Over- and Under- generalization in learning derivational morphology*

Claire Moore-Cantwell

University of Massachusetts Amherst

1. Intro

Language users' knowledge of their native language involves extensive generalization as well as memorization. In morphological patterns which are exceptionful or variable across lexical items, each form must be memorized. For example, English speakers must memorize that the past tense of 'sleep' is 'slept', but the past tense of a similar word, 'steep' is 'steeped', not '*stept'. In exceptionful or variable patterns there is no rule that language learners can induce that would predict with 100% accuracy which forms take which type of morphology. Several studies (Hayes et al. 2009, Zuraw 2000, Becker et al. 2011, among others) demonstrate that speakers nevertheless have active knowledge of probabilistic trends across these memorized derived forms.

These probabilistic trends are dependent on properties of the input. For example, Becker et al. describe a situation in which words that end in alveolars have a different probability of taking a particular affix type than words that end in labials or velars. They show that subjects produce different patterns of affixation with alveolars than with labials or velars, rather than basing their responses solely on the overall frequency of each affixation type. In this paper, I present an experiment (a wug test, Berko 1958) demonstrating that speakers of Modern Hebrew are aware of probabilistic trends in a system of denominal verb formation. They know and use two types of probabilistic trends together: a probability distribution

*Materials can be found at `http://people.umass.edu/cmooreca/hebrewdvf/index.html`. I wish to thank Joe Pater, John McCarthy, Lyn Frazier, and Kie Zuraw for lots of wonderful advice and direction. I also wish to thank Roy Becker-Kristal, Shmuel Bolozky, Rachel Borden, Will Quale, John Griffin, Chris Cantwell, and especially Michael Becker, Lena Feinleib and Aynat Rubinstein for help preparing and checking stimuli, and preparing the web interface for the experiment. Further thanks are due to Wendell Kimper, Robert Staubs, John Kingston, Anne Pycha, Tom Roeper, Iris Berent, Bruce Hayes, Anne-Michelle Tessier, audiences at UMMM, NELS 42, and the 9th OCP, and especially the UMass 2nd year seminar for helpful discussion, comments, and questions. I am grateful to Michael Lavine and Caren Rotello for invaluable statistics help. This work was supported by a National Science Foundation Graduate Research Fellowship to the author, and is dedicated to the glory of God.

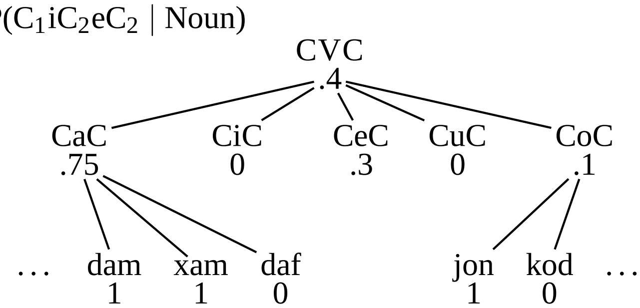
over outputs which is dependent on the type of noun, and a distribution which is dependent only on the type frequency of outputs, and not on any property of the noun.

2. Levels of Generalization

Language learners are capable of learning not only categorical generalizations, but also distributions of conditional probability over output types given an input type (P (output | input)). There are many levels of generalization over inputs which could be used by speakers to calculate the probability of an output form, some of which are more useful than others (Pierrehumbert 2006).

To give a brief preview of Hebrew denominal verb formation, suppose a morphological process where verbs are formed from nouns, and verbs must fit a CVCVC template, but nouns need not. Nouns of shape CVC can become verbs of two different shapes, say $C_1iC_2eC_2$ (ex: *dam* ∼ *dimem*) or CijeC (ex: *tik* ∼ *tijek*). Consider the diagram in (1). This illustrates the probability of a noun taking the first verbal form at different levels of generalization over noun forms. The top node represents the class of all nouns of shape CVC. At the next level, the class of nouns is broken down by vowel. The probability of taking the particular verbal form is different depending on which vowel a noun has. At the bottom are specific lexical items, which take a certain verbal form either all the time (100% probability) or none of the time (0%).

(1) $P(C_1iC_2eC_2 \mid \text{Noun})$



A learner faced with a pattern like this in her lexicon must choose which characteristics of the base are important for calculating the probability of a derived form. In order to correctly speak the language, she must learn that [dam] always becomes [dimem], but [daf] never becomes *[difef]. Beyond this, the learner has a choice of learning several probability distributions over nouns with different vowels or learning just one distribution over all nouns of shape CVC. The experiment reported here will provide evidence that learners learn both levels of generalization together.

