The role of markedness in Hebrew exceptional plurals[1]

Michael Becker, UMass Amherst (michael@linguist.umass.edu)

Highlights:

- Speakers have detailed knowledge about their lexicon. This knowledge is grammatical, or biased by UG.
- General-purpose learners fail to replicate human behavior – universal biases are necessary for learning the lexicon like humans do.
- I propose an OT model that keeps track of lexical exceptions and projects statistical generalizations from them
- I present recent and currently running experiments that aim to show the grammatical nature of lexical exceptions in Hebrew plural allomorphy

1 Hebrew plural marking: The lexicon

Hebrew has two plural suffixes, *–im* and *–ot*, with a partially-predictable distribution. Above the word level, completely regular gender agreement reveals that *–im* is masculine and *–ot* is feminine.

(1)

balšan-iy-ót	mefursam-ót	marc-ót	
linguist-f-pl	famous-pl	lecture-pl	‘Famous linguists are lecturing’

In the loanword phonology, the plural suffix selection is completely regular also at the word level: If the right edge of the word is recognizable as a feminine suffix, *–ot* is selected, otherwise it’s *–im*.

(2)

artišók	artišók-im	* artišók-ot	‘artichoke’
kolég-a	* kolég-im	kolég-ot	‘colleague’
madám	? madám-im	??? madám-ot	‘madam (in a brothel)’

At the word level, native[2] nouns can take a mismatching suffix:

(3)

xalon-ót	gvoh-ím	
window-pl	high-pl	‘higher-ups’

naš-ím	xazak-ót	ve-yod-ót	
woman-pl	strong-pl	and-knowing-pl	‘strong, knowing women’

My data comes from an electronic dictionary (Bolozky & Becker 2006). Native masculine *ot*-takers are more common than native feminine *im*-takers:

(4)

	-im	-ot
masculine	3716	173
feminine	23	1196

The choice of plural suffix is unpredictable given the singular form. There are even some minimal pairs:

(5)

himnon-ím / himnon-ót	‘national anthem’ / ‘religious hymn’
tor-ím / tor-ót	‘line, queue’, ‘appointment’ / ‘turn’
maamad-ím / maamad-ót	‘stand’ / ‘status’
mazal-ím (tov-ím) / mazal-ót	‘good luck’ / ‘astrological sign’

And some variation:

(6)

šofar-ót / šofar-ím	‘shofar’
vlad-ót / vlad-ím	‘newborn’

The label “unpredictable”, however, misses the partial phonological predictability of the suffix for masculine native nouns:

[1] Thanks to Lyn Frazier, John McCarthy and Joe Pater for valuable feedback and discussion. Thanks to the audiences at Ben Gurion University and Tromsø University for a lively discussion and helpful comments. I also owe a huge debt of gratitude to Ram Frost of the psychology department at the Hebrew University. Remaining errors, if any, are my own

[2] I use the term “native” as a label for a synchronically-defined class of nouns, ignoring etymology. Native nouns are characterized here by movement of stress to the plural suffixes. See Becker (2003) for other properties of native nouns in Hebrew.

(7)

	–im	*–ot*	
a	1136	37	3%
e	788	26	3%
i	422	9	2%
o	300	**96**	**24%**
u	1070	5	<1%
Total	3716	173	

Berent, Pinker & Shimron (1999) show that speakers project this trend onto novel items, choosing *–ot* more often with nouns that have [o] in their final syllable.

2 The role of markedness

In OT, markedness constraints have the following three properties (among others):
a. They are universal (and possibly innate)
b. Their effect is general by default
c. They assess output forms only

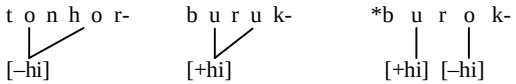
If Hebrew exceptions are organized using universal constraints, we expect to see the ***exceptional*** Hebrew pattern as a ***regular*** pattern in some other language.

In Shona (Beckman 2004), mid vowels (e, o) are licensed in initial syllables, or adjacent to another mid vowel:

(8) tonhor-, bover-, verer-, pofomar-
buruk-, simuk-, kumbir-, katuk-
*burok-, *boruk-, *burek-

Beckman analyzes the pattern using IDENT(high)$_{\sigma 1}$ » *MID » IDENT(high)

(9)

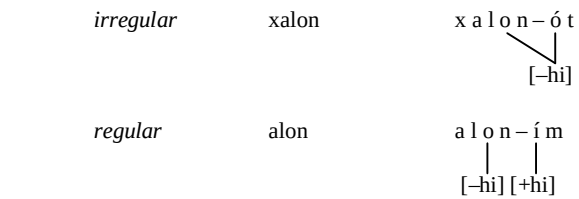


In Shona, the strong position that licenses mid vowels is the initial syllable. In Hebrew, it is the stressed syllable.

Licensing of mid vowels only in stressed syllables is very common, e.g. in Russian, many dialects of Arabic, Portuguese, and others.

