Inducing phonetic categories and phonological grammars

Joe Pater,
University of Amherst, Massachusetts

Cornell Workshop on Grammar Induction
May 16th, 2010

These comments were prepared in response to:

A single stage approach to learning phonological categories: Insights from Inuktitut

Brian Dillon, Ewan Dunbar, and William Idsardi

Ms, University of Maryland

and hopefully also serve as a suitable commentary on William Idsardi's Model-based Learning of Vowel Categories

Dillon *et al.*'s work builds an important bridge between two heretofore distinct research traditions:

1. Statistical modeling of phonetic category learning

2. Structuralist-generative analysis of phonological systems

Their proposal is a clever synthesis of the two, which results in a novel theory of the relation between phonetics and phonology.

From the MOG statistical models:

> The symbolic phoneme is replaced by a probability distribution over phonetic values.

The innovation:

> The phonological rule is replaced by a numerical transform of phonetic values.

As in exemplar-based models, there is no phonetics-phonology interface (see Kirchner and Moore 2009: ROA on phonological generalization in an exemplarist model)

Two challenges (recognized by Dillon *et al*.)

1. There is not yet a learning theory for the "rule component"

The single-stage transform from basic to context-specific values yields the effect of opaque counter-bleeding and counter-feeding rule ordering (in the manner of simultaneous application theories).

2. Languages analyzed with transparent orderings predicted to be impossible (though see Kaye on Dialect B!)

A basic assumption in Dillon *et al*.'s discussion:

Phonological generalizations require phonemes; they are done in terms of mappings from basic categories to context-specific variants

The alternative:

Phonological generalizations stated over “surface-level phones”, as in:

n-gram models of phonotactics
Finite-State models of phonotactics
constraint-based models of phonology

In the time remaining, I will:

1. Sketch an approach to constraint induction that I'll call *constraint projection*

2. Discuss an issue that category learning raises for this framework

Constraint projection (see Moreton handout for precedents, Albright slides for related work):

Constraints projected from individual pieces of learning data; generalization across data encoded in constraint weights

Constraints on surface representations:

Datum [bi]

LABIAL ↔ VOICE

Assign a reward of +1 to a voiced labial

LABIAL → VOICE

Assign a penalty of –1 to a voiceless labial

Boersma and Pater (2007: NELS), Pater, Moreton and Becker (2008: BUCLD, NECPhon)

NB Canadian raising analyzed with language-specific constraints in BP (2007)

Constraints on underlying representations:

Datum [bi] “bee”

“bee” → /bi/

Assign a reward of +1 to the UR /bet/ for meaning “bee”

or

Assign a penalty of –1 if meaning “bee” does not map to UR /bi/

Boersma (1999: Alberta), Apoussidou (2007: Amsterdam diss.), Eisenstat (2008: NECPhon), Jesney, Pater, Smith and Staubs (2010: LSA, Alberta)

A challenge inspired by Dillon *et al*.:

> If constraints are projected from individual pieces of learning data, how do learners come to generalize to appropriate language-specific categories?

For example:

> English learners exposed to word-initial “voiced” labial stops with a range of VOT values create a single category (“[b]”) that includes tokens that would be in separate [b] and [p] categories in other languages.

The constraints:

[p] Assign +1 to a word-initial voiceless unaspirated labial stop

[b] Assign +1 to a word-initial voiced unaspirated labial stop

Faith

Assign +1 if perceived voicing is represented accurately

“x” → [p], “x” → [b]

Assign +1 if meaning “x” maps to a phonological word with word-initial [p]/[b]

The problem (perception *à la* Boersma; see also Pater 2004, Pater, Stager and Werker 2004):

Input Perceived	Output "Meaning"	Hidden Categorical	Harmony	[b] 1	[p] 1	Faith 1	"bee" → [b] 1	"bee" → [p] 1
{bi}	"bee"	☞ [bi]	3	1		1	1	
{bi}	"bee"	[pi]	2		1			1
{pi}	"bee"	[bi]	2	1			1	
{pi}	"bee"	☞ [pi]	3		1	1		1

If constraints are simply given equal weight, the language has a too rich category structure.

The solution: a bias for low faithfulness

(D. Ohala 1996; Smolensky 1996; Hayes 2004; Prince and Tesar 2004; Jesney and Tessier to appear: NLLT; Jesney, Pater, Smith and Staubs 2010)

Simulation results

Initial Faith = exp(–3) = 0.05; Others = exp(1) = 2.72

Learning rate = 0.01; Noise.SD = 0.2; 10,000 iterations

Learning data:

{bi}, “bee”

{pi}, “bee”

Input Perceived	Output “Meaning”	Hidden Categorical	Harmony	[b] 3.86	“bee” → [b] 3.86	[p] 1.92	Faith 0.7	“bee” → [p] 1.92
{bi}	“bee”	☞ [bi]	8.42	1	1		1	
{bi}	“bee”	[pi]	3.83			1		1
{pi}	“bee”	☞ [bi]	7.71	1	1			
{pi}	“bee”	[pi]	4.54			1	1	1

One meaning, one category

http://people.umass.edu/pater/perceptron.R

Data: “bee” {bi} “hat” {pi}

Input Perceived	Output “Meaning”	Hidden Categorical	Harmony	Faith 5	“bee” → [b] 4.18	“hat” → [p] 4.26	[p] 2.75	[b] 2.69	“hat” → [b] 1.75	“bee” → [p] 1.75
{bi}	“bee”	☞ [bi]	11.87	1	1			1		
{bi}	“bee”	[pi]	4.5				1			1
{bi}	“hat”	[bi]	9.44	1				1	1	
{bi}	“hat”	[pi]	7.01			1	1			
{pi}	“hat”	[bi]	4.44					1	1	
{pi}	“hat”	☞ [pi]	12.01	1		1	1			
{pi}	“bee”	[bi]	6.87		1			1		
{pi}	“bee”	[pi]	9.5	1			1			1

Two meanings, two categories

For empirical evidence that category-learning infants are doing sound-object association, see Yeung and Werker (2009: Cognition); on joint inference of words and categories, see Feldman, Griffiths and Morgan (2009: CogSci).

In sum

Dillon *et al*. make an important contribution in bringing together statistical inference of hidden category structure and structuralist-generative approaches to contextual conditioning.

It seems interesting to also explore approaches that do not assume phonemic abstraction across contexts; these still need to deal with the hidden structure problem of category learning, on which I hope to have taken some baby steps here.

Thanks to Brian Dillon and the participants in Ling 751, Spring 2010 (*esp*. Karen Jesney, Kevin Mullin, and Robert Staubs) for discussion.

Funded by NSF grant BCS-0813829 to the University of Massachusetts, Amherst.