3. Hebrew Denominal Verb Formation

3.1 Semitic verbs

In Semitic languages like Hebrew, verbs are minimally disyllabic, consisting of a three consonant ‘root’ together with an interleaved vowel pattern, as well as prefixes and suffixes. This root remains constant across related words:

(2) Three-consonant root:

ga**d**a**l**	**g**i**d**e**l**	**g**a**d**o**l**	mi**gd**a**l**
he grew	*he raised*	*big*	*tower*

McCarthy (1985) explained this structure with an autosegmental account, in which the roots exist on a separate tier from the vowel-pattern morphology. His representational scheme for Semitic morphology makes a strong claim that the consonantal root is a psychologically real and phonologically necessary entity. This position has since been challenged by Bat-El (1994), Ussishkin (1999, 2003), Bat-El (2003) who argue that an account which uses an input fully specified for vowel structure is more empirically adequate than one in which surface forms are derived from an underlying 3-consonant root. They argue for this based on denominal verb formation data.

Denominal verbs typically occur in the verb class or ‘binyan’ known as ‘piʕel’, which is of the shape ‘$C_1iC_2e_3$’ in the past tense (third person, masculine, singular), and ‘lə$C_1aC_2e_3$’ in the infinitive.

(3) Denominal Verbs (Bat-El 1994, pg. 577-579)

Base		Derived verb	
telegraf	*telegraph*	tilgref	*telegraph*
sandlar	*shoemaker*	sindler	*make shoes*
blof	*bluff*	bilef	*to bluff*

When there are three or more consonants in the noun, they are simply fit into the piʕel template, either in such a way as to preserve the cluster configuration of the input, or in the phonotactically optimal configuration. However, when there are only two consonants in the noun, the pattern is more complicated. Ussishkin (1999) identified five strategies existing in the lexicon, shown in (4).

(4) Forms of verbs derived from CVC nouns:

Structure				Label
DOUBLING:				
1.	C_1VC_2	→	$C_1iC_2eC_2$	Plain Consonant Doubling (CD)
	da**m**	→	**d**i**m**e**m**	
	blood		*he bled*	
2.	C_1VC_2	→	$C_1VC_2eC_2$	Vowel Overwriting (Ov)
	ko**d**	→	**k**o**d**e**d**	
	code		*he encoded*	
GLIDE FORMATION:				
3.	C_1VC_2	→	C_1ijeC_2	Coronal Glide Formation (J)
	ti**k**	→	**t**i**j**e**k**	
	file		*he filed*	
4.	C_1VC_2	→	C_1iveC_2	Labial Glide Formation[1](V)
	su**g**	→	**s**i**v**e**g**	
	type		*he sorted*	
OTHER:				
5.	C_1VC_2	→	$C_1iC_2C_1eC_2$	Reduplication (RED)
	da**f**	→	**d**i**fd**e**f**	
	page		*he paged through*	

Ussishkin observes that the form that the verb takes depends on the vowel of the noun. Nouns whose vowel is [a] typically take Consonant Doubling, nouns whose vowel is [i] typically take Coronal Glide Formation, nouns whose vowel is [u] take either Coronal or Labial glide formation, and nouns with [o] tend to take Vowel Overwriting. In his analysis, the noun's vowel is typically overwritten by the verbal vowel pattern.

(5) Emergence of the vowel pattern

a. MAX-V-STEM: *Assign a violation to every input stem vowel without an output correspondent*

b. MAX-V-AFFIX: *Assign a violation to every input affix vowel without an output correspondent*

c. MAX-V-AFFIX ≫ MAX-V-STEM

/dam + ie/	MAX-V-AFFIX	MAX-V-STEM
a. damem	*!	
→ b. dimem		*

But the vowel of the noun (stem) is preserved, when possible, in the form of a glide in the verb.

(6) High vowels are preserved

a. ID-μ: *Assign a violation to every output segment whose value for moraicity (syllabicity) does not match its input counterpart*

b. MAX-V-STEM ≫ ID-μ

/ti$_1$k/	MAX-V-STEM	ID-μ
a. tikek	*!	
→ b. tij$_1$ek		*
/su$_1$g/		
a. sigeg	*!	
→ b. siv$_1$eg		*

Since high vowels can mutate into their corresponding glides with minimal violations of faithfulness constraints, Glide Formation is the optimal output for them. Ussishkin does not consider the possibility of low vowels mutating into high glides. Instead, he considers the possibility of a glide being epenthesized when the underlying vowel is low. This is ruled out by ranking DEP above the constraint that is violated by consonant doubling, INTEGRITY (McCarthy and Prince 1995):

If one were to consider candidates with low vowels mutated into high glides, these would presumably be ruled out by additional faithfulness constraints, ranked above MAX-V. Thus Consonant Doubling would still be the best option.

The crucial insight of this analysis is the role of MAX-V. Because MAX-V is ranked high enough to exert influence - that is, it is ranked higher than some other faithfulness constraints, the noun's vowel has the chance to surface and influence the shape of the verb.

