Over- and Under-generalization in Morphological Learning

Claire Moore-Cantwell :: University of Massachusetts, Amherst

Introduction

> Law of Frequency Matching: (Hayes et al., 2009) Speakers of languages with variable lexical patterns respond stochastically when tested on such patterns. Their responses aggregately match the lexical frequencies.

Surfeit of the Stimulus experiments demonstrate cases where speakers fail to reproduce the lexical frequencies on a wug-test. They shed light on the nature of the human language learner by demonstrating where and how it fails.

I present a natural language surfeit of the stimulus experiment, where participants **overgeneralize** one output form at the expense of others.

Hebrew Denominal Verbs

Hebrew verbs consist of three-consonant roots, with morphology expressed via vowel patterns:

gadal	**gidel**	**gadol**	mi**gdal**
he grew	*he raised*	*big*	*tower*

Ussishkin (1999) notes five ways that CVC nouns can be made into verbs, acquiring the necessary third consonant.

1. $\mathbf{C_1VC_2} \rightarrow \mathbf{C_1iC_2eC_2}$ Plain Consonant Doubling (CD)
   dam → **dimem** (*blood* → *he bled*)
2. $\mathbf{C_1VC_2} \rightarrow \mathbf{C_1VC_2eC_2}$ Vowel Overwriting (Ov)
   kod → **koded** (*code* → *he encoded*)
3. $\mathbf{C_1VC_2} \rightarrow \mathbf{C_1ijeC_2}$ Coronal Glide Formation (J)
   tik → **tijek** (*file* → *he filed*)
4. $\mathbf{C_1VC_2} \rightarrow \mathbf{C_1iveC_2}$ Labial Glide Formation (V)
   sug → **siveg** (*type* → *he sorted*)
5. $\mathbf{C_1VC_2} \rightarrow \mathbf{C_1iC_2C_1eC_2}$ Reduplication (RED)
   daf → **difdef** (*page* → *he paged through*)

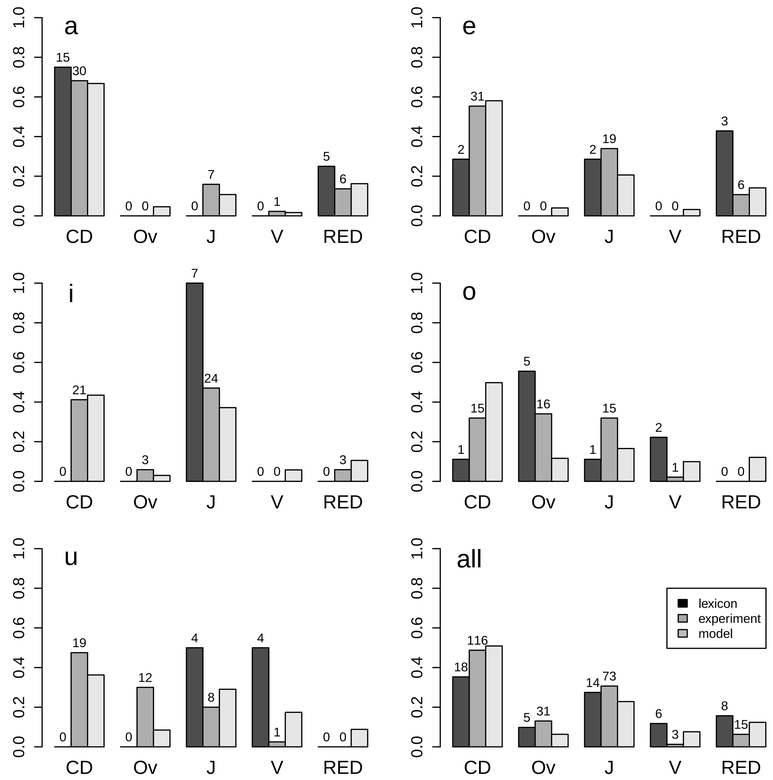
Methods

27 native Hebrew speakers performed a web-based production task. Wug-nouns were presented aurally in stories, followed by a fill-in-the-blank task.

החקלאים בתחנת החלל רגילים ___ את האדמה בחממות.

