Learning alternations

24.964—Fall 2004
Modeling phonological learning

Class 11 (2 Dec, 2004)

What we have seen so far

A variety of techniques for learning surface phonotactics, modeling the early stages of acquisition

- Statistical approaches
- Constraint demotion algorithms

But there's more to phonology…

E.g., German:

Sg.	Pl.	Gloss	vs.	Sg.	Pl.	Gloss
piːp	piːpə	‘peep’		diːp	diːbə	‘thief’
teːleːskoːp	teːleːskoːpə	‘telescope’		loːp	loːbə	‘praise’
voʁt	voʁtə	‘word’		moʁt	moʁdə	‘murder’
gʁaːt	gʁaːtə	‘edge’		gʁaːd	gʁaːdə	‘degree’
ʃtʁaɪt	ʃtʁaɪtə	‘fight’		aɪt	aɪdə	‘oath’
vɛʁk	vɛʁkə	‘work’		bɛʁk	bɛʁgə	‘mountain’
aspiːk	aspiːkə	‘aspic’		kʁiːk	kʁiːgə	‘war’

But there's more to phonology…

E.g., German:

Sg.	Pl.	Gloss	vs.	Sg.	Pl.	Gloss
piːp	piːpə	'peep'		diːp	diːbə	'thief'
teːleːskoːp	teːleːskoːpə	'telescope'		loːp	loːbə	'praise'
voʁt	voʁtə	'word'		moʁt	moʁdə	'murder'
gʁaːt	gʁaːtə	'edge'		gʁaːt	gʁaːdə	'degree'
ʃtʁaɪt	ʃtʁaɪtə	'fight'		aɪt	aɪdə	'oath'
vɛʁk	vɛʁkə	'work'		bɛʁk	bɛʁgə	'mountain'
aspiːk	aspiːkə	'aspic'		kʁiːk	kʁiːgə	'war'

But there's more to phonology…

E.g., German:

Sg.	Pl.	Gloss	vs.	Sg.	Pl.	Gloss
piːp	piːpə	'peep'		diːp	diːbə	'thief'
teːleːskoːp	teːleːskoːpə	'telescope'		loːp	loːbə	'praise'
voʁt	voʁtə	'word'		moʁt	moʁdə	'murder'
gʁaːt	gʁaːtə	'edge'		gʁaːd	gʁaːdə	'degree'
ʃtʁaɪt	ʃtʁaɪtə	'fight'		aɪt	aɪdə	'oath'
vɛʁk	vɛʁkə	'work'		bɛʁk	bɛʁgə	'mountain'
aspiːk	aspiːkə	'aspic'		kʁiːk	kʁiːgə	'war'

But there's more to phonology…

E.g., German:

Sg.	Pl.	Gloss	vs.	Sg.	Pl.	Gloss
piːp	piːpə	'peep'		diːp	diːbə	'thief'
teːleːskoːp	teːleːskoːpə	'telescope'		loːp	loːbə	'praise'
voʁt	voʁtə	'word'		moʁt	moʁdə	'murder'
gʁaːt	gʁaːtə	'edge'		gʁaːd	gʁaːdə	'degree'
ʃtʁaɪt	ʃtʁaɪtə	'fight'		aɪt	aɪdə	'oath'
vɛʁk	vɛʁkə	'work'		bɛʁk	bɛʁgə	'mountain'
aspiːk	aspiːkə	'aspic'		kʁiːk	kʁiːgə	'war'

- A central goal of phonological analysis is to figure out what the phonemic contrasts are, what the underlying form of each lexical item is, and what the grammar of alternations is

Why learning URs from alternations is hard

1. The search space is huge:

- From the earliest work in generative phonology, it has been assumed that URs may be considerably abstract, combining different features from different surface forms, and even containing representations that never appear in any surface form.

(Chomsky and Halle 1968; Hyman 1970; Schane 1974; Kenstowicz and Kisseberth 1977)

Why learning URs from alternations is hard

Quite persuasive case of URs that never surface: Fe'fe' Bamileke (Hyman 1972)

'to _'	'_ him'	'_ the child'	'_ the tree'	
vap	vab i	vab muu	vab thɯ	'whip'
kwat	kwal i	(kwa muu)	kwa thɯ	'attach'
cak	caɣ i	caɣ muu	caɣ thɯ	'seek'

Why learning URs from alternations is hard

2. Choice of UR and grammar are interdependent

- Learners do not yet know the full grammar that could confirm their choice of UR for alternating morphemes
- For example, German [loːp] ∼ [loːbə]
  - Learner may know that *[loːb] is illegal, but a grammar based on surface evidence alone may not map /b/ → [p] (might delete or epenthesize or something else instead)

(Tesar and Prince, to appear; Dresher 2004)

Why learning URs from alternations is hard

3. The data is sparse

- URs must be chosen on a morpheme-by-morpheme basis, yet for many words, learners do not yet have the forms that would reveal alternations. Thus, the possibility of alternation must be inferred in the absence of overt evidence.