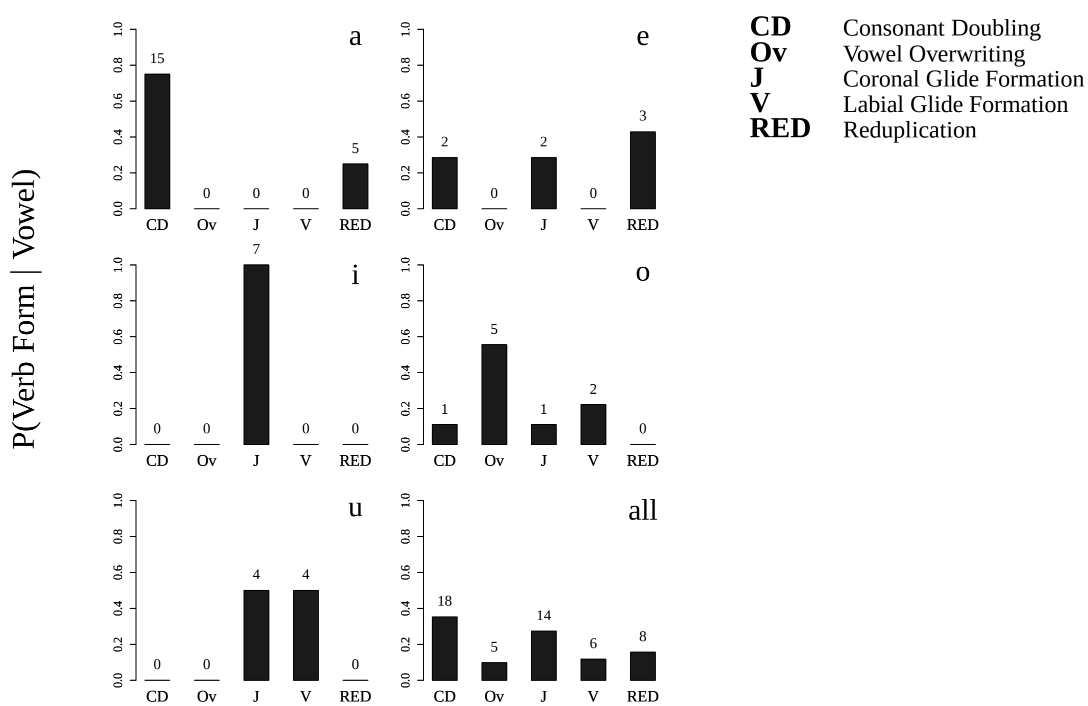
[1] Ussishkin (1999) argues that [v] 'counts' as the labial glide in Hebrew, even though it is not a sonorant. It is the consonantal counterpart of [u].

3.2 Patterns in the Hebrew lexicon

A corpus of 52 pairs of monosyllabic nouns and verbs derived from them was collected, sources for which include Ussishkin (1999), Bat-El (1994), Bolozky and Becker (2006), as well as a few Hebrew speakers. This list is very close to exhaustive, and contains many pairs in which one or both of the words are very rare, and unknown to many speakers. They also vary somewhat in semantic similarity. Some are very transparently related, such as [xam] ('hot') ∼ [ximem] ('to heat'), but others are more opaque, such as [has] ('silence') ∼ [hises] ('hesitate').

In (7), the counts of each verbal form type are plotted by noun's vowel. The height of the bars is the percentage of nouns with that vowel take that verbal form (for example, 75% of nouns with [a] take consonant doubling, and 25% of them take reduplication). Numbers above each bar indicate raw counts of that verbal form with that vowel (15 nouns with [a] take consonant doubling, 5 take reduplication).

(7) Distribution of verbal forms in the corpus



This plot illustrates the generalizations made by Ussishkin (1999), and also that they are exceptionful. Nouns with high vowels take glide formation, and the type of glide formation depends on the backness of the vowel. Nouns with the low vowel [a] preferentially take consonant doubling. Nouns with [o] take vowel overwriting, but also, though less often, plain consonant doubling and both kinds of glide formation. Nouns with [e] we would expect to take consonant doubling, and they do sometimes, but they also take coronal glide formation and reduplication.

4. Methods

4.1 Items

Twenty nonce nouns were constructed, all of the form CVC. The consonant pairs were chosen by a native speaker as pairs which do not form a real Hebrew word with any vowel. The nouns were constructed by randomly assigning a vowel to each consonant pair, such that each of the five vowels were equally represented (there were 4 each of nouns with a, e, i, o, u). These nonce words were then checked by two additional native speakers for wordhood.

