Michael Becker, Reed College
michael.becker@reed.edu
With Lena Fainleib, Tel Aviv University

UCLA Linguistics
February 2, 2009

Surface-based generalizations over lexical exceptions*

- Lexical exceptions impact the grammar. In Optimality Theory, this means that different lexical items respond to different constraint rankings.
- When markedness constraints are involved, exceptions are predicted to be learned in terms of their output properties, i.e. what Albright & Hayes (2003) call *product-oriented* generalizations.
- I offer an OT-based model that responds to exceptionality by building lexical information into the grammar. This lexically-enriched grammar allows speakers to state generalizations about their exceptions in terms of Universal constraints.
- I present results from an artificial language experiment that shows that speakers are biased to prefer product-oriented generalizations even in the absence of relevant evidence from the source language.

1 Exceptions as a part of the grammar

The observation: Speakers extract partial/statistical generalizations from their lexicon, and project them onto novel items (Bybee & Moder 1983; Pinker & Prince 1988; Prasada & Pinker 1993; Albright & Hayes 2003, and many others):

(1) I love [stɪb]-ing. I [stɪb] every day. Yesterday, I [stɪbd] / [stɪb].

(2) I love [snɛd]-ing. I [snɛd] every day. Yesterday, I [snɛdɪd] / [snɛd].

*This work benefitted greatly from the generous help of Ram Frost (Hebrew University). We thank Adam Albright, Iris Berent, Kathryn Flack, Lyn Frazier, John McCarthy, Joe Pater, and Matt Wolf for insightful comments and questions. We also thank the audience at the NYU linguistics department, especially Chris Barker, Lisa Davidson, Adamantios Gafos, Maria Gouskova, Gregory Guy, Susie Levi, Alec Marantz, and Jason Shaw. All errors remain our own.

The theory: All the words of a language can impact the grammar (Albright & Hayes 2003; Zuraw 2000), contra Pinker & Prince (1988), Berent et al. (1999, 2002). When a set of words behaves inconsistently under affixation, they cause the grammar to be inconsistent (Tesar 1998; Tesar & Smolensky 2000; Prince 2002). Inconsistent grammars are resolved by cloning (Pater 2006, 2008), which leads to lexical listing in the grammar.

(3)

/wɑnt + d/	*DD	Max	Dep
a. wɑntd	*!		
b. wɑnt		*!	
c. ☞ wɑntɪd			*

(4)

/kʌt + d/	*DD	Dep	Max
a. kʌtd	*!		
b. kʌtɪd		*!	
c. ☞ kʌt			*

The resulting grammar:

(5) *DD ≫ Dep$_{cut}$ ≫ Max ≫ Dep$_{want}$

And eventually:

(6) *DD ≫ Dep$_{cut,rid,set,spread,burst,shed,...}$ ≫ Max ≫ Dep$_{want,need,wait,fold,corrupt,pretend,...}$

A novel verb like [snɛd] can variably go with either clone of Dep.
Verbs that don't violate *DD are indifferent to the ranking of Max vs. Dep.

(7)

/stɪb + d/	*DD	Dep	Max
a. ☞ stɪbd			
b. stɪbɪd		*!	
c. stɪb			*!

2 Hebrew plurals

Hebrew has two allomorphs of the plural suffix, [-im] and [-ot].

The learner can discover that [-im] is masculine and [-ot] is feminine by looking at nouns that take different plural suffixes according to natural gender, and then by the completely regular agreement on adjectives and verbs.

(8)	sus-ím	ktan-ím	rats-ím	
	child-pl	little-pl	sing-pl	‘little horses are running’
	sus-ót	ktan-ót	rats-ót	
	child-pl	little-pl	sing-pl	‘little mares are running’

In the native vocabulary, however, masculine nouns can irregularly take [-ot], and feminine nouns can irregularly take [-im]. The true gender of the noun is revealed by agreement on adjectives and verbs (Aronoff 1994):

(9)	xalon-ót	ktan-ím	niftax-ím	
	window-pl	small-pl	open-pl	‘little windows are opening’
(10)	cipor-ím	ktan-ót	ʃar-ót	
	bird-pl	small-pl	sing-pl	‘little birds are singing’

