get now?

- Another approach, for which see Wilson (who has experimental evidence for it):
  - In a ranking system where each constraint is associated with a weight (this is different from Classic OT’s strict ranking), the learning problem involves discovering the weights.
  - We can start with each weight at zero—that is, all constraints are without effect—and promote them in response to the data.
  - Each constraint *i is also associated with a value σ_i that determines how willing the constraint is to change its weight. (Wilson derives these from Guion’s confusion rates.)
  - If we give *ke a smaller σ than *ki, then the algorithm requires more evidence in order to promote *ke than *ki.
  - So it’s possible to learn the typologically anomalous ki, tfe language, but it’s a lot easier (requires less evidence) to learn the other possibilities.
  - See White 2013, Hayes & White 2015 for an approach where constraints have same σ, but different default weights.

1.3 Constraint learning
- What about constraints themselves?
  - If the learner has to construct constraints, are all possibilities equally good?
  - There might be a criterion of formal simplicity, but, as with rules, that’s probably not enough.
    - Compare *\(\alpha_{\text{round}}\) to *\(\alpha_{\text{back}}\)–equally simple, but not equally attested
    - Same issue arises with rules: why \(\alpha_{\text{round}}\)→[\alpha_{\text{back}}] but not [\alpha_{\text{round}}]→[\alpha_{\text{voice}}]?
  - Along with constraint-learning itself, this is an open problem.
1.4 The role of phonetics

- Well-known phonetic explanation for above round/back affinity:
  - lip rounding/protrusion and tongue backing, although articulatorily independent, share an acoustic effect (lower second formant).

![F2 (Bark) diagram]

- Obviously phonetics explains a lot of observed phonology. But...
  - Does the explanatory mechanism lie in learner preferences (Hayes & Steriade 2004, Kawahara 2007) or in pathways of language change (Blevins 2003)?
  - Do grammars make literal reference to phonetic motivation (“don’t have a contour tone if the vowel is shorter than 150 msec”)
  - or do phonetic motivations get phonologized (“don’t have a contour tone except in diphthongs and final syllables”), and if so how?
  - See Hayes 1999 for this question in general; Zhang 2007 for contour tones in particular.

2. Process interaction: extrinsic ordering?

Feeding in Kalinga

<table>
<thead>
<tr>
<th></th>
<th>*o</th>
<th>σ</th>
<th>MAX-V</th>
<th>AGREEPLACE</th>
<th>IDENT(place)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/sin+pajaw/</td>
<td>a</td>
<td></td>
<td></td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>/d-in-opa/</td>
<td>c</td>
<td></td>
<td></td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td></td>
<td>*</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>e</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

- We can’t get both (b) and (d) [counterfeeding] to win, at least not with these constraints

---

1 Thanks to David Deterding’s Excel template (http://videoweb.nie.edu.sg/phonetic/vowels/measurements.html)
Bleeding in English:

<table>
<thead>
<tr>
<th></th>
<th>OBSTRUENTSAGREEVOICE</th>
<th>IDENT(voice)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/kæt+z/</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>a kætz</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>b kæts</td>
<td>!</td>
<td></td>
</tr>
</tbody>
</table>

|            | OBSTRUENTSAGREEVOICE | IDENT(voice) | DEP-V |
|------------|----------------------|--------------|
| /bænt[+z]/| !                    | *            | *     |
| c bænt[+z]| !                    |              |       |
| d bænt[+z]| !                    | *            |       |
| e bænt[+z]| !                    | *            |       |
| f bænt[+z]| !                    | *            |       |

- The counterbleeding candidate (e) can’t win—with these constraints, it’s harmonically bounded.

- Opacity is hard for standard OT to deal with, as we’ve seen! See McCarthy 2007b for a book-length discussion.