4.2 Frames

The wug nouns were presented aurally in recorded short stories of a few sentences. The stories recounted some aspect of life aboard an imaginary space station in the atmosphere of Jupiter. Each introduced the wug noun as some object, tool, or property relevant to life on the space station. The stories were recorded by a native speaker of Hebrew who had phonetic training.

Participants listened to each story, then saw a written sentence with a blank in it where a verb derived from the newly-introduced noun should go. An example of the setup is given in (8).

(8) Example item

Spoken: In Bat-Jupiter, fruit can't be grown because it takes up too much space. Fruit has to be shipped from Earth. In order to keep the fruit from weighing too much and taking up space on the ship from Earth, the fruit has to be compressed somehow. A machine called a *mok* first removes all the water from the fruit, then removes the skin, seedsł, and any other part that won't be eaten, and finally vacuum-packs it.

Written: When they're in the middle of preparing a shipment, Earth technicians __ many kilograms of fruit per day.

4.3 Distractors

Subjects were told that they would be tested on their memory of the new words. After every three trials they saw a written version of a story, randomly sampled from all the ones they had previously heard. These written versions contained blanks where the wug noun belonged, and they were asked to provide the wug noun.

4.4 Organization

There were two groups of noun-story affiliations. Group 1 was created by randomly assigning nouns to stories, and Group 2 was created by assigning nouns to the same stories, but in the reverse order from Group 1. Subjects were randomly assigned to one of the two groups, and never saw the same noun in two different stories. The order of presentation of items within the group was randomized independently for each subject, and each distractor memory task was randomly selected from the set of items shown so far.

4.5 Participants

Participants were recruited through the web. The experiment was online, and participants arrived at it through facebook and google ads, craigslist postings, mechanical turk, and emails from friends and relatives. A total of 27 adult native speakers took the survey (participants were counted as native speakers if they listed Israel as their birthplace and Hebrew as one of their household languages in an exit survey).

5. Results

5.1 Variety of responses

For the fill-in-the-blank task, there was a wide variety of responses. Many participants responded with real Hebrew verbs instead of deriving a new verb from the wug-noun. Subjects also frequently responded with a verb derived from the wug-noun, but in a binyan other than the piʕel, thus not following one of the five patterns laid out in 4. In (9) is a breakdown of response types. The category ‘Other’ contains blank answers, instances where a subject filled in the blank with the noun instead of a verb, and responses which contain different consonants than the noun had.

(9) Variety of Responses

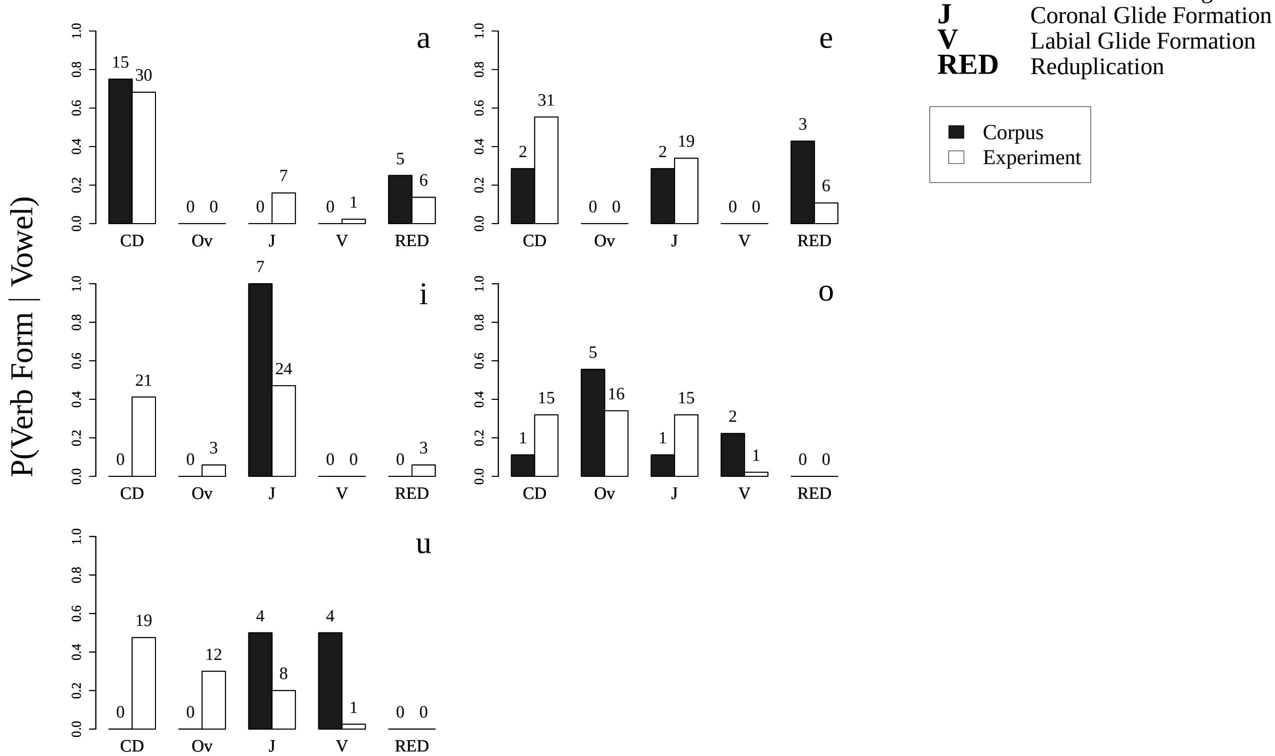
Response	Count	Percentage
Denominal verb forms:	240	44%
Other verb forms:	170	31%
Existing verbs:	85	16%
Other:	45	8%