In the loanword phonology, the plural suffixes don’t get stressed, and their selection is completely regular:

(11) blóg ∼ blóg-im ‘blog(s)’
banán-a ∼ banán-ot ‘banana(s)’

2.1 The lexicon

The masculine nouns that take [-ot] are not evenly distributed. Native masculine nouns from Bolozky & Becker (2006):

(12)

Final vowel	n	[ot]-takers	% [ot]-takers
u	1101	6	0.5%
i	464	8	1.7%
a	1349	39	2.9%
e	977	31	3.2%
o	523	146	27.9%
Total	4414	230	5.2%

Unsurprisingly, when given a masculine noun they haven’t heard before, Hebrew speakers like it better with [-ot] if it has an [o] in it (Berent et al. 1999, 2002; Becker 2009).

2.2 Learning Hebrew with Universal Grammar

The learner will identify [-im] as the masculine plural and [-ot] as the feminine plural, but will accept that the two affixes can compete for the same noun, even if its gender is known.

(13) ϕ-MATCH: Stems and affixes must agree in gender (cf. Wolf 2008 §2.4.2)

(14) LICENSE[o]: [o] must be stressed or adjacent to a stressed [o]

If a sound x is only allowed in some position, the position *licenses* the sound. Many languages require [o] to be licensed by the stressed syllable:

(15) Russian allows [o] only in the stressed syllable: *dó̲m-a* ‘at home’, *da̲m-áx* ‘at homes’.

(16) In most dialects of English, [o] can be unstressed (‘piano’, ‘fellow’), but in some dialects, unstressed [o] is not allowed (‘piana’, ‘fella’).

Other languages require [o] to be licensed by the word-initial syllable:

(17) Turkish native nouns allow [o] only in the first syllable of the word.

(18) Shona allows [o] in the root-initial syllable, and an initial [o] can license an [o] later in the word (Beckman 1997; Hayes & Wilson 2008)

Hebrew will turn out to be like Shona, but with stress: In Hebrew, [o] must be stressed, but a stressed [o] allows [o] to appear elsewhere in the word.

(19)	Regular	a l ó n	a l o n - í m	‘oak tree’
(20)	Irregular	x a l ó n	x a l o n - ó t	‘window’

The constraints ϕ-MATCH and LICENSE[o] are in conflict:

(21) Taking [-im] to satisfy ϕ-MATCH

alon$_{\text{MASC}}$ + {im$_{\text{MASC}}$, ot$_{\text{FEM}}$}	ϕ-MATCH	LICENSE[o]
a. ☞ a l o n - í m		*
b. a l o n - ó t	*!	

(22) Taking [-ot] to satisfy LICENSE[o]

xalon$_{\text{MASC}}$ + {im$_{\text{MASC}}$, ot$_{\text{FEM}}$}	LICENSE[o]	ϕ-MATCH
a. ☞ x a l o n - ó t		*
b. x a l o n - í m	*!	

When faced with conflicting evidence about the ranking, speakers will clone one of the constraints:

(23) LICENSE[o]$_{xalon}$ ≫ ϕ-MATCH ≫ LICENSE[o]$_{alon}$

And eventually, as the speaker learns all 523 Hebrew nouns with a final [o],

(24) LICENSE[o]$_{146\ items}$ ≫ ϕ-MATCH ≫ LICENSE[o]$_{377\ items}$

Given a novel native noun with [o] in it, the speaker will give it a 146/523 = 28% chance of taking [-ot].

(25) No reason to take [-ot] in loanwords

blóg$_{\text{MASC}}$ + {im$_{\text{MASC}}$, ot$_{\text{FEM}}$}	LICENSE[o]	ϕ-MATCH
a. ☞ b l ó g - i m		
b. b l ó g - o t		*!