Hebrew allows the stressed mid vowel to license an adjacent mid vowel:

(10)



Unlike in Shona, in Hebrew *alon-ím* is grammatical even with its unlicensed mid vowel. Mid vowel licensing in Hebrew is emergent.

A small number of Hebrew nouns avoid an unlicensed mid vowel by changing the root vowel: *xók ~ xuk-ím* 'law', *néc ~ nic-ím* 'hawk'.

(11)

/alon$_m$/+{im$_{mpl}$, ot$_{fpl}$}	IDENT(hi)	AGREE(gen)	*MID
alon-im			*
alon-ot		*!	
alun-im	*!		

(12)

/xalon$_m$/+{im$_{mpl}$, ot$_{fpl}$}	IDENT(hi)	*MID	AGREE(gen)
xalon-im		*!	
xalon-ot			*
xalun-im	*!		

(13)

/xok$_m$/+{im$_{mpl}$, ot$_{fpl}$}	*MID	AGREE(gen)	IDENT(hi)
xok-im	*!		
xok-ot		*!	
xuk-im			*

3 The generality of markedness

I am greatly indebted to Ram Frost of the psychology department at the Hebrew University, who generously offered to run my experiments at his laboratory for verbal information processing.

In the lexicon, feminine nouns overwhelmingly take *–ot* regardless of the stem's vowel. In fact, the single [o]-final feminine noun takes *–im* rather than *–ot*, so if anything, there is a slight preference for *–im*, not *–ot*,after [o].

(14)

	–ot	*–im*	
aa	124	3	2%
ea	62	1	2%
ia	453	7	2%
oa	42	2	5%
e	1	3	75%
ee	26	5	16%
i	1	1	50%
io	0	1	100%
Total	709	23	

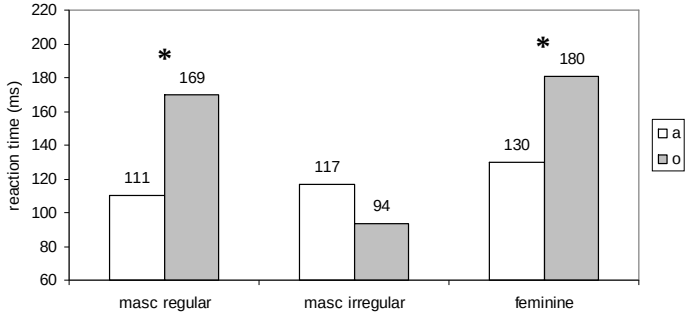
Experiment: Lexical decision task on actual roots with the correct or incorrect plural suffix:

(15)

masculine, regular		masculine, irregular		feminine, regular	
[a]	[o]	[a]	[o]	[a]	[o]
agasím	aloním	gagót	aronót	agadót	agorót
bcalím	egozím	ilanót	borót	avkót	bsorót
duxaním	egrofím	kravót	cinorót	kfafót	ofnót
kfarím	kiyorím	znavót	koxót	taxanót	tyotót
*agasót	*alonót	*gagím	*aroním	*agadím	*agorím
*bcalót	*egozót	*ilaním	*borím	*avkím	*bsorím
*duxanót	*egrofót	*kravím	*cinorím	*kfafím	*ofním
*kfarót	*kiyorót	*znavím	*koxím	*taxaním	*tyotím

Subjects: 54 native speakers of Hebrew, students at the Hebrew university.

(16) Real words



(17) ANOVA results:

	df	F	p
vowel	1	13.456	.001
vowel*gender	2	9.008	<.001

The vowel effect in the masculine nouns is expected, basically replicating the results from Berent, Pinker & Shimron (1999). The pleasant surprise is the vowel effect on the feminine nouns, since in the lexicon they overwhelmingly take *–ot*, regardless of the root's vowel.

An unbiased learner should not produce a vowel effect in the feminine nouns, since the in the lexicon, the vowel effect is limited to masculine nouns.

4 Assessing output forms

Experiment in the works: choosing plural suffixes with vowel alternations that are not attested in actual Hebrew.

		mapping	training	novel items
Language A	a.	[ao] → [ai]	acok ~ acikot apoz ~ apizot abol ~ abilim azod ~ azidim	agof, ados, axos, amox, atox, alog, aroš, adoc
	b.	[aa] → [au]	amag ~ amugot afaš ~ afušot anar ~ anurim axac ~ axucim	axaf, ayav, apas, azax, abak, ataz, adal, ayad
Language B	a.	[ai] → [ao]	acik ~ acokot apiz ~ apozot abil ~ abolim azid ~ azodim	agif, adis, axis, amix, atix, alig, ariš, adic
	b.	[au] → [aa]	amug ~ amagot afuš ~ afašot anur ~ anarim axuc ~ axacim	axuf, ayuv, apus, azux, abuk, atuz, adul, ayud

Speakers learn novel names for common nouns (all fruits and vegetables whose Hebrew name is masculine and takes *–im*). They learn the singulars and the plurals, and then asked to supply plurals for new nouns.