In the following analysis, I will focus on the denominal verb forms, of the type in (4).

5.2 Analysis

Dark bars in (10) indicate corpus counts, and light bars on the right indicate responses in the experiment. The number above each bar indicates the raw count of that verbal form derived from that vowel. The height of the bars indicates the percentage of the time that vowel takes that verbal form. For example, the top left-hand chart indicates that in the corpus, nouns with the vowel [a] take consonant doubling about 75% of the time, and reduplication about 25% of the time. The 75% consonant doubling is made up of 15 cases, and the 25% reduplication is 5 cases.

(10) Comparing the experimental results to the corpus.



This graph demonstrates that participants have mismatched the lexical frequencies in important ways. The most striking thing is that they have overgeneralized consonant doubling, and are using it in contexts in which it does not occur in the lexicon - in particular, they are using it nearly half the time with the vowels [i] and [u], whereas it never occurs with those vowels in the lexicon. Another interesting overgeneralization is vowel overwriting in nouns with [u]. In the lexicon this never happens, but in the production data, it is nearly as common as overwriting with [o].

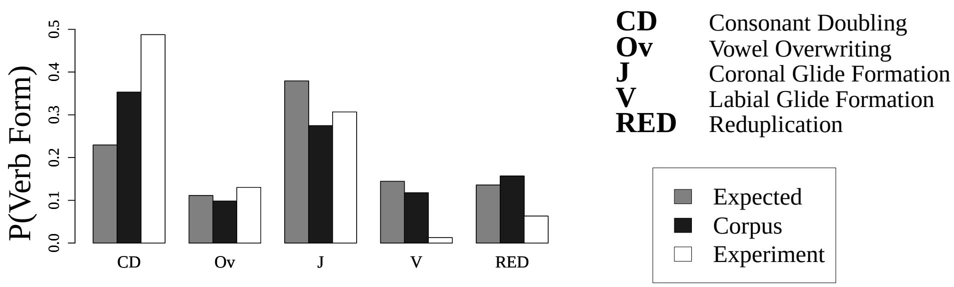
Coronal Glide formation has also overgeneralized. It occurs fairly often with [o] and [a], although it rarely or never occurs with those vowels in the lexicon. A notable undergeneralization is the use of Labial Glide formation. It was produced only 3 times in the whole experiment, a great deal less often than any other form.

The probability distribution of experimental responses is 'smoothed' with respect to the lexical distributions. There are more exceptions, and less clear generalizations. What verbal form was produced did depend on the noun's vowel significantly, however: a poisson regression was conducted on the experimental results, using verbal form and vowel as predictors of frequencies. When an interaction term was added, the model fit significantly better (AIC dropped from 175 to 125, and a χ^2 test yielded a $p<.001$).

In the lexicon, some output forms are more common than others overall, and this is entirely the result of some input vowels being more common than others. Nouns with [a] are more common than nouns with [o], which leads to Consonant Doubling being more frequent than Vowel Overwriting as an output form. In the experiment, however, all vowels occurred equally often, and were roughly equally likely to be given a denominal-type verb (The graph consists of 44 a-nouns, 56 e-nouns, 51 i-nouns, 47 o-nouns, and 40 u-nouns).

Consonant Doubling is more common in the experiment than would be expected if participants were matching the lexical probabilities of each verbal form given each vowel. (11) illustrates this. Black bars represent type frequency of each verbal form in the corpus, over all nouns with all different vowels. Grey bars represent the expected probability of each verbal form if each vowel occurred equally often. White bars represent the actual probabilities in the experiment of each verbal form over all noun types.

(11) Comparing the experimental results to the corpus.



Consonant Doubling is the most common verbal form in the lexicon (though not significantly more common that Coronal Glide Formation, p=.48), which is essentially by accident of the distribution of input types. It is also the most common in the experimental results (this time significantly, $p < .002$), but this time the input types are balanced, so it is not by accident.

