
The naturalness of product-oriented generalizations*

1	Introduction	2
2	A natural lexical trend in Hebrew	4
3	Learning natural trends with markedness constraints	9
3.1	Analysis	10
3.2	Ranking conflicts trigger the formation of generalizations	17
3.3	Learning specific patterns first	21
4	Product-orientedness in an artificial language	25
4.1	Materials	27
4.2	Methods	29
4.3	Participants	32
4.4	Transcription and encoding	32
4.5	Results	34
4.5.1	Generalization differences between the groups	34
4.5.2	No memorization differences between the groups	39
4.5.3	Bias towards [-im]	42
4.5.4	Errors and vowel perception	43
4.5.5	Summary of the experimental results	45
5	Discussion and analysis	45
5.1	The role of Universal Grammar in learning alternations	46
5.2	Stem changes and allomorph selection	51
5.3	The limited role of phonotactics	54
5.4	Learning without product-oriented generalizations	55
5.5	The role of the grammar of real Hebrew	58
5.6	Source-oriented generalizations?	60
6	Conclusions	60

*This work benefitted greatly from the generous help of Ram Frost (Hebrew University). For their careful and thorough reading of the manuscript in various stages, we are grateful to Adam Albright, Lyn Frazier, John McCarthy, Andrew Nevins, Joe Pater, and Matt Wolf. We also got thought-provoking and insightful questions from Outi Bat-El, Iris Berent, Marc Ettliger, Kathryn Flack, Bruce Hayes, and Kie Zuraw. We thank the audiences at the NYU linguistics department, the UCLA linguistics department, and the 83rd LSA annual meeting in San Francisco, especially Chris Barker, Roy Becker, Susan Curtiss, Lisa Davidson, Adamantios Gafos, Marc Garellek, Maria Gouskova, Gregory Guy, Susie Levi, Alec Marantz, Kevin Ryan, Jason Shaw, and Carson Schütze. Errors, if any, would remain our own.

1 Introduction

In Hebrew, the plural suffix for nouns has two allomorphs: [-im] for masculine nouns and [-ot] for feminine nouns. The choice of affix is completely predictable for adjectives and loanwords, but native nouns allow exceptions both ways: some masculine nouns take [-ot], and some feminine nouns take [-im].

The masculine nouns that exceptionally take [-ot] are phonologically clustered. Out of the 230 *ot*-takers in a Hebrew lexicon (Boložky & Becker 2006), 146 nouns, or 63%, have the vowel [o] in their last syllable. The results in Berent, Pinker & Shimron (1999, 2002) and in Becker (2009) show that speakers are aware of the trend for more [-ot] in nouns that end in [o], and project this trend onto novel items. In other words, speakers' choice of plural allomorph is not determined entirely by the stem's gender or morphologically idiosyncratic properties, but also by the stem's phonological shape.

In our analysis of this case of partially phonologically determined allomorph selection, *ot*-takers with [o] in them respond to a high-ranking markedness constraint that requires an unstressed [o] to be licensed by an adjacent stressed [o] (cf. similar requirement on vowel licensing in Shona, Beckman 1997; Hayes & Wilson 2008). Markedness-based accounts of allomorph selection in OT are common in the literature, starting with Mester (1994) and continuing with Mascaró (1996), Kager (1996), Anttila (1997), and Hargus (1997), among many others. More recent work includes Paster (2006), Wolf (2008), and Trommer (2008). Since the analysis crucially relies on the use of markedness constraints, i.e. constraints that assess output forms, regardless of the posited underlying representation, we set out to empirically test the adequacy of accounting for lexical trends using markedness constraints.

At issue is whether generalizations are source-oriented or product-oriented (Bybee & Slobin 1982; Bybee & Moder 1983; Bybee 2001; Albright & Hayes 2003). In the Hebrew case, one can state the correlation between a stem [o] and [-ot] in a source-oriented way, i.e. in terms of a relationship between singular and plural forms, saying that nouns that have [o] in the singular are more likely to take [-ot] in the

plural. Alternatively, one can state the generalization in a product-oriented way, i.e. in terms of conditions on the plural forms only, saying that in the plural, noun stems that have [o] in them are more likely to show up with the suffix [-ot]. In Optimality Theory, generalizations that are stated in terms of markedness constraints are product-oriented, since markedness constraints only assess outputs, or products of derivations. In contrast, rule-based theories express generalizations in terms of mappings between inputs and outputs, i.e. generalizations depend on the input to the derivation, so they are source-oriented.

The source-oriented and product-oriented generalizations are almost exactly equivalent when stated over the attested lexicon of Hebrew, since each and every noun that has an [o] in the final syllable of its plural stem also has an [o] in the singular,¹ and with the exception of five nouns,² every noun that has an [o] in its final syllable in the singular also has an [o] in the final syllable of the plural stem.

We propose that evidence in favor of product-oriented knowledge of lexical trends can be adduced by the behavior of Hebrew speakers in an artificial language setting. We present such an experiment, where speakers were taught a language that is just like Hebrew, but with two additional vowel-change rules that caused [o]'s to be present only in the singular stem or only in the plural stem, but not in both. Speakers preferred to associate the selection of [-ot] with nouns that have [o] in the plural stem rather than in the singular stem, showing that they were using surface-based, or product-oriented methods for selecting the plural allomorph.

While the experimental results are in principle compatible with a range of imaginable product-oriented interpretations, we offer our Optimality Theoretic analysis not simply for the sake of concreteness, but because it offers the best account of the results. Furthermore, our grammar-based analysis affords a clear understanding of the three-way relationship between real Hebrew, the artificial language experiment, and the known facts about vowel licensing cross-linguistically.

¹For nouns with the vowel pattern [o-e] in the singular, vowel deletion makes the [o] stem-final in the plural, e.g. *fomér* ~ *fomr-ím* 'guard, keeper'.

²Three nouns change the singular [o] to [u] (*xók* ~ *xukím* 'law', *tóf* ~ *tupím* 'drum' and *dóv* ~ *dubím* 'bear'), and two nouns change the singular [o] to [a] (*róf* ~ *rafím* 'head', *jóm* ~ *jamím* 'day').

This paper is organized as follows: §2 presents the distribution of the plural allomorphs in the lexicon, and summarizes previous work on the productivity of this distribution. §3 offers an analysis of the lexical trends in terms of markedness constraints. Support for this analysis is presented in §4, with results of an artificial language experiment that shows speakers’ preference for product-oriented generalizations. The results are discussed and analyzed in §5, and §6 concludes.

2 A natural lexical trend in Hebrew

Hebrew has two plural markers: [-im] and [-ot]. When nouns that refer to animates have both an *im*-plural and an *ot*-plural, they invariably correspond to natural gender, as in the word for horse/mare in (1).³ At the phrase level, gender agreement on adjectives and verbs is also invariably regular.

- | | | | | | |
|-----|----|----------|-----------|---------|-----------------------------|
| (1) | a. | sus-ím | ktan-ím | rats-ím | |
| | | horse-PL | little-PL | run-PL | ‘little horses are running’ |
| | b. | sus-ót | ktan-ót | rats-ót | |
| | | mare-PL | little-PL | run-PL | ‘little mares are running’ |

At the word level, native nouns can take a mismatching suffix: (2a) shows that the masculine noun *xalón* ‘window’ exceptionally takes [-ot] at the word level, but the accompanying adjective and verb take [-im], revealing the true gender of the noun (Aronoff 1994). The opposite is seen with the feminine noun *nemalá* ‘ant’ in (2b).

- | | | | | | |
|-----|----|-----------|---------|------------|---------------------------|
| (2) | a. | xalon-ót | gdol-ím | niftax-ím | |
| | | window-PL | big-PL | opening-PL | ‘big windows are opening’ |

³When nouns that refer to animates only have one plural form, the plural affix does not necessarily conform to natural gender. For example, the native noun *fulij-á* ~ *fulij-ót* ‘apprentice’ can apply to either males or females. The word for ‘baby’ has gender marking in the singular (masculine *tinók* vs. feminine *tinók-et*), but the plural is *tinok-ót* for male or female babies. Unsurprisingly, children often use the form *tinok-ím* to refer to male babies.

b.	nemal-ím	ktan-ót	nixnas-ót	
	ant-PL	small-PL	entering-PL	‘small ants are coming in’

In the loanword phonology, the plural suffix selection is completely regular even at the word level: If the right edge of the singular noun is recognizable as a feminine suffix, as in *fukátj-a*, [-ot] is selected (3a), otherwise it’s [-im], as in *blóg-im* (3b). This even applies to nouns that refer to male humans, like *kolég-a* (3c). Loanwords that refer to female humans but don’t have a plausible feminine suffix on them, like *madám*, mostly resist pluralization (3d).⁴

(3)	a.	fukátj-a	*fukátj-im	fukátj-ot	‘focaccia’
	b.	blóg	blóg-im	*blóg-ot	‘blog’
	c.	kolég-a	*kolég-im	kolég-ot	‘(male) colleague’
	d.	madám	?madám-im	*madám-ot	‘madam (in a brothel)’

A final factor that affects the distribution of the plural allomorphs is phonological. Masculine native nouns show a clustering of the *ot*-takers: about two-thirds of the masculine nouns that exceptionally take [-ot] have [o] in their final syllable (Glinert 1989; p. 454, Aronoff 1994; p. 76), and about a third of the nouns that have [o] in their final syllable take [-ot] (see 8-9 below). This preference for [-ot] in masculine nouns that end in [o] applies productively to novel nouns, as seen in Berent, Pinker & Shimron (1999, 2002) and in Becker (2009). The feminine *im*-takers don’t seem to pattern phonologically in any noticeable way, and will not be discussed further.

To summarize so far, there are three factors that determine plural allomorph selection without exception:

- (4) a. Natural gender: Whenever a noun stem has an *im*-plural and an *ot*-plural, the *im*-plural refers to males and the *ot*-plural refers to females.

