Class 19: Course summary and maybe some synthesis

To do
- Finish Manam assignment.
- Do I have all your study questions?
- Mini-conference on Tuesday, 9 AM-1:15 PM, Rolfe 3126 unless you hear otherwise
- Turn in paper on Friday

Overview: Some summarizing, some stock-taking, some prospect, and maybe some synthesis.

1. Learnability

Back to the Chomskyan basics:
- a descriptively adequate grammar labels the utterances that a typical learner would encounter as grammatical (perhaps trivially, e.g. by listing them)
- an explanatorily 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, can, given typical learning data, return an explanatorily 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.

In OT, *under the assumption of a finite, universal constraint set*, the learnability problem becomes much easier: all the learner has to do is rank the constraints.
- Given input-output pairs, it's easy: see Tesar & Smolensky book, Prince & Tesar—though see also Tesar for the problem of inaudible output structure such as feet. See also Riggle dissertation.[1]
- There are also learning algorithms for variable/stochastic constraint rankings. See Boersma & Hayes; Goldwater & Johnson; Pater, Potts & Bhatt.
- Given just outputs, the problem is not solved. A fair amount of (phonotactic) learning can be accomplished, which could be useful for later learning of alternations, though this second step remains to be implemented—see Hayes.

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 dissertation, Hayes & Wilson.

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 grammar.

[1] Apologies to the students and to authors for the inadequate bibliographic information in this handout. Time is short.

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'

- 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). See the papers by Hayes and Prince & Tesar in Kager, Pater, and Zonneveld (eds.) *Fixing Priorities: Constraints in Phonological Acquisition*. 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 2007, drawing on work by Guion: Cross-linguistically, velar palatalization (k → tʃ) 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*.

- 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] (and 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 reduces to 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 much evidence the learner needs in order to increase the weight. If we give *ke a higher σ than *ki, then it is possible to learn the typologically anomalous *ki, tʃe* 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 $*\begin{bmatrix}\alpha\text{round}\\ -\alpha\text{back}\end{bmatrix}$ to $*\begin{bmatrix}\alpha\text{round}\\ -\alpha\text{voice}\end{bmatrix}$–equally simple, but not equally attested (same issue arises with rules: why do we see [αround] → [αback] but not [αround] → [αvoice]?)

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

2. Processes and constraints

Some typological possibilities:

- languages (and phenomena within a language) are similar in the structures they avoid (constraints), but not in the changes they apply (processes)
- similarity in processes but not in constraints
- similarity in both
- similarity in neither

There seem to be different outcomes for different phenomena—see Zuraw & Lu for an example and a little discussion.

3. Process interaction: extrinsic ordering?

Feeding in Kalinga

/d-in-opa/	*o]$_\sigma$	MAX-V	NASASSIM	IDENT(place)
a di.no.pá	*!			
b din.pá		*	*!	
☞ *c* dim.pá		*		*

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

Bleeding in English:

/kæt+z/	* [–son, αvoice] [–son, –αvoice]	IDENT(voice)
a kætz	*!	
☞ *b* kæts		*

/bɹæntʃ+z/	* [–son, αvoice] [–son, –αvoice]	*[+strid][+strid]	IDENT(voice)	DEP-V
c bɹæntʃz	*!	*		
d bɹæntʃs		*!	*	
e bɹæntʃɪs			*!	*
☞ *f* bɹæntʃɪz				*

- 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. You will probably see some proposals in 201 for how to fix this—they include containment (Goldrick/Smolensky), sympathy (McCarthy), candidate chains (McCarthy), output-output correspondence (Crosswhite, Benua, Steriade, Burzio, Kenstowicz, and others), targeted constraints (Wilson), local constraint conjunction (Smolensky, Łubowicz, Kirchner), stratal OT (Kiparsky), distantial faithfulness, *MAP constraints (Zuraw), comparative markedness (McCarthy)...not all of them capture all types of opacity.

But recall discussion from the rule-based literature about eliminating extrinsic ordering: Anderson; Koutsoudas, [Gerald] Sanders & Noll. These proposals either rule out certain interactions altogether (apparent examples would all have to fall into certain categories that are amenable to other explanations), or disfavor them. See [Nathan] Sanders for an OT perspective.