6. Modeling

6.1 Maximum Entropy

In this section I will develop a model for the experimental data using a Maximum Entropy grammar (Goldwater and Johnson 2003, Hayes and Wilson 2008). MaxEnt grammars are a type of Harmonic Grammar (Smolensky and Legendre 2006, Pater 2009), and use interaction of weighted violable constraints to produce output. Instead of producing a single output for an input form, MaxEnt models produce a probability distribution over candidate outputs. I use the batch learning algorithm developed by Wilson and George (2009), which uses a hill-climbing optimization algorithm to find weights which produce the best match of its output probabilities to its training data. The constraints to be considered are given in (12).

(12) Constraints for consideration

a. MAX-V-STEM *Don't delete a vowel from the stem*

b. MAX-V-AFFIX *Don't delete a vowel from the piʕel vowel pattern*

c. ID-μ *Don't change the moraicity (syllabicity) of a segment*

d. ID-V-LO *Don't change the lowness of a vowel*
e. ID-V-HI *Don't change the height value of a vowel*
f. MAX-LAB *Don't delete a labial feature*
g. ID-V-SON *Don't change the sonority of a vowel*
h. *REDUPLICATION *Don't be reduplicated*

The roles of the first three are discussed above, in section 3. ID-V-HI penalizes changing mid and low vowels into glides, and ID-V-LO penalizes changing just low vowels into glides. Changing [a] into a glide violates both of these constraints. I've additionally included ID-V-SON because it distinguishes the coronal glide [j] from the labial 'glide' [v]. *REDUPLICATION is included to distinguish reduplicative forms from consonant doubling forms.

The optimization algorithm uses a regularization term which prevents overfitting by penalizing exhorbitantly high constraint weights. Wilson and George (2009) use an 'L2' prior, which specifies for each constraint weight a Gaussian prior with mean μ and variance σ^2. The value for μ is the 'preferred value' for the constraint, and the smaller the value of σ^2, the less the constraint weight will deviate from this value.

In order to model the smoothing that happens in the experimental data with respect to the lexicon, I employ a very strong regularization term, which restricts constraint weights to have a variance of 5 around a mean of zero (though constraint weights cannot be less than zero). I also place a specific bias on MAX-V-STEM by setting *its* σ^2 at 0.5. Violations of MAX-V-STEM are pervasive throughout the rest of the Hebrew verbal system, which constitutes a much larger dataset than the CVC-base denominal verbs. If the MaxEnt learner was supplied with data from the entire Hebrew verbal system, rather than just from the CVC-base denominal verbs, it would presumably be forced to weight MAX-V-STEM very low. Biasing the weight toward zero is an artificial way of producing this effect. (13) is a tableau illustrating the weights learned from the corpus when these regularization values are in place.

(13) Weights learned from training on the corpus compared to the experimental probabilities

	Corpus probs	MAX-V-STEM	MAX-V-AFF	IDENT-μ	IDENT-V-LOW	IDENT-V-HIGH	MAX-LAB	IDENT-V-SON	*REDUPLICATED			*Experiment probs*
Weights$_{corpus}$		**.8**	**3.0**	**0**	**2**	**1.6**	**2.6**	**1.8**	**0.8**	**H**	**P**	*p*
/$C_1u_3C_2$/												
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*

(13) illustrates the mechanics of MaxEnt. A candidate's harmony value is calculated by multiplying each constraint weight by the number of times a candidate violates that constraint, and then summing over these values (in other words, it is the dot product of the vector of weights and that candidate's vector of violation scores).

(14) $H = \sum w_i C_i(x)$
Where w is a vector of constraint weights, C is the set of constraints, so that $C_i(x)$ is the number of times x violates constraint C_i.

A candidate's probability is calculated from the harmony score by taking its negative exponent and dividing by the sum of that value for all candidates assigned to the same input.

(15) $P_x = e^{(-H_x)} / \sum_i e^{(-H_i)}$

These weights result in an output distribution for this input which is a great deal more smoothed than the training distribution (every possibility gets some probability). This distribution is still far from matching the experimental results. Whereas about half of the experimental responses were consonant doubling, this model only produces consonant doubling 6% of the time for this input. The model produces Labial Glide Formation about half the time for this input, while it barely occurs at all in the experiment.

6.2 Incorporating output type frequency

In the experiment, participants' responses showed some dependency on the noun's vowel, but the probability of each response also depended on the overall type frequency of that verb type in the lexicon. I will incorporate this knowledge into the model simply by scaling the output probabilities of the MaxEnt model by the type frequencies of each output type.