⁴Some speakers offer *madám-ij-ot* as the plural of *madám*, i.e. they add the feminine suffix [-it] to the root to make a more plausible singular feminine stem for the plural [-ot] to attach to. The change of [-it] to [-ij] before [-ot] is regular in the language (Bat-El 2008).

- b. Morpho-syntactic gender: Adjectives and verbs take [-im] with masculine nouns and [-ot] with feminine nouns. Essentially, adjectives and verbs reveal the true gender of a noun.
- c. Morpho-phonological gender: When a loanword (i.e. a noun that keeps the stress on its stem in the plural) ends in what sounds like a feminine suffix, it will be an *ot*-taker, otherwise it will be an *im*-taker.

And there are two factors that have some power in predicting the plural allomorph selection, but these allow exceptions:

- (5) a. Morpho-syntactic gender: Native nouns (i.e. nouns that lose their stress to the plural affix in the plural) are usually *im*-takers if masculine and *ot*-takers if feminine.
- b. Phonology: The majority of native masculine *ot*-taking nouns have an [o] in their stem.

From this point on, the focus will be on native masculine nouns, and the phonological effect of a stem [o] on the selection of the plural affix. The presence of a stem [o] makes the selection of [-ot] more likely, relative to the selection of [-ot] in the absence of a stem [o].

The partial predictability in the distribution of *ot*-takers is not incompatible with the existence of minimal pairs, such as those in (6), where the choice of plural affix disambiguates the meaning. Overall in the lexicon, [-ot] is more likely with a stem [o], but for any single lexical item, the selection of an affix is unpredictable.

- (6) a. himnon-ím / himnon-ót ‘national anthem’ / ‘religious hymn’
- b. tor-ím / tor-ót ‘line, queue’, ‘appointment’ / ‘turn’
- c. maamad-ím / maamad-ót ‘stand’ / ‘status’

With certain nouns, the choice of plural suffix is variable in and between speakers. Some nouns that occur variably in current usage are in (7), where the percentage

indicates the proportion of [-ot] plurals out of the total plural forms found in Google.⁵

- (7)
- | | | | |
|--|--|-----|-------------|
| | a. fofar-ím / fofar-ót | 56% | ‘shofar’ |
| | b. djokan-ím / djokna-ót <i>or</i> djokan-ót | 41% | ‘portrait’ |
| | c. kilfon-ím / kilfon-ót | 11% | ‘pitchfork’ |

For the purposes of this study, data about the distribution of [-im] and [-ot] comes from an electronic lexicon of Hebrew (Bolzky & Becker 2006) that was modeled after TELL (a Turkish Electronic Living Lexicon, Inkelas et al. 2000). The lexicon lists nouns and their plurals. The nouns are mostly collected from the Even-Shoshan (1966) dictionary, and their plurals reflect the knowledge of the second author, occasionally augmented by Google searches, in an attempt to approximate an idealized native speaker. The table in (8) lists the native masculine nouns in the lexicon, arranged by the vowel in their final syllable. Recall that in this context, ‘native’ refers to unaccented nouns (Bat-El 1993; Becker 2003), i.e. nouns that surface in the plural with the stress on the plural suffix, as opposed to loanwords, which always surface with stress on the root.

(8)

	Final vowel	<i>n</i>	<i>ot</i> -takers	% <i>ot</i> -takers
	u	1101	6	1%
	i	464	8	2%
	a	1349	39	3%
	e	977	31	3%
	o	523	146	28%
	Total	4414	230	5%

The data in (8) shows that *ot*-taking accounts for a fairly meager proportion (2.2%) of the native nouns that end in vowels other than [o], but almost a third of the nouns

⁵There are surely many more nouns that variably take either plural affix, but Hebrew orthography makes searching for them online a difficult task. The variable choice of the plural affix goes back to Tiberian Hebrew, where a considerable number of nouns are attested with two plural forms (Aharoni 2007), e.g. do:r-í:m (Isaiah 51, verse 8) vs. do:r-ó:θ (Isaiah 41, verse 4) ‘generations’.

that end in [o]. The 146 *ot*-takers that end in [o] account for 63% of the 230 *ot*-takers.

There are further morpho-phonological regularities that correlate with *ot*-taking within the set of nouns that have [o] in their final syllable. For instance, *ot*-taking is completely regular for a class of tri-syllabic masculine nouns that have a stem of the shape [CiCaC-] and the suffix [-on] (e.g. *fikar-ón* ‘state of drunkenness’). These nouns can be productively formed from verbs, to mean ‘state of X-ness’, and with this meaning, their plural is always in [-ot].⁶ Tri-syllabic nouns in [-on] account for 54 of the 146 [o]-final *ot*-takers in (8). Of the remaining 92 [o]-final *ot*-takers, 49 end in the segments [on], but in many cases, it is hard to determine whether these segments belong to the an affix⁷ or to a stem.

Having an [o] in the root is well correlated with taking [-ot] in the plural even after allowing for the effect of the suffix [-on]. In the lexicon, this can be seen most clearly with monosyllables, tabulated in (9): Of the 70 monosyllables with [o] in them, 20 are *ot*-takers (29%), and none of these *ot*-takers end in [n]. This rate of *ot*-taking is comparable to the overall rate of *ot*-taking. In addition, note that the experimental items in §4 have no final sonorants at all, to avoid any [-on] effects.

⁶The etymological data in Bolozky & Becker (2006) confirms the modern productivity of *ot*-taking for [CiCaC-on] nouns. Of the 230 *ot*-takers, 216 are attested before modern Hebrew (i.e. Biblical or Mishnaic). Of the remaining 14 *ot*-takers that were created in modern times, 13 are [CiCaC-on] nouns. The remaining modern item, prescriptively *dúax* ~ *dux-ót* ‘report’, is colloquially pronounced *dóx* ~ *dox-ót*, thus making every single modern *ot*-taker a noun with [o] in its stem.

⁷In addition to the [-on] that appears on [CiCaC-] stems to mean ‘state of’, Hebrew has at least two more [-on]’s: One [-on] is a diminutive, which always takes [-im] (e.g. *sus-ón* ~ *sus-on-ím* ‘little horse’), and another [-on] is rather vacuous semantically, and doesn’t require a certain plural affix (e.g. *fabat-ón* ~ *fabat-on-ím* ‘sabbatical’).

(9)

Stem vowel	<i>n</i>	<i>ot</i> -takers	% <i>ot</i> -takers
i	89	2	2%
a	113	5	4%
u	53	4	8%
e	39	5	13%
o	70	20	29%
Total	364	36	9%

Finally, the ability of a stem [o] to favor the selection of [-ot] is extended by speakers to novel items. This effect was observed indirectly in Berent, Pinker & Shimron (1999, 2002),⁸ while Becker (2009) explored the effect directly, and found that novel words of the form CaCoC were 24% more likely to get *ot*-responses than novel words of the form CaCaC. In other words, the correlation between having a root [o] and *ot*-taking is not confined to the lexicon, but rather influences speakers' choices when they productively derive novel forms.

3 Learning natural trends with markedness constraints

The lexicon study in §2 and the experimental results in Berent et al. (1999, 2002) and Becker (2009) show that having [o] in the root correlates with choosing the plural suffix [-ot]. This section offers an analysis of this correlation in terms of markedness constraints. The appropriateness of using markedness constraints will be motivated

⁸Experiment 1 in Berent et al. (1999) looked for the effect that the similarity of novel items to existing items has on the novel items' propensity to take [-ot]. The list of items similar to existing *ot*-takers, which was dominated by items with [o] in them, elicited up to 60% *ot*-responses, while the list of items similar to existing *im*-takers, which only had a minority of items with [o], elicited less than 10% *ot*-responses. Interestingly, in experiment 2 of Berent et al. (2002), where the items were balanced for their vowels, the difference in [-ot] responses in the "dissimilar" condition was almost completely gone (28% *ot*-responses for novel items dissimilar to real *im*-takers vs. 33.7% *ot*-responses for novel items dissimilar to real *ot*-takers). This suggests that most of the difference observed in Berent et al. (1999) was due to the effect of [o].

theoretically in this section; we offer experimental motivation for our approach in §4, using results from an artificial language experiment. The analysis in this section is extended to the experimental results in §5.

The analysis is based on Optimality Theory (Prince & Smolensky 1993/2004) with the Recursive Constraint Demotion algorithm (RCD, Tesar & Smolensky 1998, 2000; Tesar 1998; Prince 2002), augmented with a mechanism of constraint cloning (Pater 2006, 2008; Coetzee 2008; Mahanta 2007; Becker et al. 2008; Becker 2009; Tucker 2009). Cloning allows the speaker to keep track of lexical trends and build their relative strength into the grammar, as explained below.

3.1 Analysis

The preference of roots that have [o] for taking [-ot] is interpreted here as a requirement for licensing unstressed [o]'s. In native nouns, stress shows up on the root in unsuffixed forms (e.g. *xalón* 'window'), but stress moves to the right in suffixed forms, such as the plural (e.g. *xalon-ót* 'windows'). In the plural, then, the root's [o] surfaces unstressed, where it requires licensing.

Limiting [o] (and other non-high round vowels) to prominent positions is quite common in the world languages. Many languages are known to limit [o] to the stressed syllable, as in Russian *dóm* ~ *dám-á* 'home.NOM.(PL)'. Similar restrictions apply in Portuguese and elsewhere (see Crosswhite 2001 and references within).⁹

Other languages require [o] to be licensed by the word-initial syllable (Beckman 1997, 1998; Barnes 2006). Turkish native nouns, for instance, allow [o] only in the first syllable of the word. Shona allows [o] in the word-initial syllable, and more interestingly, an initial [o] can license an [o] later in the word (see also Hayes & Wilson 2008).