- You will probably see some proposals in 201A for how to fix this—most of these proposals were developed for other reasons, but as a side effect predict some opacity:
  - containment (Goldrick & Smolensky 1999)
  - sympathy (McCarthy 2003)
  - candidate chains (McCarthy 2007b, Wolf 2008)
  - output-output correspondence (Crosswhite 1998; Benua 1997; Steriade 2000; Burzio 1998; Kenstowicz 1995 and others)
  - targeted constraints (Wilson 2001)
  - local constraint conjunction (Smolensky 1997, Lubowicz 2005, Kirchner 1996)
  - Stratal OT (Kiparsky 2000)
  - distantial faithfulness (Kirchner 1996)
  - *MAP constraints (Zuraw 2007, Zuraw 2013)
  - comparative markedness (McCarthy 2002)
  - harmonic serialism (McCarthy 2000, McCarthy 2010)

- Most don’t capture all types of opacity, and whether all claimed types of opacity are learnable is debated in, (e.g., Sanders 2002).

3. Process application

3.1 Self-feeding and self-bleeding

- Recall Takelma from Anderson 1974:
  - [a] becomes [i] if followed by [i]: /alxīxamis/ → [alxīxīmis] ‘one who sees us’
  - and any preceding [a]s follow suit: /ikūmananananinkb/ → [ikūminininink] ‘he will fix it for him’ (unless a voiceless C intervenes)

  - This is expected in OT, where self-counterfeeding would be unexpected: (Kaplan 2008).
• Recall French (optional) schwa deletion from Anderson, following Dell 1973:
  \( \text{ə} \rightarrow \emptyset / \text{VC}_C(\text{r})V \)
  
  - \( /\text{ty#dəvəne}/ \rightarrow [\text{ty#dəvəne}] \) or \( [\text{ty#d_vəne}] \) or \( [\text{ty#dəv_ne}] \)
  
  - but not \( *[\text{ty#d_v_ne}] \) ‘you were becoming’
  
  - Again, expected in OT, where self-counterbleeding (Kikuyu??) would be unexpected.

3.2 Directional application

• If there is such a thing as directional rule application...
  - in the sense that the left/rightmost eligible site has priority for undergoing the rule, regardless of whether it’s stressed/unstressed, word-initial/word-final...
  - then standard OT doesn’t have much to say about it (see Hyman & VanBik 2004)

• Hypothetical case (pseudo-French—like real French except rule operates left-to-right):
  
  - only one target: \( /\text{dəvəne}/ \rightarrow [\text{dəv_ne}] \)
  
  - multiple targets: \( /\text{ty#dəvəne}/ \rightarrow [\text{ty#d_vəne}], *[\text{ty#dəv_ne}] \)

• Eisner’s (2002) directional constraint evaluation (proposed for computational reasons, not because of data like this):
  
  - Index a copy of \*SCHWA to each position (counting by segments, though other constraints might count differently) in the output string.

  - Left-to-right version:

<table>
<thead>
<tr>
<th>/\text{ty#dəvəne}/</th>
<th>*CCC</th>
<th>*ə-1</th>
<th>*ə-2</th>
<th>*ə-3</th>
<th>*ə-4</th>
<th>*ə-5</th>
<th>*ə-6</th>
<th>*ə-7</th>
<th>*ə-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a ) ( \text{[ty#d_vəne]} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( b ) ( \text{[ty#dəv_ne]} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>( c ) ( \text{[ty#dəvəne]} )</td>
<td></td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( d ) ( \text{[ty#d_v_ne]} )</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.3 Modes of variation claimed to exist (see details and references in Class 7/8 handout)

• Global: in Warao, a word has either all [p]s or all [b]s—no mixing

• Local: Vaux’s \( \text{[maikəɾəbɪləɾi]} \) ~ \( \text{[maikəɾəbɪləɾi]} \) ~ \( \text{[maikəɾəbɪləɾi]} \) ~ \( \text{[maikəɾəbɪləɾi]} \)