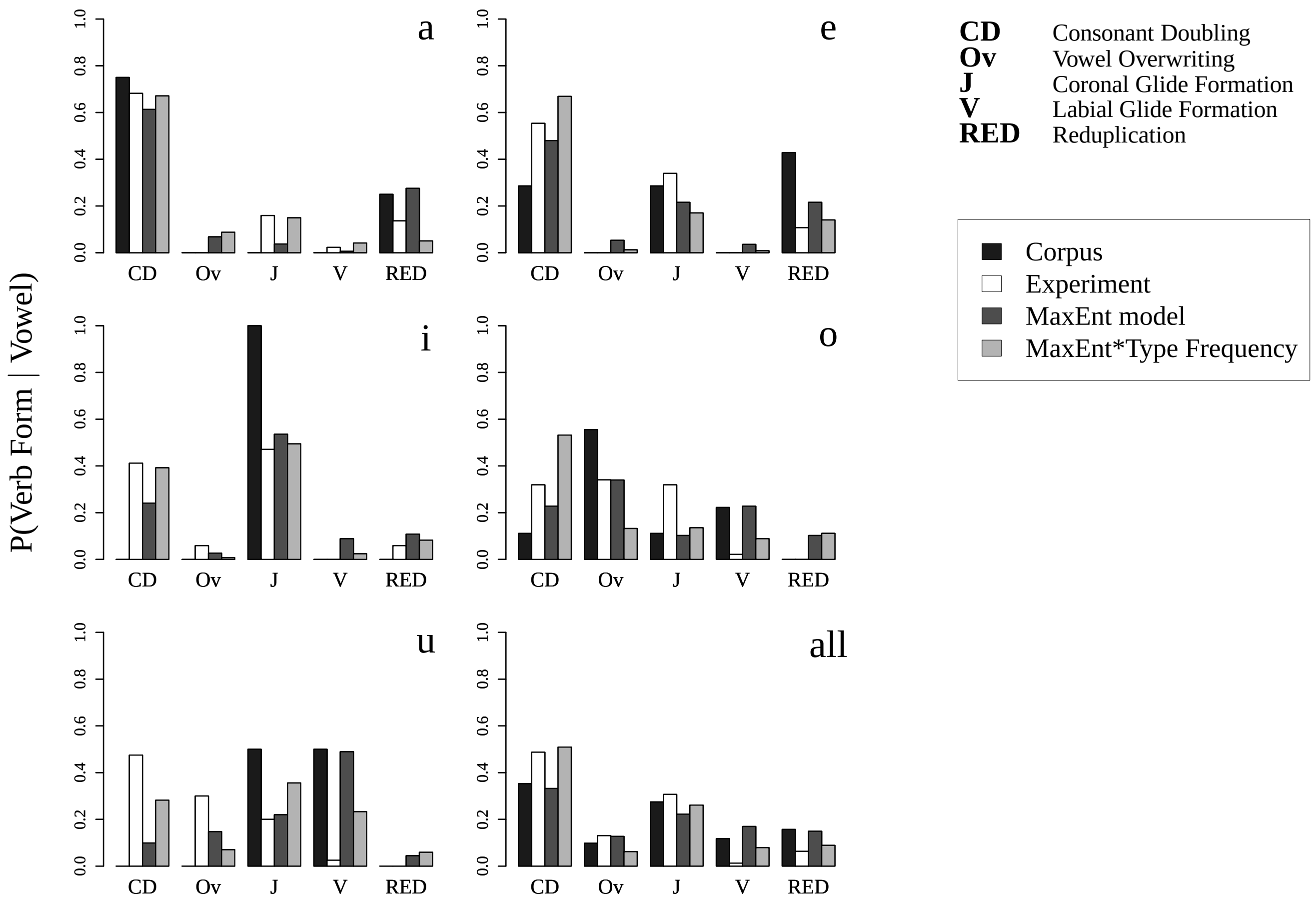
(16) $P(Verb \mid Vowel) = P_{MaxEnt}(Verb \mid Vowel) * P(Verb)$

The probability of the verbal form ($P(Verb)$) is calculated based on existing denominal verbs, as well as other existing verbs which have one of the forms in (4). These verbs were collected from Bolozky's '501 Hebrew Verbs'. Here are some examples:

(17) Denominal verb lookalikes:
a. χibev *love* (Consonant Doubling)
b. ʕoʁeʁ *wake (someone) up* (Vowel Overwriting)
c. t͡sijeʁ *draw* (Coronal Glide Formation)

Applying Bayes' theorem, we get the set of predicted distributions shown in (18).

(18) The final model



The addition of the type frequency term to the model significantly improves its fit.