2.3 Learning Hebrew without Universal Grammar

The Minimal Generalization Learner (MGL, Albright & Hayes 2002, 2003, 2006) learns morphology by creating rules of increasing generality:

(26)

	change	environment
[xalon] ∼ [xalonot]	Ø → [ot]	# x a l o n __ #
[aron] ∼ [aronot]	Ø → [ot]	# a r o n __ #
generalization:	Ø → [ot]	$\begin{bmatrix}+\text{son}\\+\text{cont}\end{bmatrix}$ o n __ #

The generalized rule can apply to any noun that ends in $\begin{bmatrix}+\text{son}\\+\text{cont}\end{bmatrix}$on, including an [im]-taker like [alon]. Each rule is associated with a success rate, or its rate of correct application.

As the MGL learns the nouns of Hebrew, it identifies two changes: Ø → [im] and Ø → [ot]. The environments for the two changes are different:

(27) Ø → [im] has a high success rate with [a], [e], [i], [u], and a somewhat lower success rate with [o]. But [a, e, i, u] don't make a natural class that excludes [o], so the general rule is: Ø → [im] / __ #.

(28) Ø → [ot] has a very low success rate with [a], [e], [i], [u], and a reasonable success rate with [o]. So we get two very general (sets of) rules:
(a) Ø → [ot] / __ # (low success rate)
(b) Ø → [ot] / o C __ # (reasonable success rate)

The learner discovers that having [o] in the root makes adding [ot] more likely.

2.4 With or without Universal Grammar?

The MGL takes a singular noun, and decides which change to apply to it. If the singular has [o] in it, it is more likely to take [-ot].

My learner creates a set of plural forms as candidates, and chooses the optimal one. If a plural has an unlicensed [o] in it, it is likely to be rejected.

The MGL makes decisions based on changes between singulars and plurals (source-oriented), and my learner makes decisions based on the plurals (product-oriented).

In real Hebrew, every noun that has [o] in its plural stem also has [o] in the singular, and almost every noun that has [o] in its singular stem keeps that [o] in the plural.

Real Hebrew is described equally well by both learners, because in real Hebrew, the [o] is present both in the singular and in the plural.

3 Artificial Hebrew

3.1 The languages

Singulars are plausible native Hebrew nouns with an [o] or an [i] in their final syllable. In the plural stems, [o] alternates with [i] and vice versa.

The choice of the plural suffix agrees with the plural stem vowel in the "surface" language and with the singular stem vowel in the "deep" language.

(29)

"surface" language		"deep" language	
amíg	amog-ót	amíg	amog-ím
axís	axos-ót	axís	axos-ím
azíx	azox-ót	azíx	azox-ím
aʃív	aʃov-ót	aʃív	aʃov-ím
adíc	adoc-ót	adíc	adoc-ím
apóz	apiz-ím	apóz	apiz-ót
agóf	agif-ím	agóf	agif-ót
acók	acik-ím	acók	acik-ót
abóʃ	abiʃ-ím	abóʃ	abiʃ-ót
alód	alid-ím	alód	alid-ót

After speakers learned one of the two languages, they were given nouns in the singular, and were asked to generate the plural.

(30) alíc, azíd, agív, apís, axíg, amíx, axíf, aníz, aʃík, afíʃ
abók, atóx, amóv, adós, aʃóg, aʃóc, aróf, ahóz, agóʃ, apód

The participants were 60 native speakers of Hebrew, who were students at the Hebrew University or at Tel Aviv University (average age: 23.4). Each participant was randomly assigned to one of the two languages.

3.2 The predictions

Prediction of my markedness-based approach: When a speaker creates a plural form, the Universal constraint LICENSE[o] wants [o] to be licensed. It doesn't care what any vowels might have been in the singular.

(31) In the "surface" language, LICENSE[o] is always satisfied, because all plural stem [o]'s go with [-ot]. LICENSE[o] helps the speakers make the right choice.

(32) In the "deep" language, plural stem [o]'s go with [-im], causing a violation of LICENSE[o]. LICENSE[o] discourages the speakers from making the right choice.

The "surface" language is predicted to be easier to generalize than the "deep" language.