• Iterational: Vata /\( \text{ɔ̍ o̍ ká zā pīī īī}/ \rightarrow /\text{ɔ̍ o̍ kά ẓ ĩ īī} \) ~ /\text{ɔ̍ o̍ kά zʌ̄ʌ̄ ʌ̄ʌ̄ pīī īī} \) ~ /\text{ɔ̍ o̍ o̍ kʌ́ʌ́ ʌ́ʌ́ zʌ̄ʌ̄ ʌ̄ʌ̄ pīī īī} \) ~ /\text{ɑ̍ ɑ̍ o̍ o̍ kʌ́ʌ́ ʌ́ʌ́ zʌ̄ʌ̄ ʌ̄ʌ̄ pīī īī} \)

• At-most-one-target: Dominican Spanish \( \text{hablar fisno} \) style \( \text{as.bo.ga.do} \) ~ \( \text{a.bos.ga.do} \) ~ \( \text{a.bo.gasdo} \) ~ \( \text{a.bo.gasdo} \) ~ \( \text{a.bo.ga.dos} \), but \*\( \text{as.bo.gas.do} \), \( \text{(a.bos.ga.dos)} \), etc.

• At-least-one-target: Munro & Riggle 2004
  
  - In Pima [Uto-Aztecan, Arizona; Ethnologue groups it with Tohono O’odham, for 9,600 speakers], reduplication marks plurality, but in compounds plurality is expressed by reduplicating any non-empty subset of the conjuncts:
4. Derivational look-ahead

- Crowhurst & Michael 2005, Nanti [Arawakan, Peru, 480 speakers]:
  - an iterative rule shifting stress within a foot can be triggered by a violation of *CLASH:
    \[(\text{o.kò})(\text{ri.kì})(\text{tá.ka}) \rightarrow (\text{ò.ko})(\text{rì.kì})(\text{tá.ka})\]
    ‘she wore a nose-disk’
  - but stress can’t shift to a less-prominent (e.g., higher) vowel:
    \[(\text{i.kà})(\text{tsi.tò})(\text{ká.kse})\]
    ‘he held (it) in his talons’
  - What do you think of this form? How could it be analyzed with rules? OT?
    \[(\text{no.tà})(\text{me.sè})(\text{tá.kse})\]
    ‘I scraped (it)’

- OT may go too far with its look-ahead ability (see Kaplan 2011 for discussion)...
5. Constraint violability

- In a rules+constraints analysis of Nanti, for instance, we could have *CLASH
  - it’s frequently violated, though, so we have to restrict its power, either by giving it a limited set of rules to trigger, or by stipulating that some other constraint can block its triggered rules.
- In OT, at least the theory makes it clear how this kind of interaction works:

*CLASH >> RhType=IAMB...

<table>
<thead>
<tr>
<th>okorikjitaka</th>
<th>NonFinality</th>
<th>PROMINENCE</th>
<th>*CLASH</th>
<th>RhType=IAMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>!</td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...but PROMINENCE=INFoot >> *CLASH

<table>
<thead>
<tr>
<th>nosamerejaka</th>
<th>NonFinality</th>
<th>PROMINENCE</th>
<th>*CLASH</th>
<th>RhType=IAMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>!</td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>f</td>
<td></td>
<td>!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>!</td>
<td>!</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. Representations

6.1 Autosegmentalism

* We saw
  - features’ independence from segments (especially tone)
  - long-distance interactions between certain types of segments (e.g., sibilant harmony)
  - group behavior of certain features (e.g., place)

* Open questions
  - Is locality really all-or-nothing? Recall Martin’s Navajo sibilant harmony case:
    - The autosegmental account predicts that it doesn’t matter how much material intervenes between the two stridents—they are still adjacent as far as the [anterior] tier is concerned.
    - But Martin found that, in compounds, agreement is gradient: the more material intervenes between the two sibilants, the less likely they are to agree:
• (Additional twist, explored further in Martin 2007: most of the agreement in compounds is already there in the underlying forms!)

• Do we need a geometry to group features, or do we include in the evaluation metric principles that decide which features are favored to be referred to together in a spreading rule or an AGREE constraint?