3.1 Interesting opacity example from last night's phonology seminar by Ania Łubowicz:

In Polish, coronals palatalize before front vowels, seemingly neutralizing with the underlying prepalatals:

(1) Coronal Palatalization: /t d n s z/ → [ć dź ń ś ź]/_e

		nominative sg.	locative sg.	gloss
t → ć	:	lis[t]	o liś[ć] + e	'letter'
d → dź	:	obia[d]	o obie[dź] + e	'dinner'
n → ń	:	ok[n] + o	o ok[ń] + e	'window'
s → ś	:	bruda[s]	o bruda[ś] + e	'dirty man'
z → ź	:	łobu[z]	o łobu[ź] + e	'troublemaker'

(copied from 1st page of Łubowicz's WCCFL 25 proceedings paper)

But underlying prepalatals take a different allomorph of the locative suffix:

(2) Original prepalatals

		nominative sg.	locative sg.	gloss
ć	:	liś[ć]	o liś[ć] + u	'leaf'
dź	:	narzę[dź] + e	o narzę[dź] + u	'tool'
ń	:	ko[ń]	o ko[ń] + u	'horse'
ś	:	łoso[ś]	o łoso[ś] + u	'salmon'
ź	:	pa[ź]	o pa[ź] + u	'type of butterfly'

(same source)

If there were a *e→u* rule, this would be counterfeeding; if *u→e*, it would be counterbleeding.

Łubowicz takes a different approach: she proposes a version of OT in which not just candidates but *scenarios* are evaluated (sets of input-output pairs), and there are constraints that enforce contrast between outputs of certain inputs. For example, since 'letter' and 'leaf' contrast underlyingly (in [high]), they must contrast on the surface too (in this case, the surface contrast is also in the feature [high], but realized on a following vowel instead of on the stem).

4. Process application

4.1 Repeated application (self-feeding)

Recall Takelma from Anderson.

[a] becomes [i] if followed by [i]: /alxīxamis/ → [alxīx**i**mis] 'one who sees us'

and any preceding [a]s follow suit: /ikūmanananink^h/ → [ikūm**i**n**i**n**i**nink^h] 'he will fix it for him'

unless a voiceless C intervenes: /alsegesak^{h}sanik^h/ → [alsegesa**k**^{h}s**i**nik^h] 'we keep nodding to one another'

- This is expected in OT (where non-self-feeding would be unexpected).

4.2 Multiple sites for application

Recall French (optional) schwa deletion from Anderson, following Dell:

ə → Ø / VC__C(r)V

/ty#dəvəne/ → [ty#dəvəne] or [ty#dvəne] or [ty#dəvne], but not *[ty#dvne] 'you were becoming'

- In OT, it's expected that you can't delete both—that would violate the very constraint that prevents deletion in /ʒɑk#dəve#partir/ → [ʒɑk#dəve#partir] 'Jacques had to go'
- We saw that the mechanics of embedding a principle like "don't apply a process in two sites if one of the applications would have bled the other" in a rule-based theory are difficult.

4.3 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.

Hypothetical case (pseudo-French):

only one target: /dəvəne/ → [dəvne]
multiple targets: /ty#dəvəne/ → [ty#dvəne], *[ty#dəvne]
/ty#vudre#kə#sə#kə#lə#pɔlisje…/ → [ty#vudre#k#sə#k#lə#pɔlisje], *[ty#vudre#kə#s#kə#l#pɔlisje]

Eisner's proposal of directional constraint evaluation is relevant, though. We index a copy of *SCHWA to each position (counting by segments, though other constraints might count differently) in the output string. Here's the left-to-right version:

	/ty#dəvəne/	*CCC	*ə-1	*ə-2	*ə-3	*ə-4	*ə-5	*ə-6	*ə-7	*ə-8
☞ *a*	[ty#dvəne]						*			
b	[ty#dəvne]					*!				
c	[ty#dəvəne]					*!		*		
d	[ty#dvne]	*!								

5. Derivational look-ahead

Nanti (Crowhurst & Michael):

- an iterative rule shifting stress within a foot can be triggered by a violation of *CLASH:
  (o.kò)(ri.kʃì)(táka) → (ò.ko)(rì.kʃi)(tá.ka) 'she wore a nose-disk'

- 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 (we can imagine some crazy repairs that never happen...can you add or subtract syllables to move stress onto a more-prominent vowel?), but in most cases look-ahead capability is a welcome result.