Prediction of the MGL: In real Hebrew, the two available changes are Ø → [im] and Ø → [ot]. In each artificial language, there are two different changes:

(33)

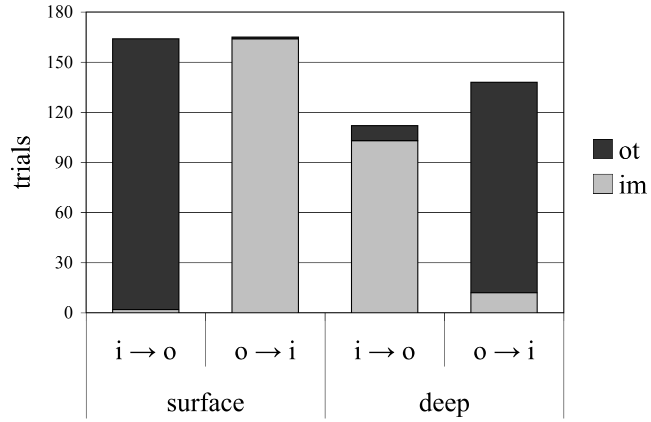
"surface" language	"deep" language
[o C] → [i C im]	[o C] → [i C ot]
[i C] → [o C ot]	[i C] → [o C im]

The changes of the artificial languages are not found in real Hebrew, and vice versa. Both the "surface" and the "deep" languages are equally distant from Real Hebrew, so they should be equally easy/hard to generalize.

3.3 Results

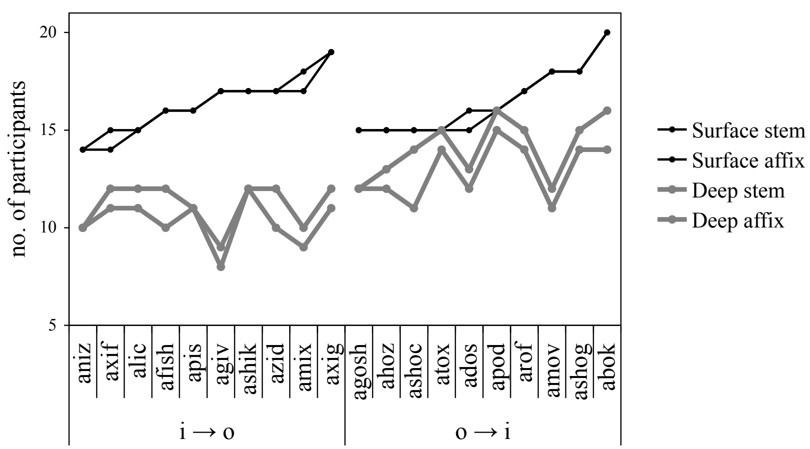
Speakers were more successful generalizing the "surface" language than the "deep" language. In the "deep" language, speakers were less successful with the [i] → [o] mapping relative to the [o] → [i] mapping.

(34) Trials with successful vowel changes



By-item analysis: same picture. Significantly more success with the "surface" group, for the stem vowel change (paired t-test, $t(19) = 7.36$, $p < .001$) and for the stem vowel change and affix selection (paired t-test, $t(19) = 9.25$, $p < .001$).

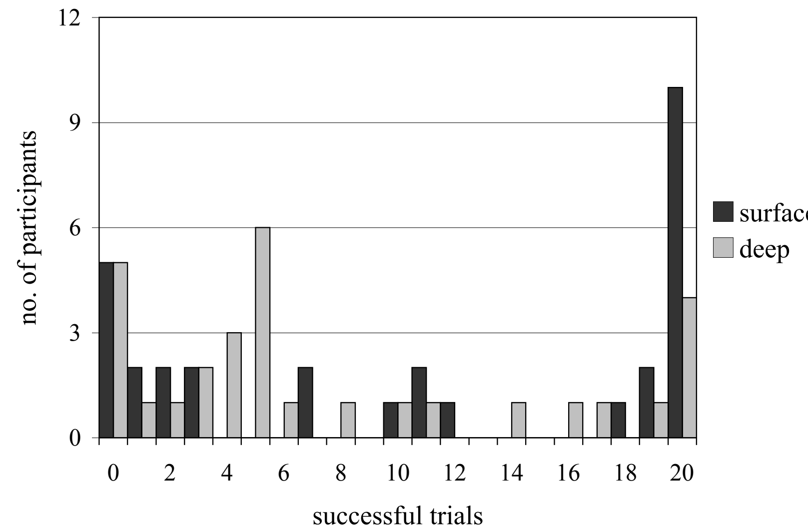
(35) Successful stem vowel change, with and without successful affix selection, by item



The "surface" group treats both vowel changes equally ($t(17.67) = .268$, $p > .1$). The "deep" group is significantly worse with [i]→[o] ($t(17.17) = 4.430$, $p < .001$).

