Class 19: Retrospective and prospective course wrap-up

To do
- Samoan due Friday
- Work on presentation and paper—do you have a name for the conference yet?

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

1. Learnability

Back to 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, will, given typical learning data, return an 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.
- ...and given just outputs (with or without their inaudible parts): it’s a lot harder. See Tesar et al. 2003.
  • A fair amount of phonotactic learning can be accomplished, which could later be used to learn alternations, though that step remains to be implemented (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.
  Try it out, using the same format for input files as in Bruce’s software, at web.linguist.umass.edu/~halp/ (Potts, Becker, Bhatt & Pater)

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 explanatory adequate theory.

The subset problem—say you are exposed to the following language:

<table>
<thead>
<tr>
<th>English</th>
<th>Tagalog</th>
<th>English</th>
<th>Tagalog</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘goat’</td>
<td>tagu</td>
<td>‘goats’</td>
<td>tagune</td>
</tr>
<tr>
<td>‘mango’</td>
<td>ale</td>
<td>‘mangos’</td>
<td>alene</td>
</tr>
<tr>
<td>‘corkscrew’</td>
<td>siri</td>
<td>‘corkscrews’</td>
<td>sirine</td>
</tr>
<tr>
<td>‘my goat’</td>
<td>taguba</td>
<td>‘my mango’</td>
<td>aleba</td>
</tr>
<tr>
<td>‘my corkscrew’</td>
<td>siriba</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- In a rule framework, what grammar would you learn?
- How do you think you would then react to the word sirab? Is this predicted by the grammar?
- 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 relevant alternations (Prince & Tesar 2004, Hayes 2004). The main idea is to force markedness constraints to be ranked as high as is consistent with the data.

1.2 Ranking bias within markedness or faithfulness constraints?

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

- 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 $\sigma_i$ 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 $\sigma$ than *ki, then we require 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.
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 $\begin{pmatrix} \alpha & \text{round} \\ -\alpha & \text{back} \end{pmatrix}$ to $\begin{pmatrix} \alpha & \text{round} \\ -\alpha & \text{voice} \end{pmatrix}$—equally simple, but not equally attested

Same issue arises with rules: why $[\alpha\text{round}] \rightarrow [\alpha\text{back}]$ but not $[\alpha\text{round}] \rightarrow [\alpha\text{voice}]$?

Along with constraint-learning itself, this is an open problem.

1.4 ☯ The role of phonetics

Well-known phonetic explanation for round/back affinity: lip rounding/protrusion and tongue backing, although articulatorily independent, share an acoustic effect (lower second formant).\(^1\)

---

\(^1\) Thanks to David Deterding’s Excel template (http://videoweb.nie.edu.sg/phonetic/vowels/measurements.html)
Obviously phonetics explains a lot of 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. Processes and constraints—some typological possibilities
   a. languages (and phenomena within a language) are similar in the structures they avoid (constraints), but not in the changes they apply (processes): e.g., *NČ, diverse repairs
   b. similarity in processes but not in constraints? not sure—how many different “problems” is, say, C-deletion a “solution” to?
   c. similarity in both: *VOICEOBRSTRENT#, devoicing only
   d. similarity in neither: ?? I guess very idiosyncratic phenomena like Palauan s→k / _l

   o What do you think about SPE’s and OT’s predictions here?

3. Process interaction: extrinsic ordering?

Feeding in Kalinga

<table>
<thead>
<tr>
<th></th>
<th>/d-in-opa/</th>
<th>*o]σ</th>
<th>MAX-V</th>
<th>NASASSIM</th>
<th>IDENT(place)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>di.no.pá</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>din.pá</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>dim.pá</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

   - In OT, it’s hard to get the counterfeeding candidate (b) if the language has nasal assimilation.