⁹In standard American English, and other dialects, [o] can be unstressed ('piano', 'fellow') word-finally, but in some dialects, especially in the South, unstressed [o] is not allowed ('piana', 'fella'). This restriction on [o] in English, however, is just a part of a wider ban on unstressed full vowels in these dialects.

In the analysis proposed here, Hebrew is like Shona, but with stress: In Hebrew, [o] must be stressed, but a stressed [o] allows [o] to appear elsewhere in the word. The licensing of [o] is not a categorical restriction in Hebrew, as unstressed [o]’s are tolerated. The licensing effect emerges when selecting [-ot] allows the stressed [o] of the suffix to license the unstressed [o] in a stem.

Regular masculine nouns, as in (10a), allow [o] to surface unlicensed in the plural. But nouns with [o] in them can have that [o] licensed in the plural if they are *ot*-takers, as in (10b). The representations in (10) are auto-segmental, but other theories that regulate the relations between vowels would work equally well, either with domains or spans (as in, e.g. Cassimjee & Kisseberth 1998 or McCarthy 2004) or by correspondence (Walker 2006).

(10)		Singular	Plural	
		$\begin{bmatrix} -\text{high} \\ +\text{back} \end{bmatrix}$ 	$\begin{bmatrix} -\text{high} \\ +\text{back} \end{bmatrix}$ $\begin{bmatrix} +\text{high} \\ -\text{back} \end{bmatrix}$ 	
a.	Regular	a l ó n	a l o n - í m	‘oak tree’
		$\begin{bmatrix} -\text{high} \\ +\text{back} \end{bmatrix}$ 	$\begin{bmatrix} -\text{high} \\ +\text{back} \end{bmatrix}$ /	
b.	Irregular	x a l ó n	x a l o n - ó t	‘window’

Among nouns that have [o] in their roots, only those that surface stressless in the plural, i.e. native nouns, could benefit from taking [-ot] in the plural. Loanwords, i.e. nouns that keep their stress on the root, would not benefit from taking [-ot], since there is no [o] that needs licensing, and indeed loanwords do not allow exceptional *ot*-taking.

In terms of Optimality Theory (Prince & Smolensky 1993/2004), taking [-im] or [-ot] can be fruitfully understood as responding to the satisfaction of different markedness constraints.

The requirement for the masculine [-im] on masculine nouns is enforced by a

morphological constraint, ϕ -MATCH, which demands gender features to match in poly-morphemic words (cf. Wolf 2008 §2.4.2 for a more complete view of the role of morphology in the Hebrew plural). For an *im*-taker like *alón* (11), ϕ -MATCH outranks the constraint LICENSE(o), which requires [o]-licensing (defined in 13 below):

(11)

$\text{alon}_{\text{MASC}} + \{\text{im}_{\text{MASC}}, \text{ot}_{\text{FEM}}\}$	ϕ -MATCH	LICENSE(o)
a. alón-ím		*
b. alon-ót	*!	

Conversely, an *ot*-taker like *xalón* requires a high-ranking LICENSE(o):

(12)

$\text{xalon}_{\text{MASC}} + \{\text{im}_{\text{MASC}}, \text{ot}_{\text{FEM}}\}$	LICENSE(o)	ϕ -MATCH
a. xalón-ím	*!	
b. xalon-ót		*

(13) LICENSE(o)

An [o] must be licensed by virtue of being stressed, or by virtue of being auto-segmentally associated to a stressed [o] in an adjacent syllable.

In a small number of nouns, LICENSE(o) forces the change of a root [o] to [u], as in (14a). Ranking ϕ -MATCH and LICENSE(o) over IDENT(Hi) would give rise to the vowel alternation, as shown in (15). The number of words involved, however, is very small: It's the nouns *xok* 'law', *tof* 'drum' and *dov* 'bear', the quantifiers *kol* 'all' and *rov* 'most', and a dozen adjectives. There are only two words that display an *o* ~ *a* alternation: *róf* 'head' and *jóm* 'day' (14b).

(14)

a.	<i>xók</i>	<i>xuk-ím</i>	'law'
b.	<i>róf</i>	<i>raf-ím</i>	'head'

(15)

$xok_{MASC} + \{im_{MASC}, ot_{FEM}\}$	ϕ -MATCH	LICENSE(o)	IDENT(Hi)
a. xok_{MASC} -ím			*
b. xok_{MASC} -ím		*!	
c. xok_{MASC} -ót	*!		

An additional effect that follows from the use of constraints that license [o] by the stressed syllable is the regularity of the plural affix selection in loanwords. In these words, stress stays on the root,¹⁰ so any [o] in the stem would be equally licensed in the singular and the plural. The tableau in (16) shows the noun *blóg* ‘blog’, where the presence of the [o] cannot trigger selection of [-ot], since LICENSE(o) is equally satisfied by either plural affix.

(16)

$blóg_{MASC} + \{im_{MASC}, ot_{FEM}\}$	ϕ -MATCH	LICENSE(o)
a. $blóg_{MASC}$ -im		
b. $blóg_{MASC}$ -ot	*!	

Similarly, if a loanword has an unstressed [o] in it, like *kétfop* ‘ketchup’, LICENSE(o) is again unable to prefer the selection of [-ot].

¹⁰If suffixation puts the stressed syllable more than three syllables away from the edge, the stress (optionally) shifts two syllables (=one trochee) to the right (Bat-El 1993; Becker 2003), but never off the root. For example, the plural of *béjbisiter* ‘male babysitter’ is either *béjbisiter-im* or *bejbisiter-im*, but never **bejbisiter-im*. Similarly, the plural of *béjbisiter-it* ‘female babysitter’ is either *béjbisiter-ij-ot* or *bejbisiter-ij-ot*, but never **bejbisiter-ij-ot*.

(17)

kétfop _{MASC} + {im _{MASC} , ot _{FEM} }	ϕ-MATCH	LICENSE(o)
a. kétfop-im		*
b. kétfop-ot	*!	**

The regular selection of [-ot] with feminine loanwords, as in *fukátfa* ~ *fukátf-ot* ‘focaccia’, does indeed introduce an unlicensed [o]. Since our analysis allows LICENSE(o) to dominate ϕ-MATCH for some nouns, one would expect that some feminine loanwords would choose [-im], contrary to fact. However, recall that the selection of [-ot] in loanwords is not based on morpho-syntactic gender (i.e. the gender that is revealed by agreement on adjectives and verbs), but rather on apparent morpho-phonological gender: All and only the nouns that *appear* to be feminine by virtue of sounding like they have a feminine suffix on them take [-ot], including masculine nouns that end in [-a], such as *kolég-a* ‘(male) colleague’. We are assuming that just like ϕ-MATCH enforces morpho-syntactic gender agreement, some other constraint enforces morpho-phonological gender agreement, and this constraint categorically outranks both LICENSE(o) and ϕ-MATCH. We call this constraint μ-MATCH, as shown in (18).¹¹

(18)

fukátf-a _{FEM} + {im _{MASC} , ot _{FEM} }	μ-MATCH	ϕ-MATCH	LICENSE(o)
a. fukátf-im	*!	*	
b. fukátf-ot			*

¹¹We leave the formal definition of μ-MATCH to future work; in essence, μ-MATCH is a constraint about the morphological parsing of words. We suggest that a final unstressed [a] is simply less plausible as a part of the root, and therefore more strongly requires interpretation as a suffix, possibly in concert with pressure from FINAL-C, which demands that roots be consonant-final. A different possibility is suggested by Faust (2008), where the plural suffix [-ot] phonologically contains the singular [-a], thus involving faithfulness constraints in the selection of the plural suffix.

Returning to native masculine nouns now, there is still the problem of selecting [-ot] for those *ot*-takers that don't have [o] in them, such as *kír* ~ *kír-ót* 'wall'. Since LICENSE(o) cannot help with selecting [-ot] in the absence of a root [o], some other mechanism must be involved.

We propose that *ot*-taking can be attributed to a constraint that doesn't refer to the root vowel, but rather penalizes some aspect of the [-im] suffix itself, e.g. * \acute{o} /HIGH, which penalizes stressed high vowels (Kenstowicz 1997; de Lacy 2004). A constraint such as *LAB would work equally well – neither constraint is otherwise clearly active in the language.¹²

(19)

$\text{kír}_{\text{MASC}} + \{\text{im}_{\text{MASC}}, \text{ot}_{\text{FEM}}\}$	* \acute{o} /HIGH	ϕ -MATCH	LICENSE(o)
a. kír-ím	*!		
b. kír-ót		*	

Regular nouns without [o] in them, like *gír* ~ *gír-ím* 'chalk', are derived as in (20).

(20)

$\text{gír}_{\text{MASC}} + \{\text{im}_{\text{MASC}}, \text{ot}_{\text{FEM}}\}$	ϕ -MATCH	* \acute{o} /HIGH	LICENSE(o)
a. gír-ím		*	
b. gír-ót	*!		

This use of * \acute{o} /HIGH, which attributes the selection of [-ot] to marked structure that happens to appear in the suffix [-im], makes no reference to the phonological shape

¹²Arguably, both constraints are relevant for Hebrew phonology in general: * \acute{o} /HIGH could be used to derive the distribution of stressed vowels in segholates, where only non-high stressed vowels are allowed, producing alternations like the one in *kétsev* ~ *kitsb-í* 'rhythm/rhythmic'. Self-conjunction of *LAB could account for the restrictions on the distribution of labials in roots (as in, e.g. de Lacy 1997), although this isn't how root co-occurrence restrictions are handled in Optimality Theory more recently; see Coetzee & Pater (2008) and references within.

of the root. This is in line with the rest of the analysis, which assumes that any vowel other than [o] is inert with respect to plural allomorph selection.