My prediction: When deriving novel nouns, speakers will form a strategy for choosing the plural suffix based on the vowel of the stem (either in the input or in the ouput).

My success will be devastating for any general-purpose learner that simply learns the lexicon without universal biases. Since no Hebrew noun has either [i] → [o] or [o] → [i], plural suffixes are predicted to be chosen at chance level.

5 Analyzing exceptions in Optimality Theory

I assume that UG acts as a filter on learning the lexicon. UG constrains the learning process, making speakers notice phonologically-motivated generalizations and ignore others.

When speakers derive novel forms, they **do not access their lexicon**. They only use their grammar, which has the phonologically-motivated aspects of the lexicon built into it.

5.1 Lexical statistics are kept following Inconsistency Detection

I propose a learning model in which speakers detect inconsistency in the grammar (Pater 2006) and then start keeping track of the behavior of individual items:

(18)

/alon$_{m}$/+{im$_{mpl}$, ot$_{fpl}$}	AGREE(gender)	*MID
→ alon-ím		*
alon-ót	*!	

(19)

/xalon$_{m}$/+{im$_{mpl}$, ot$_{fpl}$}	*MID	AGREE(gender)
→ xalon-ót		*
xalon-ím	*!	

(20) $*MID_{xalon}$ » AGREE(gender) » $*MID_{alon}$

As more words are learned, each instance of *MID will accumulate "weight", and this "weight" is projected onto novel words:

(21) $*MID_{xalon,\ bor,\ cror\ldots}$ » AGREE(gender) » $*MID_{alon,\ pašoš,\ šaon,\ egrof,\ kipod,\ sfog,\ xelbon\ldots}$

Thus, the ratio of *im*-takers and *ot*-takers is built into the constraint ranking. A novel word like *dóf* will be attracted by the heavier low-ranking *MID, so *dof-ím* is more likely than *dof-ót*.

5.2 Generalizations in terms of constraints

Plausible constraints of CON can keep track of plural suffix choice regardless of the root's vowel, so all vowels other than [o] should behave the same:

(22)

/gag$_m$/+{im$_{mpl}$, ot$_{fpl}$}	*MID	*ó/HI	AGREE(gender)
→ gag-ót			*
gag-ím		*!	

(23)

/dag$_m$/+{im$_{mpl}$, ot$_{fpl}$}	*MID	AGREE(gender)	*ó/HI
→ dag-ím			*
dag-ót		*!	

(24) *ó/HI$_{gag, gvul, kir…}$ » AGREE(gender) » *ó/HI$_{dag, xacav, kfar, perur, xut, bul, gir, kafil, mexir…}$

To make sure that roots with [o] in them are listed using *MID rather than *ó/HI, cloning must be biased towards more specific constraints, i.e. constraints that cover a smaller set of the data.

(25)

/xalon$_m$/+{im$_{mpl}$, ot$_{fpl}$}	*MID	*ó/HI	AGREE(gender)
xalon-ím	*	*	
→ xalon-ót			*

Identifying the most specific constraint in a set is a tricky, language-specific task, but see Tessier (2006) for a way to do it.

6 Conclusions

- Speakers use their universal grammar when they learn the words of their language.
- Lexical exceptions are learned in terms of rankings of universal constraints, and these rankings can be projected unto novel nouns.
- I am out to show that UG-less learning algorithms fail to model human behavior in a range of different ways.

References

Albright & Hayes (2002). Modeling English Past Tense Intuitions with Minimal Generalization. In Maxwell, Michael (ed) *Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology*. Philadelphia, July 2002. ACL.

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

Inkelas, Sharon, Aylin Kuntay, John Lowe, Orhan Orgun & Ronald Sprouse (2000). Turkish electronic living lexicon (TELL). Website, http://socrates.berkeley.edu:7037/.

Inkelas, Sharon & Cemil Orhan Orgun (1995). Level ordering and economy in the lexical phonology of turkish. *Language* **71**. 763–793.

Inkelas, Sharon, Cemil Orhan Orgun & Cheryl Zoll (1997). The implications of lexical exceptions for the nature of the grammar. In Iggy Roca (ed.) *Derivations and Constraints in Phonology*, Oxford: Clarendon. 393–418.

Pater, Joe (2004). Exceptions and optimality theory: Typology and learnability. Talk given in the Conference on Redefining Elicitation: Novel Data in Phonological Theory. New York University.

Tessier, Anne-Michelle (2006) *Biases and stages in phonological acquisition*. Ph.D. dissertation, University of Massachusetts, Amherst.

Vaux, Bert (2005). Formal and empirical arguments for morpheme structure constraints. Talk given at LSA Annual Meeting, San Francisco.