Bleeding in English:

<table>
<thead>
<tr>
<th></th>
<th>/kæt+z/</th>
<th>OBSTRUENTS AGREE VOICE</th>
<th>IDENT(voice)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>kætz</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>kæts</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>/bænd[z]/</th>
<th>OBSTRUENTS AGREE VOICE</th>
<th>IDENT(voice)</th>
<th>DEP-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>bænd[z]</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>bænd[ʃ]s</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>e</td>
<td>bænd[ʃ]s</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>f</td>
<td>bænd[ʃ]z</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

   - In OT, it’s hard to get the counterbleeding candidate (e)—here, 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 201 for how to fix this (not all of these proposals were developed with opacity in mind):

- containment (Matt Goldrick & Smolensky 1999)
- sympathy (McCarthy 1998)
- candidate chains (McCarthy 2007b)
- 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)
- 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 (Sanders 2002).

4. Process application

4.1 Self-feeding and self-bleeding

Recall Takelma from Anderson 1974:

- [a] becomes [i] if followed by [i]: /alxixamis/ → [alxīxīmis] ‘one who sees us’
- and any preceding [a]s follow suit: /ikūmananānikh/ → [ikūmininīnik] ‘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:

\[\text{œ} \rightarrow \emptyset / VC_\text{C(r)}V\]

/ty#davônè/ → [ty#davônè] or [ty#d_vônè] or [ty#dâv_ne], but not *[ty#d_v_ne] ‘you were becoming’

Again, expected in OT, where self-counterbleeding (Kikuyu??) would be unexpected.

4.2 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), standard OT doesn’t have much (plausible) to say about it (see Hyman & VanBik 2004)

Hypothetical case (pseudo-French):

only one target: /davônè/ → [davne]

multiple targets: /ty#davônè/ → [ty#d_vônè], *[ty#d_v_ne]

/ty#vudre#kã#sødã#lø#polisje.../ → [ty#vudre#k_k_sødã_lø_polisje], *[ty#vudre#k_s_k_l#polisje]
Eisner's (2002) proposal of 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></th>
<th>/ty#davone/</th>
<th>*CCC</th>
<th>*a-1</th>
<th>*a-2</th>
<th>*a-3</th>
<th>*a-4</th>
<th>*a-5</th>
<th>*a-6</th>
<th>*a-7</th>
<th>*a-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>[ty#d_vone]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>[ty#dav_ne]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>[ty#davone]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>[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>

4.3 Modes of variation claimed to exist (see details and references in Class 8/9 handout)

- **Global**: in Warao, a word has either all [p]s or all [b]s—no mixing
- **Local**: Vaux’s \[ma\text{ik}\text{t}^{\text{b}}\text{b}\text{u}\text{t}^{\text{b}}\text{il}^{\text{h}}]\] ~ \[ma\text{ik}\text{t}^{\text{h}}\text{b}\text{u}\text{t}^{\text{h}}\text{il}^{\text{i}}\] ~ \[ma\text{ik}\text{t}^{\text{b}}\text{b}\text{u}\text{t}^{\text{i}}\text{il}^{\text{h}}\] ~ \[ma\text{ik}\text{t}^{\text{h}}\text{b}\text{u}\text{t}^{\text{i}}\text{il}^{\text{h}}\] ~ \[ma\text{ik}\text{t}^{\text{b}}\text{b}\text{u}\text{t}^{\text{i}}\text{il}^{\text{h}}\] ~ \[ma\text{ik}\text{t}^{\text{h}}\text{b}\text{u}\text{t}^{\text{i}}\text{il}^{\text{h}}\]
- **Iterational**: Vata /\text{ká rá p\text{i}} → /\text{á ká zá p\text{i}} ~ /\text{á ká zá p\text{i}} ~ /\text{á ká zá p\text{i}} ~ /\text{á ká zá p\text{i}}
- **At-most-one-target**: Dominican Spanish hablar fisno style as.bo.ga.do ~ a.bos.ga.do ~ a.bo.gasdo ~ a.bo.ga.dos, but *as.bo.gas.do, (a.bos.ga.dos), etc.
- **At-least-one-target**: Munro & Riggle 2004: In Pima [Uto-Aztec, 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:

(3rd page of manuscript version)
5. 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:
  \((o.kò)(ri.kį)(tá.ka) \rightarrow (ò.ko)(ri.kį)(tá.ka)\) ‘she wore a nose-disk’
- but stress can’t shift to a less-prominent (e.g., higher) vowel:
  \((i.kà)(tsi.tò)(ká.kse)\) ‘he held (it) in his talons’
  - What do you think of this form? How could it be analyzed with rules? OT?
  \((no.tà)(me.sè)(tá.kse)\) ‘I scraped (it)’

OT may go too far with its look-ahead ability, but in most cases look-ahead capability is a welcome result.