In principle, the selection of [-ot] with nouns that don't have [o] in them could be done with a purely arbitrary diacritic, with no phonological substance at all. In the analysis proposed in (19) above, however, it is hard to see why the learner would fail to notice the preference that * $\acute{\sigma}$ /HIGH makes, if this constraint is indeed universal and available to the learner “for free”.

We leave open the possibility that in some cases, learners are left with no phonological mechanism for making the right choice in allomorph selection, and they are forced to simply list the exceptional affix-takers. Suppose that a constraint such as * $\acute{\sigma}$ /HIGH is unavailable to the speaker for some reason, making the observed form *kir-ót* harmonically bounded, as in (21).

(21)

$\text{kir}_{\text{MASC}} + \{\text{im}_{\text{MASC}}, \text{ot}_{\text{FEM}}\}$	ϕ -MATCH	LICENSE(o)
a. kir-ím		
b. ☹ kir-ót	*!	

Faced with a situation as in (21), the speaker will simply list the form *kir-ót* in their lexicon, i.e. the learner will resort to listing to bypass the OT grammar, as in Zuraw (2000); Hayes & Londe (2006); Hayes et al. (to appear), see also Tessier (2008). Once listed in the lexicon, this form will have no effect on the grammar and thus no effect on the treatment of novel nouns.

To summarize the point so far: Most masculine native nouns in Hebrew select the plural [-im] due to a high ranking morphological constraint, ϕ -MATCH. A phonological constraint, LICENSE(o) prefers the selection of [-ot] when there is an [o] in the stem. Different Hebrew nouns are subject to different constraint rankings: Nouns that take [-im] are associated with a high-ranking ϕ -MATCH, while nouns with [o] in them that take [-ot] are associated with a high-ranking LICENSE(o). Finally,

ot-takers that don't have [o] in them are associated with a different high-ranking phonological constraint, * \acute{o} /HIGH.

3.2 Ranking conflicts trigger the formation of generalizations

The analysis must now be completed with a mechanism that allows speakers to do three things: (a) learn the correct affix to choose with existing nouns, (b) learn the relative frequency of *ot*-taking in the lexicon in the presence and in the absence of a stem [o], and (c) project the frequencies of the lexicon onto novel items. Such a mechanism is outlined here; the full proposal is detailed in Becker (2009).

The analysis relies on learners' ability to identify cases where there is no single grammar that can apply successfully to all of the words of their language. The Recursive Constraint Demotion algorithm (RCD, Tesar & Smolensky 1998, 2000; Tesar 1998; Prince 2002) allows language learners to collect ranking arguments from different lexical items and find conflicting rankings.

The use of RCD is most clearly illustrated with comparative tableaux (Prince 2002), where pairs of winners and losers are compared as to how they fare on various constraints. For example, the plural form of *xalón* 'window' is *xalon-ót*, so the learner has to make sure that *xalon-ót* wins over the intended loser *xalon-ím*. The constraint ϕ -MATCH prefers *xalon-ím*, while the constraint LICENSE(o) prefers *xalon-ót*, so if *xalon-ót* is to win, the constraint that prefers the winner must be ranked over the constraint that prefers the loser. This situation is shown with the winner-loser pair in (22a), with LICENSE(o) assigning a W ("Winner preferring") to the winner-loser pair, and ϕ -MATCH assigning an L ("Loser preferring") to it.

Similarly, the winner-loser pair in (22b) shows the *im*-taker *alón* 'oak tree', which requires the ranking of ϕ -MATCH over LICENSE(o) to make *alon-ím* win over *alon-ót*.

(22)

	LICENSE(o)	ϕ -MATCH
a. <i>xalón-ót</i> \succ * <i>xalón-ím</i>	W	L
b. <i>alón-ím</i> \succ * <i>alón-ót</i>	L	W

Given a comparative tableau, the learner can extract a constraint ranking from it by finding columns that have only W's or empty cells in them, and *installing* the constraints in those columns. Installing a constraint means that it is added to the constraint ranking below any constraints that are already in it, and any winner-loser pairs it assigns a W to are removed from the tableau. Installing constraints continues until all winner-loser pairs are removed. In the case of (22), however, there are no constraints to install, since all the columns have both W's and L's in them.

The solution to this situation was offered by Pater (2006, 2008), who suggested that a constraint can be *cloned* to solve the inconsistent ranking of the constraints. Cloning a constraint means that the learner makes two copies, or clones, of the constraint, and makes both clones lexically-specific. Clones are lexically-specific in the sense that they apply only to the list of lexical items that are associated with them. When a constraint is cloned, every lexical item it assigns a W to is associated with one clone, and every lexical item it assigns an L to is associated with the other clone.¹³

In the case at hand, suppose the learner decided to chose LICENSE(o) for cloning (see §3.3 below about the choice of constraint to clone). One clone would be associated with *xalón*, and the other would be associated with *alón* (23).

¹³This last point is a departure from Pater (2006, 2008). Making both clones of a constraint lexically-specific allows the grammar to be frequency-matching; see below for discussion.

(23)

	LICENSE(o) _{xalon}	ϕ-MATCH	LICENSE(o) _{alon}
a. xalon-ót \succ *xalon-ím	W	L	
b. alon-ím \succ *alon-ót		W	L

Now there is a column that only has W's in it, and there is a constraint to install: LICENSE(o)_{xalon}. Once installed, the *xalon-ót* \succ **xalon-ím* winner-loser pair in (23) is removed, which leaves the column of ϕ-MATCH with only W's in it. ϕ-MATCH is installed and added to the constraint ranking below LICENSE(o)_{xalon}, and the second and last winner-loser pair in (23) is removed. The remaining constraint, LICENSE(o)_{alon} is added to the ranking below ϕ-MATCH.

The result is the grammar in (24), where there are no longer any ranking conflicts.

(24) LICENSE(o)_{xalon} \gg ϕ-MATCH \gg LICENSE(o)_{alon}

As the learner encounters more nouns with [o] in their final syllable, the conflict between ϕ-MATCH and LICENSE(o) will cause more nouns to be associated with one of the clones of LICENSE(o). Nouns that take [-ot] will be associated with the higher ranking clone, and nouns that take [-im] will be associated with the lower ranking clone.

(25) LICENSE(o)_{xalon, makom, ...} \gg ϕ-MATCH \gg LICENSE(o)_{alon, faon, pagof, ...}

Since nouns like *kír*, which don't have [o] in them, are neither preferred nor dispreferred by LICENSE(o), they will not be assigned a W or an L by LICENSE(o), and thus will not be associated with either clone.

Of the nouns with [o] in their final syllable in Bolozky & Becker (2006), 146 are *ot*-takers and 377 are *im*-takers. A speaker who learns all of them will end up with a grammar such as the one in (26).¹⁴

¹⁴This picture is somewhat simplified, since the set of *ot*-takers with a final [o] is not homogeneous,

(26) LICENSE(o)_{146 items} >> ϕ -MATCH >> LICENSE(o)_{377 items}

The grammar in (26) achieves two goals at once: It encodes the behavior of the existing nouns of Hebrew by associating them with one of the clones of LICENSE(o), and since it has a list of *im*-takers and a list of *ot*-taker, the grammar makes available to the speaker the relative proportions of *im*-takers and *ot*-takers among the the nouns that have [o] in them. This information, in turn, can be used to project these proportions onto novel nouns, or in other words, the grammar allows the speaker to be a frequency-matcher (Ernestus & Baayen 2003; Hayes et al. to appear)

Once LICENSE(o) is cloned, and each clone is made lexically-specific, there is no longer a general LICENSE(o) constraint that can apply to novel items. When faced with a novel noun that has [o] in its final syllable, the speaker must decide which clone of LICENSE(o) to associate it with, and this decision will be influenced by the number of items associated with each clone. Since 28% of the nouns associated with clones of LICENSE(o) in (26) are associated with its higher ranking clone, the learner will have a 28% chance of choosing [-ot].

(27) LICENSE(o)_{28%} >> ϕ -MATCH >> LICENSE(o)_{72%}

There is another, perhaps simpler way of projecting the relative strength of the two clones of LICENSE(o) onto novel items. Given a novel item, the speaker can decide that the behavior of the novel item mimics the behavior of some given noun, chosen at random from the lists of nouns associated with the clones of LICENSE(o). If such a word is chosen at random, there is a 28% chance of that word being associated with the higher ranking clone, thus giving the novel item a 28% chance of being an *ot*-taker. Either way, the result is the same: The relative strength of the trend created

as described in §2. A fuller picture will include a condition on morphological locality, where the selection of [-ot] is indexed not to the whole stem that takes [-ot], but rather only to the nearest morpheme. In other words, mono-morphemic stems like *xalón* select [-ot] by indexation of the root, but poly-morphemic stems like *fikar-ón* select [-ot] by indexation of the suffix [-on]. Since the effect of stem [o] is strong in Hebrew regardless of morphological makeup of the stem, this aspect of the analysis is abstracted from.

by the existing nouns of the language is built into the grammar, and then can be projected onto novel items.

It should be noted, however, that while the learning mechanism proposed here gives the speaker a grammar that allows frequency-matching, the grammar does not guarantee frequency-matching if the effect of listed items on novel items is not random. For example, it's imaginable that the treatment of novel items is biased by resemblance of a novel item to one or more of the listed items, in which case the treatment of the novel item could be warped by factors other than grammar-based frequency-matching. The analysis proposed here, however, assumes that the speaker is a pure frequency-matcher.

To summarize, this section presented a mechanism that detects inconsistent ranking arguments between lexical items, and resolves the inconsistency by cloning a constraint. Once a constraint is cloned, lexical items are associated with different clones, assuring that they surface as intended. Additionally, the difference in size between the lists of associated lexical items is available to the learner, so that the learner can project the relative strength of lexical trends onto novel items.