(Kiparsky 1998; Harrison and Kaun 2000; McCarthy, to appear)

- As the dates in the references suggest, recent proposals have begun to take on problems (2) and (3), but have not really addressed the harder issue of constructing truly abstract representations

An important insight

Sommerstein (1974) "On phonotactically motivated rules" (*Journal of Linguistics* 19, 71-94)

- Segments often alternate in order to obey a phonotactic principle
- Example: German final devoicing satisfies whatever constraint bans voiced obstruents in codas
- (Also Kisseberth (1970), on conspiracies)

What this means for learning alternations

- In learning alternations, it could help to know first what the surface phonotactics are
- This is a strength of OT (at least as ranking algorithms are currently formulated)
- See also Hayes (1999) on this point

Using phonotactic knowledge to learn alternations

Tesar & Prince (to appear)

- Step 1: Learn some surface phonotactics, using IN=OUT assumption
- Step 2: Discover that some forms are morphologically related, and compare them, seeing what they have in common
- Step 3: Explore values of features that differ, to see which yield successful grammars

Tesar & Prince (to appear)

Example: a language like German

Root	Suffixed
tat	tat-e
tat	tad-e
dat	dat-e
dat	dad-e

Tesar & Prince (to appear)

Step 1: learn ranking to generate forms as individual words

- tat, dat, tate, tade, date, dade

Using the BCD algorithm:

$*\textsc{VoiCoda} \gg \textsc{Ident}(\text{voi}) \gg *\textsc{VoiObstr}, *[+\text{voi}][-\text{voi}][+\text{voi}]$

Tesar & Prince (to appear)

Step 2:

- Learn morphological relations between pairs of words (not formalized)
- Compare related forms to see what they have in common
  - [tat] $\sim$ [tade] share [taT] (T = coronal stop), differ in [$\pm$voi]
- Possible URs: /tat/, /tad/

Tesar & Prince (to appear)

Step 3: test possible URs, to see which one works (p. 16)

Input	Derives	Observed
/tat/	✓[tat]	[tat]
/tat-e/	*[tate]	[tade]

vs.

Input	Derives	Observed
/tad/	✓[tat]	[tat]
/tad-e/	✓[tade]	[tade]

- We pick /tad/

Tesar & Prince (to appear)

Life is not always so easy

- In example above, grammar from BCD happened to already derive devoicing, so choice of /tad/ automagically succeeds in deriving [tat]

- Unfortunately, real phonology usually allows more repairs than this

Tesar & Prince (to appear)

Throwing MAX into the mix

- New candidates: /tat/ $\rightarrow$ [ta], [at], etc.
- New ranking that emerges from BCD:

*VOICODA $\gg$ IDENT(voi), MAX $\gg$ *VOIOBSTR, *[+voi][−voi][+voi]

Tesar & Prince (to appear)

Now things do not sail through as easily

Input	Derives	Observed
/tat/	✓[tat]	[tat]
/tat-e/	*[tate]	[tade]

vs.

Input	Derives	Observed
/tad/	✓[tat], *[ta] (tie)	[tat]
/tad-e/	✓[tade]	[tade]

- Neither hypothesis works!

Tesar & Prince (to appear)

Proposal: in stalemate situations, try to use incorrect predictions as errors to drive learning

Table 16 Inconsistent winner-loser pairs for /tat/$_1$.

Image removed due to copyright considerations.

- Hypothesis /tat/ leads to inconsistency (see last two rows)

Tesar & Prince (to appear)

Proposal: in stalemate situations, try to use incorrect predictions as errors to drive learning

Table 17 The hypothesis /tad/$_1$ yields a consistent support.

Image removed due to copyright considerations.

- Hypothesis /tad/ leads to consistent ranking:

  *VOICODA $\gg$ MAX $\gg$ IDENT(voi) $\gg$ *VOIOBSTR, *[+voi][−voi][+voi]

- N.B.: BCD starts afresh each time the support changes! This is how MAX can get wedged in above where it used to be

Tesar & Prince (to appear)

Some issues:

- Reranking from scratch each time we need to discover something new seems undesirable
  - Perhaps not such a problem under the BCD, which is provably efficient (but also unable to learn some languages); would be more of a problem for an algorithm like the GLA

- This worked OK for choosing between /t/ and /d/; what about more complex cases? E.g., Bamileke
  - [cak] $\sim$ [caɣ]: /cak-/, /cag-/, /cax-/, /caɣ-/
  - [kwat] $\sim$ [kwal]: a number of featural differences
  - Number of hypotheses = $2^{|f|}$ ($|f| =$ number of features)

Tesar & Prince (to appear)

Some issues:

- Learner never stores the knowledge that [t] $\sim$ [d] alternations were resolved by setting up /d/
  - Subsequent learning is easier because grammar is in place, but $2^{|f_{\Delta}|}$ hypotheses must still be checked each time a new alternating pair is found
- Learner can never infer /d/ in the absence of overt [d]
  - Though see McCarthy (2004) regarding "free rides"

McCarthy (2004)

"Taking a free ride in morphophonemic learning"

(student presentation)