Results

About 45% of all responses were of one of the expected five types. Below, the percentage (y axis) of each output type is plotted by noun's vowel. Numbers represent raw counts.

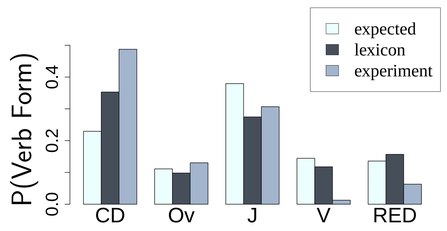


The frequency of verbal form depends on the noun's vowel in both the lexicon and the production data (a Poisson regression shows that this dependency is significant, $p < .05$).
Consonant Doubling is the most frequent form in participants' responses ($p < .05$), but **not** in the lexicon.

Discussion

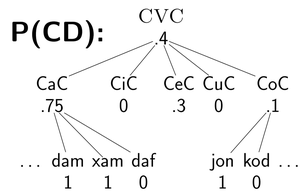
Consonant Doubling is **overgeneralized**: It occurs more often, and in more contexts in participants' responses than it does in the lexicon.

If participants matched exactly the conditional probabilities P(Verb Form | Vowel) in the lexicon, each verb form would occur with the probability in the 'expected' column →



Discussion

P(CD):



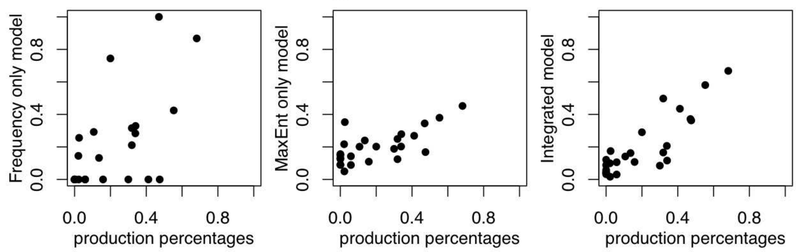
Learners of Hebrew are faced with many different levels of generalization over which to calculate probability. Neither the top level nor the by-vowel level matches participants' behavior. I argue that both levels of generalization are learned and used in production.

Modeling

I present a two-part model incorporating **phonological learning** as a Maximum Entropy grammar (Goldwater and Johnson, 2003), and **output type frequency**.

	Corpus probs	Max-V-Stm	Max-V-Aff	Ident-μ	Ident-V-lo	Ident-V-hi	Max-Lab	Ident-V-son	*Redupl.	H	P	*p*
Weights$_{corpus}$		**.8**	**3.0**	**0**	**2**	**1.6**	**2.6**	**1.8**	**0.8**			
/$C_1u_3C_2$/												
a. $C_1iC_2eC_2$	**0**	1					1			**-3.4**	**0.09**	*.48*
a. $C_1u_3C_2eC_2$	**0**		1							**-3**	**0.14**	*.30*
a. $C_1ij_3eC_2$	**.5**			1			1			**-2.6**	**0.23**	*.2*
a. $C_1iv_3eC_2$	**.5**			1				1		**-1.8**	**0.49**	*.03*
a. $C_1iC_2C_1eC_2$	**0**	1					1		1	**-4.2**	**0.04**	*0*

The output probabilities of the MaxEnt model are scaled by the overall probability of each output (verbal form) type. ($\chi^2_{diff} = 35.8$, $df = 4$, $p < .001$)



The generalization of CD as a default is consistent with Kam and Newport (2009), who find that adults tend to regularize inconsistent linguistic input. In the model, this form is chosen because it is (a) frequent and (b) applicable to all noun types.

References

Sharon Goldwater and Mark Johnson. Learning ot constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson, and Osten Dahl, editors, *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, pages 111–120, 2003.

Bruce Hayes, Kie Zuraw, Péter Siptár, and Zsuzsa Londe. Natural and unnatural constraints in Hungarian vowel harmony. *Language*, 85(4):822–863, December 2009.

Carla L. Hudson Kam and Elissa L. Newport. getting it right by getting it wrong: when learners change languages. *Cognitive Psychology*, 59:30–66, 2009.

Adam Ussishkin. The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output-output correspondence. *Phonology*, 16(3):410–442, 1999.

cmooreca@linguist.umass.edu