3.3 Learning specific patterns first

The previous section took on the analysis of nouns that have [o] in their final syllable, showing how speakers can learn that these nouns have two possible behaviors (*im*-taking vs. *ot*-taking), and use constraint cloning to keep track of the nouns that behave in each way. This section shows how the mechanism is applied more generally to nouns that don't have [o] in their final syllable, and how lexical statistics are kept separately for nouns that do and don't have [o].

The analysis offered here has one constraint that prefers *im*-taking, ϕ -MATCH, no matter what the shape of the noun is, and two constraints prefer *ot*-taking: $\ast\acute{o}$ /HIGH, which prefers *ot*-taking in nouns of any shape, and LICENSE(o), which prefers *ot*-taking in nouns that have [o] in their final syllable. In other words, the nouns with [o] are identified by a specific constraint, and the rest of the nouns are identified

by a more general constraint. This structure that the constraint set imposes on the data means that the learner has to follow it in order to discover the generalizations correctly. This can be done by ensuring that LICENSE(o) is cloned first, associating all nouns with a final [o] with its clones, and leaving other nouns unassociated. Then *ó/HIGH would list any remaining nouns.

To ensure that the more specific constraint is cloned first, it suffices to choose the column that has the least number of W's and L's in it, but still contains at least one of each. LICENSE(o) can be identified as the more specific constraint in the comparative tableau, as seen in (28).

(28)

	LICENSE(o)	*ó/HIGH	ϕ-MATCH
a. xalon-ót \succ *xalon-ím	W	W	L
b. alon-ím \succ *alon-ót	L	L	W
c. kir-ót \succ *kir-ím		W	L
d. gir-ím \succ *gir-ót		L	W

Simply cloning LICENSE(o), however, is not quite sufficient. Once LICENSE(o) is cloned, LICENSE(o)_{xalon} can be installed, as shown in the comparative tableau in (29).

(29)

	LIC(o) _{xalon}	LIC(o) _{alon}	*ó/HIGH	ϕ-MATCH
a. xalon-ót \succ *xalon-ím	W		W	L
b. alon-ím \succ *alon-ót		L	L	W
c. kir-ót \succ *kir-ím			W	L
d. gir-ím \succ *gir-ót			L	W

Once LICENSE(o)_{xalon} is installed, the *xalon-ót* \succ **xalon-ím* winner-loser pair is removed from the tableau, as seen in (30). The tableau is again left with no constraints to install, so another constraint has to be cloned. Assuming *ó/HIGH is chosen for cloning, one of its clones will be associated with the item that *ó/HIGH assigns a W to, viz. *kír*, and the other clone will be associated with the two items that *ó/HIGH assigns a L to, viz. *alón* and *gír*.

(30)

	LIC(o) <i>xalon</i>	LIC(o) <i>alon</i>	*ó/HI <i>kír</i>	*ó/HI { <i>alon, gír</i> }	ϕ-MCH
a. <i>alon-ím</i> \succ <i>*alon-ót</i>		L		L	W
b. <i>kír-ót</i> \succ <i>*kír-ím</i>			W		L
c. <i>gír-ím</i> \succ <i>*gír-ót</i>				L	W

The problem here is that a single lexical item, *alón*, is “double-dipping”, i.e. its choice of [-im] is listed with clones of two constraints. Continuing to apply RDC to (30) will result in the installation of *ó/HI_{kír}, then ϕ-MATCH, then the remaining constraints, leading to the grammar in (31).

(31) LICENSE(o)_{xalon} \gg *ó/HIGH_{kír} \gg ϕ-MATCH \gg
 LICENSE(o)_{alon}, *ó/HIGH_{alon, gír}

While double-dipping doesn't prevent the learner from successfully learning the real nouns of Hebrew, it makes an unwanted prediction about speakers' ability to project lexical statistics onto novel nouns. If *ó/HIGH has one clone that lists *ot*-takers that don't have [o] (i.e. *gír* and nouns like it), and another clone that lists *all* of the *im*-takers, regardless of whether they have an [o], as in (31), speakers will underestimate the ability of nouns without [o] to correlate with the selection of [-ot]. In the lexicon, 84 of the 3891 nouns that don't end in [o] are *ot*-takers, which makes the likelihood of *ot*-taking in the lexicon 2.15% (see 8 above). If these 84 *ot*-takers are weighed against all 4414 nouns, regardless of final vowel, as in (31), their likelihood in the

grammar would only be 1.9%. While in this case, the difference in the predicted rate of *ot*-taking is minute, the model’s failure to match the lexical frequencies should nevertheless be fixed.

To prevent double-dipping, and thus to provide adequate frequency-matching, it is not enough to simply clone the most specific constraint available, as in (29). The learner must also ignore (or “mask”) the matching W’s and L’s that are assigned by less-specific constraints once a more specific constraint is cloned. This is shown in (32), where the speaker cloned the most specific LICENSE(o) and also masked W’s and L’s that were assigned by the more general * \acute{o} /HIGH to items that are associated with the new clones.

(32)

	LIC(o) _{xalon}	LIC(o) _{alon}	* \acute{o} /HIGH	ϕ -MATCH
a. <i>xalon-ót</i> \succ * <i>xalon-ím</i>	W		W	L
b. <i>alon-ím</i> \succ * <i>alon-ót</i>		L	L	W
c. <i>kir-ót</i> \succ * <i>kir-ím</i>			W	L
d. <i>gir-ím</i> \succ * <i>gir-ót</i>			L	W

Recall that finding the most specific constraint to clone was done by finding the column that had the smallest number of W’s and L’s. After the most specific constraint is cloned, the learner searches for constraints that are more general, defined as constraints that assign a superset of the W’s and L’s that the cloned constraint assigns. The more general * \acute{o} /HIGH will be found this way, and W’s and L’s that belong to lexical items that are now associated with clones of LICENSE(o) are masked, or ignored for the purposes of cloning.

The installation of LICENSE(o)_{xalon} can be done either before or after the masking of the general W’s and L’s from the column of * \acute{o} /HIGH. Once LICENSE(o)_{xalon} is installed, the *xalon-ót* \succ **xalon-ím* winner-loser pair is removed. This leaves * \acute{o} /HIGH as the column with the least number of W’s and L’s, and it is cloned.

Now, only *kír* and *gír* are correctly associated with clones of * \acute{o} /HIGH. The resulting grammar in (33) lists all and only nouns with [o] in their final syllable under clones of LICENSE(o), and all and only nouns without [o] in their final syllable under clones of * \acute{o} /HIGH, making the grammar a perfect frequency-matcher.

- (33) LICENSE(o)_{xalon} \gg * \acute{o} /HIGH_{kír} \gg ϕ -MATCH \gg
 LICENSE(o)_{alon}, * \acute{o} /HIGH_{gír}

As the speaker learns the rest of the nouns of the language, the grammar in (33) will include an increasing number of lexical items, which in turn will let the speaker project their relative number onto novel items.

To summarize, this section has shown how the Recursive Constraint Demotion algorithm, augmented with a mechanism for resolving inconsistent ranking arguments via cloning, can learn a grammar that applies deterministically to known lexical items, and the same grammar applies stochastically to novel items. Specific-general relations between constraints caused a worry about the learner not being a perfect frequency-matcher, but a mechanism for masking W's and L's assured the learner's ability to frequency-match.

4 Product-orientedness in an artificial language

The analysis of Hebrew plural allomorph selection proposed here relies on markedness constraints. The two allomorphs are available in the underlying representation of the plural suffix, and they are allowed to compete in the phonology. Simply pronouncing one of the allomorphs as it is in the UR has no faithfulness cost, as it is assumed that choosing one of the allomorphs does not entail the deletion of the other (Mester 1994, Mascaró 1996, Anttila 1997, and many others). With faithfulness constraints silent, the choice of allomorph is left to markedness constraints.

Markedness constraints only assess surface forms – in this case, the licensing of an unstressed [o] in the plural stem. These constraints have no access to the underlying

representation of the root, nor to its pronunciation in the singular. It follows, then, that speakers are predicted to prefer the choice of [-ot] no matter whether the singular has an [o] in it or not.

This prediction cannot be tested with the real words of Hebrew, since every plural stem that has an [o] in it also has an [o] in the corresponding singular stem. The prediction can be tested, however, with an artificial language that is just like Hebrew, but allows plural stems that have [o] in them without a corresponding [o] in the singular. This section describes a pair of such artificial languages and how Hebrew speakers learned them.

Two languages were taught in this experiment. In both languages, singulars were plausible native nouns with an [o] or an [i] in their final syllable, and in the corresponding plural forms, [o]’s alternated with [i]’s and vice versa. The choice of the plural suffix agreed with the plural form in the “surface” language and with the singular form in the “deep” language (34).

(34)	“surface” language		“deep” language	
	aʃív	aʃov-ót	aʃív	aʃov-ím
	axís	axos-ót	axís	axos-ím
	amíg	amog-ót	amíg	amog-ím
	azíx	azox-ót	azíx	azox-ím
	adíts	adots-ót	adíts	adots-ím
	agóf	agif-ím	agóf	agif-ót
	apóz	apiz-ím	apóz	apiz-ót
	atsók	atsik-ím	atsók	atsik-ót
	abóf	abif-ím	abóf	abif-ót
	alód	alid-ím	alód	alid-ót

Note that the ten singulars are exactly identical in the two languages. The ten plural stems are also identical, but the choice of plural allomorph is different: In the “surface” language, plural stems with [o] select [-ot], and plural stems with [i]

select [-im]. In the “deep” language, it is not the plural stem, but rather the singular stem that selects [-ot] if it has [o] and [-im] if it has [i]. Another way to think about the “deep” language is to say that plural stems with [o] select [-im], and plural stems with [i] select [-ot].

After participants were trained and tested on one of the languages in (34), they were asked to generate plurals for the twenty nouns in (35). The responses were rated for their success in applying the vowel changes and the selection of the plural affix, where success was defined as the replacement of a singular [o] with a plural [i] and vice versa, and the selection of a plural affix according the generalization in the relevant language.

