Surfeit of the stimulus experiments (Moreton, 2008; Hayes et al., 2010; Becker et al., 2011), in which data contain patterns apparent to the analyst, but human learners fail to learn them, shed light on the nature of the human language learning faculty by demonstrating where and how it fails. This paper presents the results of a natural language surfeit of the stimulus experiment, in which speakers undergeneralize some relationships apparent in the lexical patterns, but overgeneralize others.

In Modern Hebrew, verbs must be minimally of the shape CVCVC. Verbs derived from nouns of shape CVC can take five shapes, depending on the quality of the noun's vowel. For a noun $C_1VC_2$:

| | | | |
|---|---|---|---|
| Coronal glide formation: | $C_1ijeC_2$ | (tik ∼ tijek) | V = [i,u] |
| Labial glide formation: | $C_1iveC_2$ | (sug ∼ siveg) | V = [u,o] |
| Consonant doubling: | $C_1iC_2eC_2$ | (dam ∼ dimem) | V = [a] |
| Vowel overwriting: | $C_1VC_2eC_2$ | (kod ∼ koded) | V = [o] |
| Reduplication: | $C_1iC_2C_1eC_2$ | (daf ∼ difdef) | V = [a,e] |

In the lexicon, the noun's vowel predicts what form the verb will take, but there are exceptions.

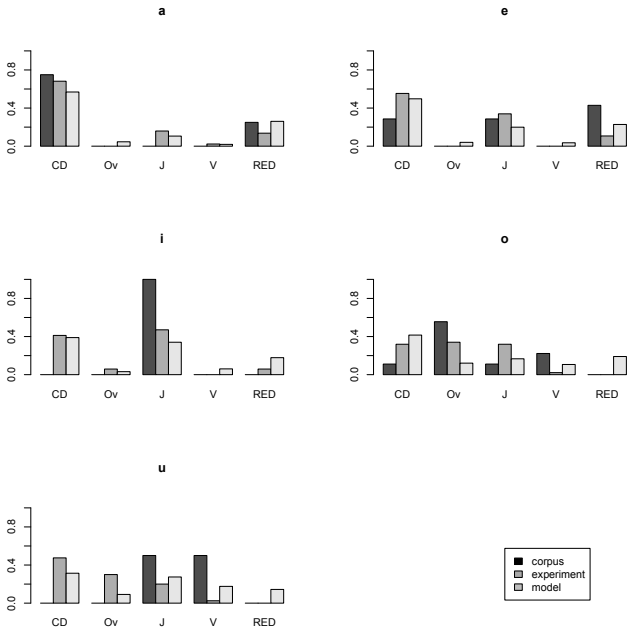
A web-based wug-test (Berko, 1958) was conducted in which 27 native Hebrew speakers produced verbs for 20 novel nouns. Nouns were presented aurally in stories, and participants filled in the blank of a written sentence with a verb. A Poisson regression showed that the noun's vowel predicted the frequency of responses of each verbal type, as in the lexicon, but also that Consonant Doubling was significantly more common in the participants' responses than all other verbal forms, though this is not true in the lexicon.

A graph (1, below) illustrates that although the frequency of verbal form produced in the experiment (the dark grey bars) does depend on the noun's vowel, this dependency is not as robust as in the lexical pairs (the black bars). The main differences can be characterized as 1) Consonant Doubling is overgeneralized. It occurs more often and in more contexts in the production data than in the lexicon, and 2) Glide Formation from high vowels is undergeneralized. The two high vowels take glide formation of some kind exceptionlessly in the lexicon, but only half the time or less in the production data.

I present a Bayesian model of these data, of the type proposed in Wilson and Davidson (2009). The model uses both phonological knowledge, modeled with a Maximum Entropy learning algorithm (Goldwater and Johnson, 2003) with inductive biases, and the type frequency of the various possible output forms. Its predicted probabilities correlate well with the probabilities observed in the experiment (r=.79).

I conclude that participants use both their phonological knowledge, and their knowledge of output type frequency to produce new words.

(1) Probability of verbal form by vowel. CD= Consonant Doubling; Ov= Vowel Overwriting; J= Coronal Glide Formation, V= Labial Glide Formation




# References

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.

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*, 2010.

Elliott Moreton. Analytic bias and phonological typology. *Phonology*, 25:83–127, 2008.

Colin Wilson and Lisa Davidson. Bayesian analysis of non-native cluster production. In Seda Kan, Claire Moore-Cantwell, and Robert Staubs, editors, *Proceedings of the North East Linguistics Society*, Amherst, MA, 2009. North East Linguistics Society, GLSA Publications.