The by-subject analysis shows an advantage for the "surface" participants, but the distribution is bimodal in both groups, so the stats are tricky.

(36) Successful stem vowel change and affix selection, by participant



At a cut-off point of 17, the difference between the groups is significant (Fisher exact test: odds ratio 3.736, $p < .05$). The choice of 17 for the cut-off point comes from the "surface" group, where no participant scored in the 13–17 range.

Mixed-effects logistic regression model, with *participant* and *item* as random effect variables:

(37)

	Estimate	SE	z	p
(Intercept)	0.761	0.723	1.054	0.292
"deep" language	−1.859	1.010	−1.843	0.065
singular [o]	0.091	0.286	0.317	0.752
"deep":singular [o]	0.658	0.374	1.760	0.078

The difference between the groups was not due to the random smartness of the "surface" language participants: Participants in both groups performed equally well on memorizing the singulars ($t(57.14) = .61$, $p > .1$).

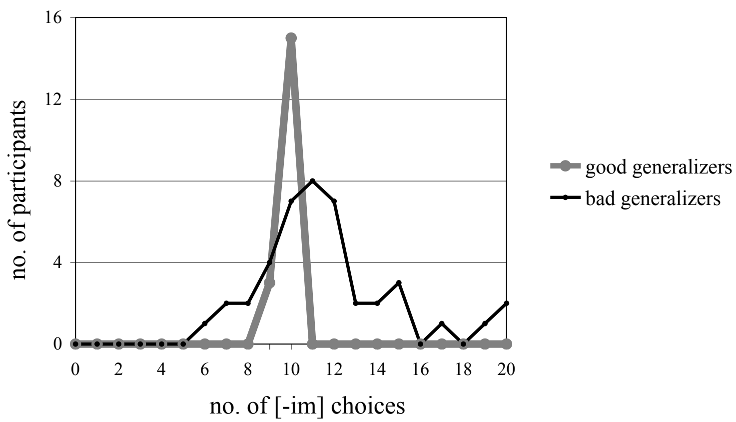
3.4 The relevance to natural language

There is good reason to think that the participants were not only treating the artificial languages as an extension of real Hebrew, but specifically that they were treating the artificial items as masculine native nouns of Hebrew.

(38) Speakers invariably generated plural forms with *final stress*, as in the Hebrew native phonology, and unlike Hebrew loanword phonology.

(39) Speakers chose [-im] 55.5% of the time, which is significantly more often than the expected 50% ($n = 60$, $M = .555$, Wilcoxon test with $\mu = 50\%$, $V > 1200$, $p < .05$). This shows that speakers treated the new words as masculine nouns, which are heavily biased towards [-im] in real Hebrew.

(40)



3.5 A role for phonotactics?

In native masculine plural forms, [-im] is by far the most common. In the native plurals, disharmonic vowel sequences are more common than harmonic sequences.

(41) Prefinal vowels in native plural nouns

	Masculine	Feminine	Total
...o-ím	527	5	532
...i-ím	437	7	444
...o-ót	147	178	325
...i-ót	6	1070	1076

The greater frequency of disharmonic vowel sequences is apparent in roots, too:

(42) Vowel sequences in native singulars

	All singulars	Di-syllabic masculines
i-o	286	107
o-i	132	8
i-i	126	2
o-o	21	8

If speakers went by raw phonotactics, they would have preferred the "deep" language, which has disharmonic sequences on the surface.