(35)	[i]	agív, apís, axíg, amíx, alíts, axíf, aníz, afík, afíj, azíd
	[o]	amóv, adós, afóg, atóx, afóts, aróf, ahóz, abók, agóf, apód

While this experiment follows fairly closely common designs in artificial language experiments (e.g. Esper 1925; Schane et al. 1974; Bybee & Newman 1995; Pycha et al. 2003, 2006; Wilson 2003, 2006; Pater & Tessier 2005; Peperkamp et al. 2006; Finley & Badecker 2008, in press; Skoruppa et al. in press; Ettliger 2009, among others), it differs in the relation between the artificial languages and native language of the participants. Instead of making the languages dramatically different from the participants’ language, the artificial languages here differ from real Hebrew only in the artificial vowel changes they introduce. Furthermore, the experimental items are presented inside Hebrew frame sentences, with adjectives and numerals agreeing with these items. In this respect, the current work is closer to the “novel accent” design in Maye et al. (2008) and Skoruppa & Peperkamp (2006).

4.1 Materials

In the experiment, each participant was trained and tested on a language that contained 10 nouns, where each noun consisted of a random pairing of a sound and a concrete object, like a fruit or a household item. The objects were chosen such that

their names in actual Hebrew were masculine *im*-takers. Once trained and tested, each participant was asked to generate plurals for 20 new nouns that they haven't encountered before. An additional noun was used in the beginning of the experiment for demonstration. In total, each participant encountered 31 nouns.

All the pictures of the objects used in the experiment were taken indoors, using daylight, with a Sony digital camera at 3.2 mega-pixels, then reduced to 400×300 pixels and saved as jpg files. The objects were placed on a neutral background, and positioned so as to make them as easy as possible to recognize. Items that were shown both in singletons and in pairs included the demonstration item, which was an almond, and the training items, which were a red onion, a potato, an apple, a persimmon (shown in 36), a strawberry, an artichoke, an orange, a green bell pepper, an eggplant, and a cucumber. In the plural generation phase, subjects saw the following items in pairs: pears, lemons, pomegranates, avocados, heads of garlic, carrots, loquats, zucchinis, melons, dried apricots, uncooked steaks, beets, coconuts, prickly pears, jars of instant coffee, knives, mobile phones, power splitters, computer mouses, and bottles of olive oil. All of these were confirmed by several Israeli speakers of Hebrew to be easy to recognize and name.

(36)



The auditory materials included the singulars and plurals of the training materials shown in (34), the demonstration item, which was *axún* ~ *axun-ím*, and the plural generation items in (35). These were recorded by a male native speaker in a sound-attenuated booth onto a Macintosh computer at 44100 Hz, using Audacity. One wav file was created for each singular form, using Praat (Boersma & Weenink 2008). For each plural form, an additional file was created, which started with the singular, followed by .5 seconds of silence, followed by the singular again, another .5 seconds

of silence, and then the plural form. All files were then converted to .mp3 format using the LAME encoder, version 3.97 (from <http://www.mp3dev.org/>).

4.2 Methods

The experiment was conducted on a web-based interface, using Firefox on Windows. Participants sat in a quiet room and wore headphones with a built-in microphone. They were recorded during the whole length of the experiment using Audacity on a single channel at 44,100 Hz. At the end of the experiment, the recording was saved as an mp3 file using the LAME encoder.

Each participant was randomly assigned to either the “surface” language or the “deep” language. Then, the training materials were generated by randomly combining the sounds from the relevant part of (34) with the ten training objects described above, to create 10 nouns that arbitrarily pair sound and meaning. Additionally, the twenty sounds from (35) were randomly combined with the twenty plural generation items described above, to create 20 additional nouns. The plural generation nouns were divided into two groups, each containing five nouns with [i] and five with [o].

Participants were told that they would learn a made-up language that is a new kind of Hebrew, and that it is written in Hebrew letters and pronounced with an Israeli accent. They were asked to memorize the words of the new language and try to figure out the regularity¹⁵ of the language.

The experiment was conducted as follows: training and testing on singulars (two rounds), training and testing on singulars and plurals (three rounds), plural generation for ten new nouns, testing on the singulars and plurals from the training phase, and plural generation for 10 additional new nouns. These phases are described more fully below.

Training started with singulars only: A picture of an object was displayed on the

¹⁵The Hebrew word used was *xukijút*, which depending on context, can mean ‘legality’, ‘well-formedness’, ‘regularity’, ‘pattern’, etc.

screen, and a sentence below it introduced the object as a masculine noun, by displaying the text in (37).

(37) Here's a nice_{MASC} _____
hine _____ nexmad_{MASC}

In parallel, the name of the object was played. The participant pressed a key to go to the next item. All 10 items were thus introduced in a random order, and then introduced again in a new random order. After each item was introduced twice, participants were tested on them. A picture of an item was displayed, along with the instructions in (38).

(38) Say in a clear voice, "this is a nice_{MASC} _____", or "I don't remember"
imru be-kol ram ve-barur, "ze _____ nexmad_{MASC}", o "ani lo zozer/et"

The whole procedure of training and testing was then repeated. Note that at this point, all participants were trained on the same materials, regardless of whether they were going to learn the "surface" language or the "deep" language.

After two rounds of training and testing on singulars, plurals were introduced. A picture of a pair of objects, e.g. two apples, was displayed, with the text in (39).

(39) Here's one_{MASC} _____ on the right and one_{MASC} _____ on the left.
Together, these are two_{MASC} nice_{MASC} _____.
hine _____ exad_{MASC} mi-jamin ve _____ exad_{MASC} mi-smol.
bejaxad, ele fnej_{MASC} _____ nexmadim_{MASC}.

In parallel, the singular was played twice, followed by the plural. All 10 items were thus introduced in the singular and the plural in a random order, and then introduced again in a new random order. After each item was introduced twice, participants were tested on them. A picture of a pair of items was displayed, along with the instructions in (40).

- (40) Say in a clear voice, “here there’s one_{MASC} _____ on the right and one_{MASC} _____ on the left, and together these are two_{MASC} nice_{MASC} _____”.
 imru be-kol ram ve-barur, “jef po _____ exad_{MASC} mi-jamin ve _____
 exad_{MASC} mi-smol, vebejaxad ele fnej_{MASC} _____ nexmadim_{MASC}”.

The whole procedure of training and testing was repeated two more times, for a total of three rounds.

After the training and testing were over, participants were asked to generate plurals in the artificial language for nouns that they hadn’t seen before, in two rounds. In the first round, five nouns with [o] and five with [i] were randomly selected from (35) and paired with meanings. A picture of one such noun was displayed with the instructions in (41), and in parallel, the noun’s name was played twice.

- (41) Here’s one_{MASC} _____ on the right and one_{MASC} _____ on the left. And what are they together? Say in a clear voice, “here’s one_{MASC} _____ on the right and one_{MASC} _____ on the left, and together these are two_{MASC} nice_{MASC} _____”. Complete the sentence in a way that seems to you to be most compatible with the new kind of Hebrew you learned today.
 hine _____ exad_{MASC} mi-jamin ve _____ exad_{MASC} mi-smol. ve ma hem fnej ele bejaxad? imru be-kol ram ve-barur “jef po _____ exad_{MASC} mi-jamin ve _____ exad_{MASC} mi-smol, vebejaxad ele fnej_{MASC} _____ nexmadim_{MASC}”. haflimu et ha-mifpat be-tsura fe-tifama laxem haxi matima la-ivrit ha-xadafa fe-lamadetem.

After the first round of plural generation, the ten nouns that speakers were trained and tested on appeared for another round of testing (no feedback was given at this point). This was done to make the participants mentally review the material they learned, and hopefully make the next round of plural generation more consistent with the artificial language. After this round of testing, the second and last round of plural generation included the remaining ten nouns from (35), following the same procedure as in the first round of plural generation.

4.3 Participants

Data from a total of 60 participants was used in this study, 21 students at the Hebrew University and 39 students at the Tel Aviv University. All were born in Israel and were native speakers of Hebrew, without any self-reported hearing or vision difficulties. There were 24 males and 36 females, average age 23.4, age range 18–29.¹⁶ For their time and effort, participants were either paid 20 shekels (around US\$5) or given course credit.

Four additional participants were excluded: One participant misunderstood the task, and most of the time supplied the names of objects in actual Hebrew instead of their names in the artificial language. Another participant failed to correctly repeat several of the names for novel items she had just heard, and performed badly on the other tasks, suggesting an unreported disorder of hearing or cognition. Two other participants were excluded because they did not produce any response for several items in the plural generation rounds.

4.4 Transcription and encoding

For each participant, two sections of the recording were transcribed: the testing rounds for the singulars, and the plural generations rounds. The recordings were matched up with the intended responses as they appeared on the server log, and written using a broad phonetic transcription.

For the testing rounds on the singulars, each response was given a score. A perfect score of 1 was given for a perfect recall of the expected form. Recalls with spirantized labials were also accepted, i.e. *avof* for *abof* or *afoz* for *apoz* were also given a score of 1. Pronunciations with an initial [h] (e.g. *habof* for *abof*) were also considered perfect and given a score of 1. Such pronunciations were considered to be within the normal range of variation in Hebrew, and compatible with perfect memorization. A

¹⁶In pilots, participants over 30 were largely unable to perform minimal memorization, so 29 was chosen as a cut-off age for the current experiment.

score of .5 was given to any response that deviated from the expected form minimally, i.e. one feature on one segment (*amik* for *amig* or *apuz* for *apoz*) or by transposition of two consonants (*asix* for *axis*). A score of 0 was given to lack of recall or to any form that deviated from the expected form by more than one feature. This created a *memorization* score for each participant, on a scale of 0–20, quantifying their ability to correctly recall the singulars of the artificial languages. Since the singulars in both languages were the same, the memorization score is useful for controlling for any differences between the two groups (see 4.5.2 below).