6.2 Metrical stress theory

We saw...
• that stress is not like “real” features, not even autosegmental ones
• to deal with this, grids+feet

We didn’t really get to (among other things)
• proposals for additional hierarchical structure in phonological representations: feet grouped into prosodic words, then phonological phrases, then larger intonational phrases... (e.g., Selkirk 1978; Nespor & Vogel 1986; Hayes 1989).

7. The role of morphology

We looked at matters like...
• Cyclicity: derived words sometimes retain characteristics of their morphological predecessors
• Non-derived environment blocking: some processes apply only when triggered by morphology or (perhaps) other phonology
• Levels: within a language, subsets of the phonological processes are associated with subsets of the word-formation rules
• and relatedly, Lexical vs. post-lexical: there seem to be two syndromes—productive vs. not as much, gradient vs. categorical, carrying over into L2 vs. not, applying across word boundaries vs. not...

Martin 2004, p. 23
8. The role of syntax—which we didn’t talk about

8.1 Syntax influencing phonology
Kisseberth 2000, Chimwiini (dialect of Swahili formerly w/ 40,000 speakers in Somalia; most have emigrated to Kenya)

- Long vowels allowed only in the penult and antepenult of a “phonological phrase”.

- 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): ALIGN(XP,R,PPhrase,R)

  Why is the vowel of /maayi/ short in the first tree but long in the second?

  /maayi malada/          /maayi ni malada/
  NP
  |                     IP
  N’
  |                     D’
  N’ AP
  |           I
  N’ A
  ma.yi ma.la.da
  water fresh
  ‘fresh water’

  Most approaches to syntax’s influence on phonology focus on how syntactic structure defines domains like the phonological phrase, which phonology then refers to.

8.2 Phonology influencing syntax? Or at least word order...

- Embick & Noyer 2001, Latin: the clitic –que ‘and’, attaches after 1st word of 2nd conjunct:

  [boni pueri] [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 rebus–que
  around–and those places in things–and

  contræ–que lægem  dē prövinciæ–que
  against–and law from province–and  (p. 576)

- For more cases, and reviews of the large literature, see Schütze 1994, Shih et al. 2015
9. Some of my favorite things to think about in phonology, besides the above

- What is stored in the lexicon and what is computed online?
- How detailed is the lexical representation (Bybee 2001; Pierrehumbert 2002; Gahl 2008)? Can it contain redundant information?
- What is the phonology-processing interface like?
  - How does lexical retrieval for production influence pronunciation, e.g. single word vs. morpheme string (Hay 2003, but see Fiorentino 2006) priming and competition from other words (Baese-Berk & Goldrick 2009 and refs. therein, Martin 2007, (Smolensky, Goldrick & Mathis 2014))?
  - How does word recognition influence perception and lexicalization?
- What are the limits of learnability? Within the learnable, are some patterns more learnable than others?
- How can we get good data about competence? Especially, how can we tell what’s lexicon and what (if anything) is grammar?

10. Phonological things you can do after this course

- Take Ling 201A (Phonological Theory II) with Bruce Hayes next quarter (required for most of you)
- Check the phonology seminar (261ABC) schedule and feel free to drop in for whatever talks interest you: [www.linguistics.ucla.edu/talksaevents](http://www.linguistics.ucla.edu/talksaevents)
- Courses with a phonological element that are offered this year, but not every year (so take advantage)
  - Ling 202, Language Change
  - Ling 205, Morphology
  - Ling 211, Intonation, an in-depth look at the highest levels of the prosodic hierarchy
- Courses not offered this year that you can look out for in future
  - Ling 217, Experimental Phonology
  - Ling 236, Computational Phonology
- Look out for phonetics and phonology proseminars (251A/B). These are courses that focus on a special topic

References


Munro, Pamela & Jason Riggle. 2004. Productivity and lexicalization in Pima compounds. BLS.


Sanders, Nathan. 2002. Opacity and Sound Change in the Polish Lexicon. UC Santa Cruz.