3.6 Summary

- Real Hebrew provides evidence for a correlation between having an [o] in the stem and taking the plural [-ot]. Real Hebrew does not provide evidence about the level of the generalization: Singulars, plurals, or the mappings between them.
- My markedness-based analysis predicts that speakers state the generalization over output forms, or plurals. I contrasted this with the MGL's rule-based analysis, which states the generalization over singular-plural mappings.
- When given an artificial language that puts [o]'s only in the singulars or only in the plurals, speakers prefer the language that pairs [-ot] with plural [o]'s.

4 Using the grammar to learn lexical statistics

For nouns that have [o] in their final syllable, there is a conflict between ϕ-MATCH and LICENSE[o]:

(43)

	LICENSE[o]	ϕ-MATCH
a. xalon-ót ≻ *xalon-ím	W	L
b. alon-ím ≻ *alon-ót	L	W

Speakers know two things about the distribution of [-ot]: It accounts for ∼30% of the nouns that have [o] in their stem, and for ∼2% of the nouns that don't.

What about [ot]-takers that don't have an [o] in their final syllable?

(44) gír ∼ gir-ím 'chalk(s)'
kír ∼ kir-ót 'wall(s)'

(45)

	License[o]	ϕ-Match
a. xalon-ót ≻ *xalon-ím	W	L
b. alon-ím ≻ *alon-ót	L	W
c. kir-ót ≻ *kir-ím		L
d. gir-ím ≻ *gir-ót		W

The learner can use any Universal constraint that happens to prefer [kir-ót] to [kir-ím], such as *σ́/High (Kenstowicz 1997; de Lacy 2004).

(46)

	License[o]	*σ́/High	ϕ-Match
a. xalon-ót ≻ *xalon-ím	W	W	L
b. kir-ót ≻ *kir-ím		W	L
c. alon-ím ≻ *alon-ót	L	L	W
d. gir-ím ≻ *gir-ót		L	W

The analysis uses constraints to partition the lexicon:

(47) License[o] identifies nouns that have [o] in their final syllable, and distinguishes [im]-takers from [ot]-takers.

(48) *σ́/High distinguishes [im]-takers from [ot]-takers no matter what the shape of the stem is.

To make sure that *σ́/High doesn't account for *all* [ot]-takers, License[o] must be allowed to list [ot]-takers *first*.

The nouns that don't have an [o] in them don't form a natural class ([a e i u]). They are the nouns that are left over after the nouns with [o] in them were taken care of.

(49) Cloning License[o] first

	Lic[o]$_{xalon}$	Lic[o]$_{alon}$	*σ́/High	ϕ-Match
a. xalon-ót ≻ *xalon-ím	W		W	L
b. kir-ót ≻ *kir-ím			W	L
c. alon-ím ≻ *alon-ót		L	L	W
d. gir-ím ≻ *gir-ót			L	W

(50) Cloning *σ́/High next

	Lic[o] *xalon*	Lic[o] *alon*	*σ́/High *kir*	*σ́/High *gir*	ϕ-Mtch
a. xalon-ót ≻ *xalon-ím	W				L
b. kir-ót ≻ *kir-ím			W		L
c. alon-ím ≻ *alon-ót		L			W
d. gir-ím ≻ *gir-ót				L	W

The resulting grammar:

(51) License[o]$_{xalon}$ ≫ *σ́/High$_{kir}$ ≫ ϕ-Match ≫ License[o]$_{alon}$, *σ́/High$_{gir}$

And as all the native nouns of Hebrew are learned,

(52) License[o]$_{146\ items}$ ≫ *σ́/High$_{84\ items}$
≫ ϕ-Match ≫ License[o]$_{377\ items}$, *σ́/High$_{3807\ items}$

When given a novel noun and asked to supply its plural, a speaker has a 28% chance of choosing [-ot] if the novel noun has an [o] in its final syllable, and a 2% chance of choosing [-ot] otherwise.

5 OT analysis of the artificial languages

(53)

"surface" language		"deep" language	
amíg	amog-ót	amíg	amog-ím
apóz	apiz-ím	apóz	apiz-ót

In the "surface" language, stems with [o] always take [-ot], [-im] otherwise:

(54) Grammar for the "surface" language:
LICENSE[o] ≫ ϕ-MATCH ≫ *σ́/HIGH

The "surface" language is a simplified, regular version of real Hebrew.

