From the lexicon to a stochastic grammar

Michael Becker, University of Massachusetts, Amherst

The problem: Getting from a lexicon to a grammar

Phonological processes that are restricted to certain lexical items typically apply stochastically to novel items.

The behavior of novel items reflects lexical trends (Hayes & Londe 2006, Albright & Hayes 2003, Zuraw 2000, and several others)

→ **We need a way to project a stochastic grammar from the lexicon**

Case study: Hebrew plurals

Hebrew has two plural markers: *–im* on most masculine nouns
–ot on most feminine nouns

Most of the masculine nouns that exceptionally take *–ot* have [o] in their final syllable. The preference for *–ot* in masculine nouns that end in [o] applies productively to novel nouns, as seen in Berent, Pinker & Shimron (1999).

Analysis: Regular nouns allow mid vowels freely; irregular nouns want mid vowels to be licensed by an adjacent stressed mid vowel.

	Singular	Plural	
Regular	***alón***	***a l o n - ím*** [–hi] [+hi]	‘oak tree’
Irregular	***xalón***	***x a l o n - ót*** [–hi]	‘window’

A morphological constraint, φ-MATCH, requires the masculine *–im* on masculine nouns. φ-MATCH conflicts with mid vowel licensing:

/alon$_{MASC}$ + {im$_{MASC}$, ot$_{FEM}$}/	φ-MATCH	*MID
☞ a. alon-ím		*
b. alon-ót	*!	

/xalon$_{MASC}$ + {im$_{MASC}$, ot$_{FEM}$}/	*MID	φ-MATCH
a. xalon-ím	*!	
☞ b. xalon-ót		*

The solution: Clone a constraint, then keep track of lexical items

When lexical items demand conflicting rankings, BCD (Prince & Tesar 1999) detects inconsistency and stalls:

	φ-MATCH	*MID
alon-ím ~ alon-ót	W	L
xalon-ót ~ xalon-ím	L	W

The Pater (2006) solution: Clone a constraint to resolve the inconsistency. My proposal: make **both** clones lexically specific.

	*MID$_{\{xalon\}}$	φ-MATCH	*MID$_{\{alon\}}$
alon-ím ~ alon-ót		W	L
xalon-ót ~ xalon-ím	W	L	

Result: A categorical grammar for listed lexical items:

*MID$_{\{xalon, makom, ...\}}$ » φ-MATCH » *MID$_{\{alon, šaon, pagoš, ...\}}$

The relative number of lexical items on each clone defines a stochastic grammar:

*MID $_{24\%}$ » φ-MATCH » *MID $_{76\%}$

References
Albright, Adam and Bruce Hayes (2003) *Learning nonlocal environments*. Talk given at the LSA yearly meeting in Atlanta, Jan 4.
Becker, Michael, Christopher Potts and Joe Pater (2007) JavaTableau. Open-source software, UMass Amherst
Becker, Michael & Joe Pater (2007) OT-Help. Open-source software, UMass Amherst.
Beckman, Jill (2004) Positional Faithfulness, in *Optimality Theory in Phonology: A Reader*, John J. McCarthy (ed), 310–342.
Berent, Iris, Steven Pinker & Joseph Shimron (1999). Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition* **72**. 1–44
Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **21**. 43–58.
Hayes, Bruce, Bruce Tesar and Kie Zuraw (2004) OTSoft 2.1. Software package, http://www.linguistics.ucla.edu/people/hayes/otsoft/
Hayes, Bruce and Zsuzsa Londe (2006) Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* **23**. 59–104
Pater, Joe (to appear) Morpheme-Specific Phonology: Constraint Indexation and Inconsistency Resolution. In Steve Parker, (ed.) *Phonological Argumentation: Essays on Evidence and Motivation*. London: Equinox.
Prince, Alan and Bruce Tesar (1999) *Learning Phonotactic Distributions*. ROA-353.
Zuraw, Kie (2000). *Patterned Exceptions in Phonology*. Ph.D. dissertation, UCLA.

michael@linguist.umass.edu

Cloning specific constraints early

Exceptions without [o] in them are selected using a constraint that doesn’t depend on the root vowel, e.g. *STRESS/HI

/gag$_{MASC}$ + {im$_{MASC}$, ot$_{FEM}$}/	*MID	*ó́/HI	φ-MATCH
a. gag-ím		*!	
☞ b. gag-ót			*

*MID accounts for fewer lexical items, i.e. it is more specific:

	*MID	*ó́/HI	φ-MATCH
alon-ím ~ alon-ót	L	L	W
xalon-ót ~ xalon-ím	W	W	L
gag-ót ~ gag-ím		W	L

If we clone *STRESS/HI first, it will account for all exceptions, and the mid vowel effect will be lost:

☹ *ó́/HI$_{\{xalon, gag\}}$ » φ-MATCH » *MID, *ó́/HI $_{\{alon\}}$

We must clone *MID first to list words with [o] in them, then clone *STRESS/HI to account for words without [o]:

*MID$_{\{xalon,\}}$,*ó́/HI$_{\{gag\}}$ » φ-MATCH » *MID$_{\{alon\}}$, *ó́/HI $_{\{alon\}}$

Check out the implementation!

The input: A list of OTSoft (Hayes, Tesar & Zuraw 2004) tableaux, each representing a lexical item.

I use RCD to detect inconsistency, then clone a constraint that assigns the non-zero minimum of both W’s and L’s to the set of inconsistent ERC’s. This continues recursively, until the data becomes consistent, or can’t be made consistent by cloning.

The output: a single grammar that is categorical relative to existing lexical items, but can apply stochastically to novel items.

*MID$_{\{xalon,\}}$,*ó́/HI$_{\{gag\}}$ » φ-MATCH » *MID$_{\{alon\}}$, *ó́/HI $_{\{alon\}}$

The program uses code from JavaTableau (Becker, Potts & Pater 2007) and OT-Help (Becker & Pater 2007).