6. Process competition: Kiparsky's Elsewhere Condition

Recall: Depending on the formulation, this said something like "if two rules want to do different things to a form, and one rule is more specific than the other, apply only the more specific rule."

Example of a non-trivial case from Prince & Smolensky:

(a) V → [+stress] / __ C_0 V C_0 #
(b) V → [+stress] / __ C_0 #

- If both rules are applicable, which should be chosen, by the Elsewhere Condition?
- Give an example of a word where (b) would get to apply.
- How can we capture the priority of (a) in OT? What happens if we change the ranking?

Other cases are trivial in OT: if the specific constraint is ranked lower, it just becomes invisible.

7. Constraint violability

In Nanti, for instance, one might want to posit a *CLASH constraint—the constraint is 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 is able to block its triggered rules.

In OT, it's spelled out how this kind of interaction works:

*CLASH >> RHTYPE=IAMB...

	okorikʃitaka	NONFINALITY	PROMINENCE INFOOT	*CLASH	RHTYPE= IAMB
a	(o.kò)(ri.kʃì)(tá.ka)			*!	*
b	(o.kò)(rì.kʃi)(tá.ka)			*!	**
☞ *c*	(ò.ko)(rì.kʃi)(tá.ka)				***
d	(o.kò)(ri.kʃì)(ta.ká)	*!			

...but PROMINENCEINFOOT >> *CLASH

	nosamerejaka	NONFINALITY	PROMINENCE INFOOT	*CLASH	RHTYPE= IAMB
e	(nò.sa)(mè.re)(já.ka)		*!		***
f	(no.sà)(mè.re)(já.ka)			*	**!
☞ *g*	(no.sà)(me.rè)(já.ka)			*	*
h	(no.sà)(me.rè)(ja.ká)	*!			

8. The role of morphology

8.1 Cyclicity

Why do derived words sometimes retain characteristics of their morphological predecessors? There have been various proposals in OT, such as Output-Output correspondence, and cyclic application of the constraint ranking.

8.2 Lexical phonology

How can we capture phonology that happens with some Word Formation Rules (affixes, etc.) and not others? Non-Derived Environment Blocking (e.g., in the Polish example above, it's fine to have an alveolar before a front vowel within a root)?

Proposals include…

- Stratal OT: as in rule-based Lexical Phonology, there's a separate grammar for each morphological level, with a postlexical grammar applying last.
- Faithfulness constraints that apply not between input and output but between different outputs. These can be indexed to the morphological level at which one form is derived from the other, thus getting some lexical phonology effects.

NDEB is harder…

9. Representations

9.1 Autosegmentalism

The independent behavior of tone (and sometimes other features too), the existence of long-distance interaction between certain types of segments remain issues, and the group behavior of features are issues that tend to be addressed by the theory of representations rather than by the grammatical architecture.

Some remaining issues:

- Is locality really an all-or-nothing matter? Recall Martin's Navajo sibilant harmony case.
- What about the tendency of certain features to group together? Do we need a feature geometry (recall those trees we saw in class), or do we include in the evaluation metric principles that decide which features are favored to, say, spread together?

9.2 Metrical stress theory

We seem to need additional representational machinery to deal with stress and prominence; it's not like other features, not even autosegmental ones. We saw...

- grids
- feet
- grids with feet, grids with trees

It was mentioned in class but not really discussed that there are proposals for much more hierarchical structure in phonological representations: feet grouped into prosodic words, then phonological phrases, then larger intonational phrases... (e.g., Selkirk; Nespor & Vogel; Hayes).

10. What to 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 come to whatever talks interest you. www.linguistics.ucla.edu/colloquia
- Ling 205, Morphology is not a yearly event, so take advantage when it comes around; same goes for Ling 211, for an in-depth look at the highest levels of the prosodic hierarchy.
- Look out for phonetics and phonology proseminars (251). Pat Keating has one in winter (not sure what topic is), and Aaron Kaplan has one in spring on iterativity and related issues. These are courses that focus on one topic, often the professor's current research interest, and typically involve presentation of readings by students.
- Come to the UCLA-UC Berkeley Conference on Languages of Southeast Asia, Jan. 30-Feb. 1 in Royce Hall. There will be some pretty good phonology and phonetics talks (OK, and some good talks in other areas too)! See http://www.international.ucla.edu/cseas/ for updates.