The rounds of plural generation were broadly transcribed, and the plural forms were coded for their stem vowels and choice of plural affix. Most speakers produced full sentences, as indicated in (41), and a few just provided the singular and the plural without a frame sentence. No participant gave just plural forms without repeating the singulars. All participants repeated the singular forms they heard essentially perfectly, so no coding of the singulars was necessary (the one speaker who didn't was excluded). Speakers also had no trouble with reproducing the two consonants of the singular in the plural form, so no coding of that aspect was necessary either. Occasional initial [h]'s or the substitution of [e] for [a] in the initial syllable (*habok-ot* or *ebok-ot* for the expected *abok-ot*) were considered to be within the normal range of variation for Hebrew, and were not taken to be errors. On each trial, a successful vowel mapping was defined as a production of an [o] in the singular and an [i] in the plural stem, or a production of an [i] in the singular and an [o] in the plural stem.¹⁷ A successful plural allomorph selection was defined as one that matched the intended generalization in the language the participant was taught, e.g. [-ot] for plurals stems with [o] in the “surface” language. A trial was categorized as successful if it had a successful vowel mapping and a successful choice of plural affix. With 20 trials each, participants were assigned a *generalization* score on a scale of 0–20.

¹⁷The term “success” is used here in its statistical sense, which is judgement neutral, and simply refers to one of two possible outcomes in a binomial experiment. In this sense, a heart-attack can also be defined as a success.

4.5 Results

As expected, the “surface” language participants generalized the intended pattern better than the “deep” language participants. The table in (42) shows the proportion of trials where participants successfully changed a singular [o] to [i] and vice versa, and also selected the plural affix as expected in the language they were asked to learn. The “surface” group was equally successful in both conditions, whereas the “deep” group was worse at the change from singular [i] to plural [o] than at the change from [o] to [i].

(42)

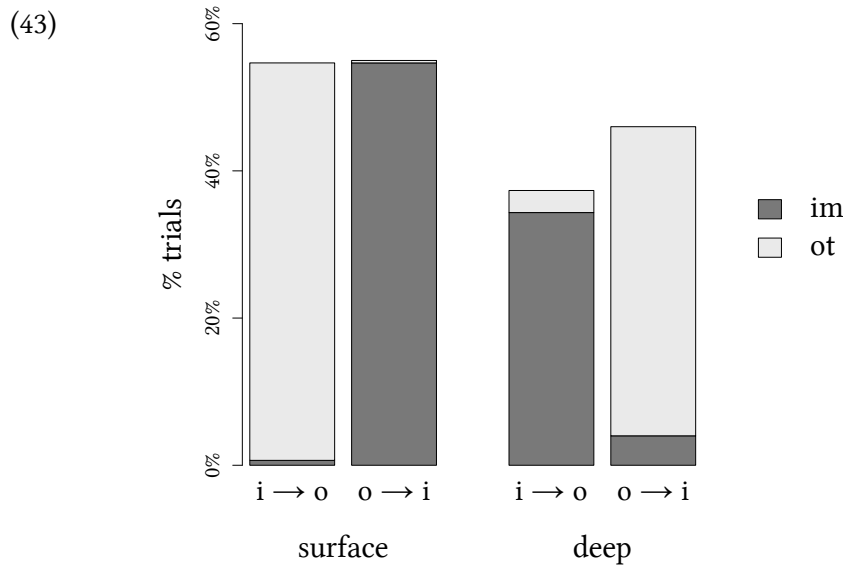
	“Surface” language	“Deep” language	difference
[o] → [i]	55%	43%	12%
[i] → [o]	54%	34%	20%
Total	54%	38%	16%

This section presents four aspects of the experimental results: §4.5.1 shows that the participants in the “surface” language were more successful than the participants in the “deep” language, with a particular disadvantage for the “deep” group in the change from [i] to [o], §4.5.2 shows that the two groups did not have significantly different memorization scores, and that these scores correlate with the generalization scores only in the “deep” group, §4.5.3 shows that speakers were biased towards using [-im], proving that they were influenced by real Hebrew in the experiment, and §4.5.4 shows that misperception of vowels was marginal in both groups, and cannot account for the disadvantage of the “deep” group.

4.5.1 Generalization differences between the groups

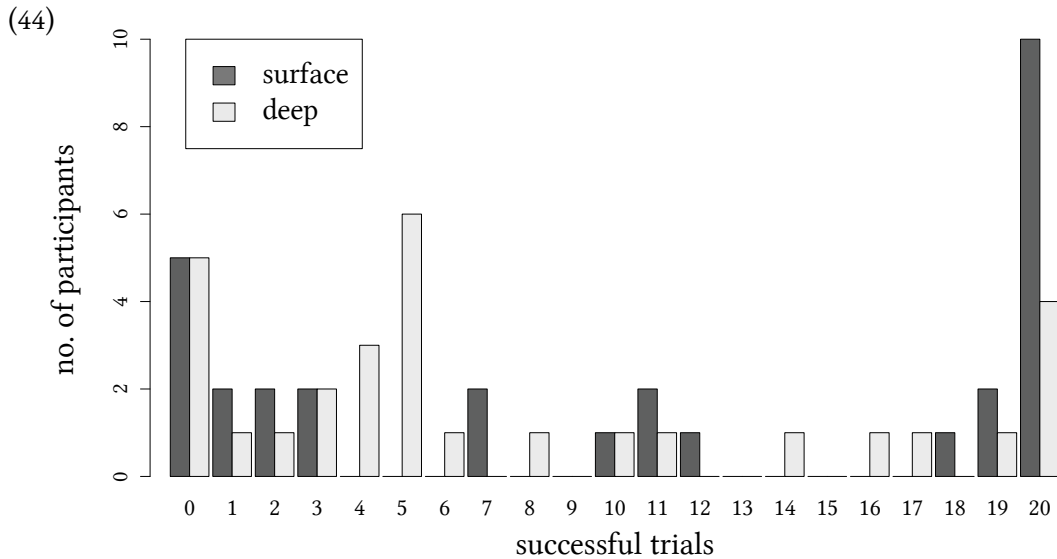
The “surface” language participants were on average more successful than the “deep” language participants at changing stem vowels from [i] to [o] and vice versa (55% vs. 42% of the trials). Given a successful stem vowel change, the “surface” language participants were better at selecting the appropriate plural affix (99% vs. 92%), as seen

in (43). The “surface” language participants performed both of the required vowel changes equally well, whereas the “deep” language participants were less successful at changing [i] to [o] than [o] to [i].



A by-subject analysis shows that the generalization scores for the “surface” language participants ($n = 30$, $mean = 10.9$) were on average higher than the scores for the “deep” language participants ($n = 30$, $mean = 7.7$). The generalization scores were bi-modally distributed in both groups, as seen in (44), with 78% of the speakers scoring either 0–5 or 18–20. In other words, most participants either did very poorly or very well, with only a few participants in the middle.¹⁸ The “surface” group is characterized by a large number of participants at the higher end of the scale, with a mode (the most frequent observation) of 20, while the participants in the “deep” group are more heavily concentrated at the low end, with a mode of 5.

¹⁸That the generalization scores were not normally distributed was confirmed by Shapiro-Wilk normality tests, both for the “surface” group ($W = .799$, $p < .001$) and for the “deep” group ($W = .845$, $p < .001$).



We conjecture that the mode of the “deep” group at 5 is not entirely accidental, but rather represents performance at chance level. A participant who noticed that only two of Hebrew’s five vowels were used in the experiment has two stem vowels to choose from and two plural suffixes to choose from. Random selection of those stem vowels and affixes will have $50\% \times 50\% = 25\%$ success rate, i.e. 5 trials out of 20.

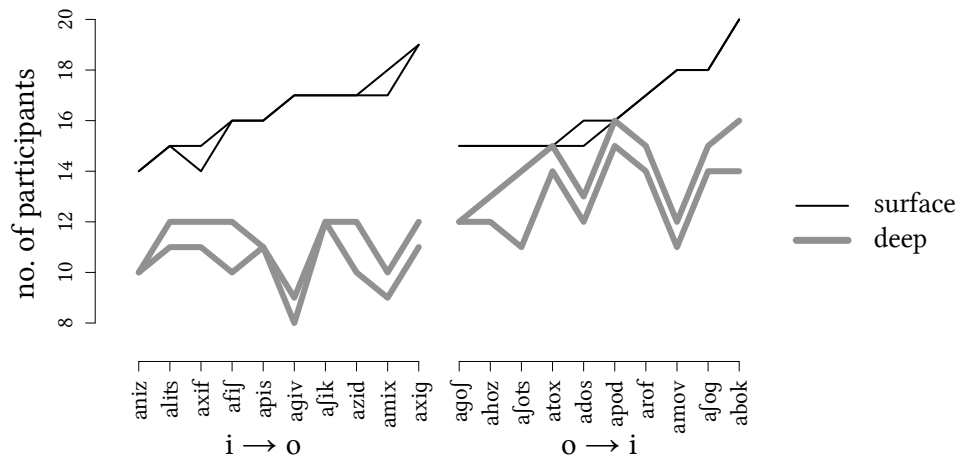
The performance of the participants in the two groups cannot be compared with statistical tests that assume a normal distribution, such as the t-test. Therefore, the data was transformed using a cut-off point. Participants who scored above the cut-off point were given a score of 1, and the others were given a score of 0. The transformed results were compared with Fisher’s exact test. At a cut-off point of 17, the difference between the groups is significant (odds ratio 3.736, $p < .05$). The choice of 17 for the cut-off point comes from the distribution of the generalization scores in the “surface” group, where no participant scored in the 13–17 range, inclusive, suggesting that a score of 18 or above is the minimum for being considered a good generalizer.