In the "deep" language, stems with [o] take [-im]: ϕ-MATCH ≫ LICENSE[o].
But if stems without [o] take [-ot], then *σ́/HIGH ≫ ϕ-MATCH.
Which predicts that all nouns take [-ot], so there is no consistent grammar for the "deep" language.

(55) Possible grammar for the "deep" language:
σ́/HIGH$_{\{\textit{aʃiv, axis, amig, azix, adic}\}}$ ≫ ϕ-MATCH ≫ *σ́/HIGH$_{\{\textit{agof, apoz, acok, aboʃ, alod}\}}$, LICENSE[o]

This grammar correctly predicts that [-im] will be selected 50% of the time, but this prediction is true *regardless of the vowel of the stem*. This grammar cannot be used to correlate the choice of plural suffix with the choice of stem vowel.

6 Conclusions

- Hebrew speakers know that having [o] in the stem is conducive to choosing [-ot], but real Hebrew doesn't tell them whether this generalization is stated over singulars or singular-plural mappings (source-oriented) or over plurals (product-oriented).
- In an artificial language experiment that put [o] only in the singular or only in the plural, speakers preferred the correlation of [-ot] with a plural [o], i.e. they preferred a product-oriented generalization.
- The Universal bias for product-oriented generalizations follows naturally from an OT-based analysis that uses markedness constraints.

References

Albright, Adam & Bruce Hayes (2002). Modeling English past tense intuitions with minimal generalization. In Michael Maxwell (ed.) *Proceedings of the sixth meeting of the ACL special interest group in computational phonology*. Philadelphia: ACL, 58–69.

Albright, Adam & Bruce Hayes (2003). Rules vs. Analogy in English past tenses: a computationalexperimental study. *Cognition* **90**. 119–161.

Albright, Adam & Bruce Hayes (2006). Modeling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.) *Gradience in Grammar*, Oxford University Press. 185–204.

Aronoff, Mark (1994). *Morphology by itself*. MIT Press.

Becker, Michael (2009). *Phonological Trends in the Lexicon: The Role of Constraints*. Ph.D. dissertation, UMass Amherst.

Beckman, Jill (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* **14**. 1–46.

Berent, Iris, Steven Pinker & Joseph Shimron (1999). Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition* **72**. 1–44.

Berent, Iris, Steven Pinker & Joseph Shimron (2002). The nature of regularity and irregularity: Evidence from Hebrew nominal inflection. *Journal of Psycholinguistic Research* **31**. 459–502.

Bolozky, Shmuel & Michael Becker (2006). Living Lexicon of Hebrew Nouns. Ms. UMass Amherst.

Bybee, Joan & Carol Lynn Moder (1983). Morphological classes as natural categories. *Language* . 251–270.

de Lacy, Paul (2004). Markedness conflation in optimality theory. *Phonology* **21**. 145–199.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* **39**. 379–440.

Kenstowicz, Michael (1997). Quality-Sensitive Stress. *Rivista di Linguistica* **9.1**. 157–187.

Pater, Joe (2006). The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In Leah Bateman & Adam Werle (eds.) *UMOP: Papers in Optimality Theory III*, Amherst, MA: GLSA. 1–36.

Pater, Joe (2008). Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.) *Phonological Argumentation: Essays on Evidence and Motivation*, Equinox. 1–33.

Pinker, Steven & Alan Prince (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* **28**. 73–193.

Prasada, Sandeep & Steven Pinker (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* . 1–56.

Prince, Alan (2002). Arguing optimality. ROA 562-1102.

Tesar, Bruce (1998). Using the mutual inconsistency of structural descriptions to overcome ambiguity in language learning. In P. Tamanji & K. Kusumoto (eds.) *Proceedings of NELS 28*. Amherst, MA: GLSA, 469–483.

Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Wolf, Matthew (2008). *Optimal Interleaving: Serial Phonology-Morphology Interaction in a Constraint-Based Model*. Ph.D. dissertation, UMass Amherst.

Zuraw, Kie (2000). *Patterned Exceptions in Phonology*. Ph.D. dissertation, UCLA.