The problematic predictions usually seem to involve two different phenomena (instead of a single phenomenon, stress, as in Nanti): e.g., does any language add or subtract syllables to move stress onto a more-prominent vowel??
The problem here may be not look-ahead but which processes can solve which kinds of problems. See Blumenfeld 2006 for examples and a theory.

6. 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, the theory makes it clear how this kind of interaction works:

<table>
<thead>
<tr>
<th>*CLASH &gt;&gt; RHType=IAMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>okorikjítaka</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>b</td>
</tr>
<tr>
<td>c</td>
</tr>
<tr>
<td>d</td>
</tr>
</tbody>
</table>

...but PROMINENCE INFOOT >> *CLASH

<table>
<thead>
<tr>
<th>nosamerejaka</th>
<th>NONFINALITY</th>
<th>PROMINENCE INFOOT</th>
<th>*CLASH</th>
<th>RHType=IAMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>((nò.sa)(mè.re)(já.ka))</td>
<td></td>
<td>*!</td>
<td>***</td>
</tr>
<tr>
<td>f</td>
<td>((no.sà)(mè.re)(já.ka))</td>
<td></td>
<td>*</td>
<td>**!</td>
</tr>
<tr>
<td>g</td>
<td>((no.sà)(me.rè)(já.ka))</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>h</td>
<td>((no.sà)(me.rè)(ja.ká))</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>
7. Representations

7.1 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:

\[
\text{\% agreement}
\]

\[
\begin{array}{c|c|c}
\text{sinbicans in adjacent} & \text{sinbicans in non-adjacent} & \text{\%}\text{\%} \\
\hline
\text{agree} & 42 & 30 \\
\text{disagree} & 58 & 70 \\
\end{array}
\]

- (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?

7.2 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 get to
- 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).
8. 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...

9. The role of syntax—which we didn’t talk about

9.1 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): $\text{ALIGN}(XP,R,P\text{Phrase},R)$

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

```
/maayi malada/  /maayi ni malada/
NP          IP
  |    |    |    |    |    |
N'    DP  I'  VP
   |      |
N' AP   D'   I
|      |      |
N A    NP V'
|      |      |
ma.yi ma.la.da V A ma.aa.yi ni ma.la.da
water fresh water cop. fresh
‘fresh water’       ‘water is fresh’
```

Most approaches to syntax’s influence on phonology focus on how syntactic structure defines domains like the phonological phrase, which phonology then refers to.
9.2 Phonology influencing syntax? Or at least word order...
Embick & Noyer 2001, Latin: the clitic –que ‘and’, attaches after first word of second conjunct:

\[ \text{good boys and girls} \]

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

\[ \text{in things-and} \]

10. Some of my favorite things to think about in phonology, besides the above

- What is lexicalized and what is computed online?
- How detailed is the lexical representation (Bybee 2001; Pierrehumbert 2002; Gahl 2008)?
- 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)?
  - 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?

11. Phonological things you can do after the winter break

- Take Ling 201 (Phonological Theory II) with Bruce Hayes next quarter (required for linguistics grad students)
- Check the phonology seminar (261ABC) schedule and feel free to drop in for whatever talks interest you www.linguistics.ucla.edu/Events.htm
- Ling 205, Morphology, is not a yearly event, so take advantage when it comes around; same goes for Ling 211, Intonation, an in-depth look at the highest levels of the prosodic hierarchy.
- Look out for phonetics and phonology proseminars (251). These are courses that focus on one topic, often the professor’s current research interest, and typically involve presentation of readings by students.

References [see web version for the rest of the pages]


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


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


Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. Language 83. 277-316.