(19) Chi square value for various models of the experimental data

Model	χ^2	df	χ^2_{diff}	df_{diff}	p
MaxEnt	163.5	8			
MaxEnt + Output freq.	127.7	12	35.8	4	$< .001$

7. Conclusion

7.1 Implications for Semitic morpho-phonology

In this test of the productivity of denominal verb formation, it was found that speakers attend to the noun's vowel very much less than expected given the lexical distributions. The MaxEnt model produced a good match to the experimental data when the weight of MAX-V-STEM, the constraint which forced the preservation of the noun's vowel, was made arbitrarily low.

Although preservation of the noun's vowel is an apparent goal in denominal verbs in the lexicon, speakers do not apparently have this goal when producing novel verbs. This constitutes an argument for the psychological reality of the consonantal root in this case - speakers pay more attention to the consonants of an input than to its vowel.

7.2 Implications for morpho-phonological productivity

This study builds on previous work such as Hayes et al. (2009), Becker et al. (2011), Pierrehumbert (2006) which shows that language learners are able to reproduce probability distributions present in their lexicons on wug-tests. However, speakers are only sensitive to some levels of generalization and not to others. In this study I demonstrate that Hebrew speakers use conditional probability based on generalization over a property of the input form, but also that they are sensitive to output type frequency and use this as well to produce novel words.

Speakers demonstrated underlearning of the conditional probabilities present in the lexicon, and overgeneralization of one form (Consonant Doubling) at the expense of others. The model successfully reproduced this overgeneralization via the strong regularization term and the incorporation of output type frequency. The regularization resulted in overall 'smoothing' of the predicted output distributions relative to the training data. Likewise, overgeneralization of consonant doubling was especially strong, because the constraint militating against it, MAX-V-STEM, was given an arbitrarily low weight. Since Consonant Doubling is the most common form in the lexicon, multiplying the MaxEnt frequencies by the output type frequencies amplified the degree of overgeneralization.

Questions future work must address include: Why did speakers underlearn the conditional probabilities in the lexicon? and also, Why are speakers sensitive to output type frequency in this case, but not (apparently) in others? The answers may be related. It could be, for example, that speakers are sensitive to output type frequencies just in case the input-sensitive mapping frequencies are difficult to learn. The mapping frequencies reported here may be difficult to learn for a few reasons. First, the body of evidence for the pattern in the lexicon is very small - only a total of 52 forms. Second, in the denominal verb formation pattern, speakers have to choose between five possibilities, not just two (as in the Hungarian genitive or Turkish possessive). Kam (2009) discusses statistical learning study in which participants are capable of matching a distribution over two forms, but when give more than two they employ an 'always choose the most common' strategy.

References

Outi Bat-El. Stem modification and cluster transfer in modern hebrew. *Natural Language and Linguistic Theory*, 12:571–596, 1994.

Outi Bat-El. The fate of the consonantal root and the binyan in optimality theory. In Jean Lowenstam, editor, *Grammaire et gabarits*. Presses universitaires de Vincennes, Saint-Denis, France, 2003.

Michael Becker, Nihan Ketrez, and Andrew Nevins. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, 87(1):84–125, March 2011.

Jean Berko. The child's learning of english morphology. *Word*, 14:150–77, 1958.

Shmuel Bolozky and Michael Becker. Living lexicon of hebrew nouns. ms. UMass Amherst, 2006.

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 and Colin Wilson. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39:379–440, 2008.

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. More than words: Adults learn probabilities over categories and relationships between them. *Language learning and development*, 5(2):115–145, April 2009.

John J. McCarthy. *Formal Problems in Semitic Phonology and Morphology*. Garland, New York, 1985.

John J. McCarthy and Alan Prince. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, 18, 1995.

Joe Pater. Weighted constraints in generative linguistics. *Cognitive Science*, 33:999–1035, 2009.

Janet Pierrehumbert. The statistical basis of an unnatural alternation. In Louis Goldstein, D.H. Whalen, and Catherine Best, editors, *Laboratory Phonology 8: Varieties of phonological competence*, pages 81–106. Berlin: Mouton de Gruyter, 2006.

Paul Smolensky and Geraldine Legendre. *The harmonic mind: from neural computation to optimality-theoretic grammar*. MIT Press, Cambridge, Massachusetts, 2006.

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

Adam Ussishkin. Templatic effects as fixed prosody: The verbal system in semitic. In Jean Lowenstam Jacqueline Lecarme and Ur Shlonsky, editors, *Research in Afroasiatic Grammar III*, pages 511–530. John Benjamins, Amsterdam, 2003.

Colin Wilson and Benjamin George. The maxent grammar tool. Department of Cognitive Science, Johns Hopkins University and Department of Linguistics, UCLA (http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/), 2009.

Kie Zuraw. *Patterned Exceptions in Phonology*. PhD thesis, University of California, Los Angeles, 2000.

Department of Linguistics
South College
University of Massachusetts
Amherst, MA 01003
cmooreca@linguist.umass.edu