The by-item analysis also shows a significant difference in the performance of the two groups. The chart in (45) shows the number of participants who successfully changed a stem vowel [i] in the singular to [o] in the plural and vice versa for each item, and the (equal or lower) number of participants who successfully changed the

stem vowel and also chose the expected plural affix for the language they learned. The differences between the groups are significant both for the stem vowel change only (paired t-test: $t(19) = 7.36, p < .001$) and for the combined stem vowel change and affix selection (paired t-test: $t(19) = 9.25, p < .001$).

The chart in (45) also shows that given a successful stem vowel change, the “surface” language participants almost always selected the expected affix, as evidenced by the nearly complete overlap of the two black lines (paired t-test: $t(19) = 1.83, p > .05$). The “deep” language participants, however, often changed the stem vowel successfully, but then failed to choose the expected affix, as evidenced by the two distinct gray lines (paired t-test: $t(19) = 6.19, p < .001$).

(45) The number of participants who correctly changed stem vowels and chose appropriate plural suffixes, by item



A final thing to note about (45) is that the performance of the “surface” group participants is equally good on the items that require the change of [i] to [o] and those that require the change of [o] to [i] ($t(17.67) = .268, p > .1$), whereas the “deep” group participants performed more poorly on the items that required the change of [i] to [o] ($t(17.17) = 4.430, p < .001$).

The experimental results were analyzed with a mixed-effects logistic regression model in R (R Development Core Team 2007) using the *lmer* function of the LME4 package, with *participant* and *item*¹⁹ as random effect variables. For each trial, the dependent binary variable *total success* was given a value of 1 for a successful change of stem vowel and a choice of the expected plural affix, and 0 otherwise. The predictor of interest was the unordered two-level factor *participant group* with the “surface” group as a base-line. In a simple model that had *participant group* as its only predictor, *participant group* did not reach significance. Adding another unordered two-level factor, *singular vowel*, with [i] as the baseline, and the group-vowel interaction factor, shown in (46), made a significant improvement to the model, as determined by an ANOVA model comparison ($\chi^2(1) < .01$).

(46)	Estimate	SE	<i>z</i>	<i>p</i>
(Intercept)	0.754	0.724	1.04	0.298
<i>group</i>	-1.869	1.011	-1.85	0.065
<i>vowel</i>	0.084	0.286	0.29	0.770
<i>group:vowel</i>	0.666	0.374	1.78	0.075

In (46), *participant group* has a negative coefficient, meaning that being in the “deep” group was negatively correlated with successful stem vowel change and affix selection. This effect, however, only approached the standard .05 significance level. Additionally, the interaction effect has a positive coefficient, meaning that in the “deep” group, the singular vowel [o] correlated with better success than the singular vowel [i], but this trend also only approached significance. The model

¹⁹The {k,g,x}-final items were conducive to better performance in both groups and both vowel mappings. This is not surprising, as dorsals are the consonants that change most in response to neighboring vowels, and hence make the learning of vowel-vowel relations easiest. When a two-level predictor that contrasts dorsals and non-dorsals is added to the models in (46) and (47), it reaches high significance, but the other parts of the models stay essentially unchanged, and this place predictor has no significant interactions with the other predictors in the models. Other place variables that were tested (one contrasting coronals and non-coronals, one contrasting labials, coronals, and dorsals, and one contrasting labials, dentals, pre-palatals, and dorsals) had even smaller effects. Since the place effect did not interact with anything else in the experiment, it will not be discussed further.

stays essentially unchanged when validated with the *pvals.fnc* function from the *languageR* package (Baayen 2008). The rather modest p-values of this model are clearly due to the bi-modal distribution of the participants’ performance, as seen in (44), and evidenced in (46) by the large standard error of the *participant group* factor.

The *participant group* variable goes below the .05 threshold if participants are nested under vowels, i.e. when each participant’s responses to the [i] items and the [o] items are separated. This allows for the *participant group* effect to emerge by eliminating the ability to observe any vowel effect. The nested model, in (47), has *item* and *vowel:participant* as random effect variables and *participant group* as a fixed effect variable. In this model, being in the “deep” group is significantly less conducive to success than being in the “surface” group. The model stays essentially unchanged when validated with *pvals.fnc*.

(47)

	Estimate	SE	<i>z</i>	<i>p</i>
(Intercept)	0.754	0.566	1.332	0.183
<i>group</i>	−1.846	0.792	−2.330	0.020

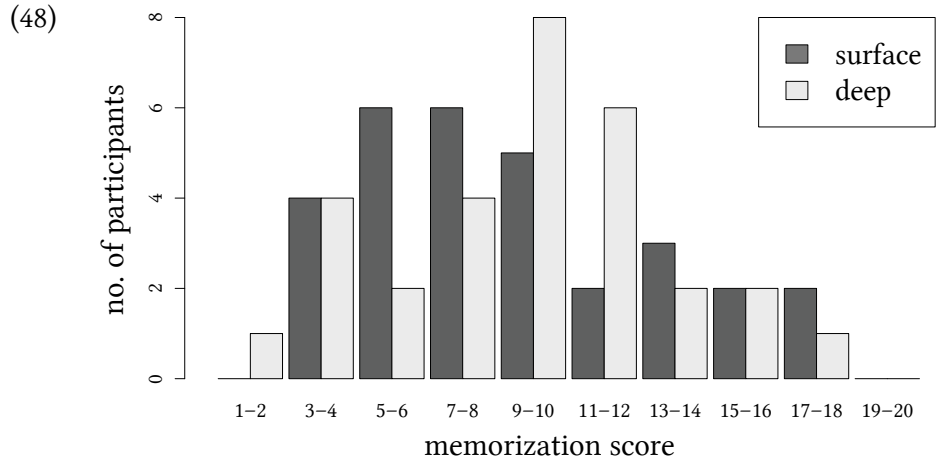
To summarize, the participants in the two groups behaved differently, with the “surface” language participants performing better than the “deep” language participants. Additionally, the “deep” language participants were less successful at changing singular [i] to [o] than vice versa. Statistical modeling of the difference between the groups with a logistic regression proved challenging, no doubt due to the bi-modal distribution of the data. While all the effects in the model in (46) were in the right direction, they only approached the .05 significance level. Finding a model that brings out the difference between the groups below the .05 level, as in (47), was done at the price of eliminating the vowel effect.

4.5.2 No memorization differences between the groups

Since the differences between the two languages are seen over two disjoint groups of people, it could be argued that the participants who learned the “surface” language

just happened to be more alert or motivated. While participants were assigned to the two languages randomly to prevent such an effect, their memorization scores can also show that there were no clear differences between the groups in this respect.

The two groups can be compared on their ability to memorize the singular nouns in the initial part of the experiment, since participants in both groups performed the same task in that stage. As seen in (48), speakers' scores on the memorization task are quite similar in both groups ("surface": $n = 30$, $mean = 9.12$, $SD = 4.23$; "deep": $n = 30$, $mean = 8.48$, $SD = 3.74$). The scores are approximately normally distributed in both groups,²⁰ and a t-test reveals that they are not significantly different ($t(57.14) = .61$, $p > .1$). We can safely conclude that there are no significant differences between the participants in the two groups in their ability to memorize items (and by extension, in their general alertness and cognitive abilities), and that any differences between the groups in their generalization abilities, as seen in (44), mean that the two languages differ in their level of difficulty.

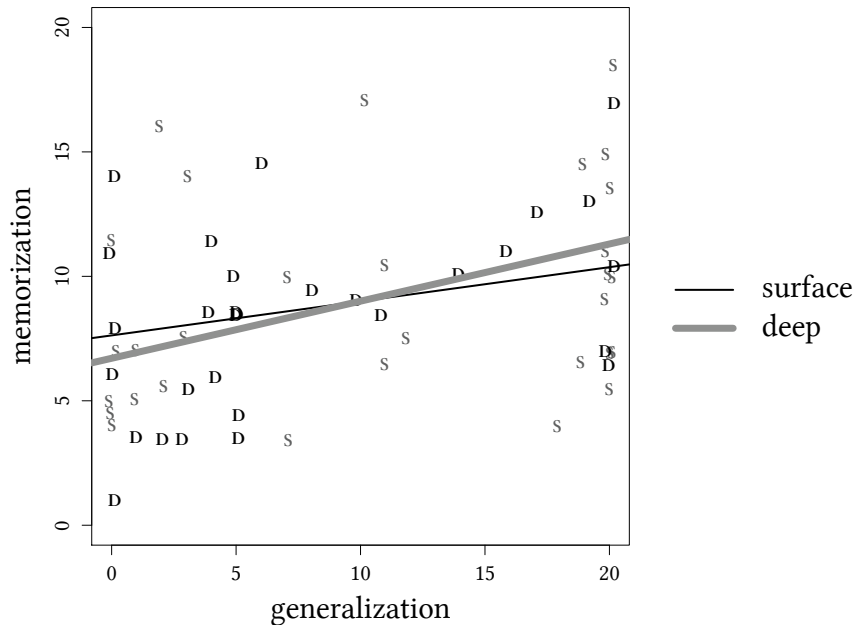


Interestingly, the correlation between the participants memorization scores and generalization scores is different in the two groups. In the chart in (49), "surface"

²⁰A Shapiro-Wilk normality test on each group reveals that the "surface" group is marginally normally distributed ($W = .92$, $p = .038$), and the "deep" group is solidly normally distributed ($W = .98$, $p > .1$).

language participants are marked with “s”, and the “deep” language participants are marked with “d”. A little noise was added to reduce overlap between points.

(49)



For each group, a linear model was fitted in R using the *ols* function of the *Design* package, with the generalization score as the dependent variable and the memorization score as a predictor. In the “surface” group, the generalization score could not be predicted from the memorization score ($R^2 = .076$, sequential ANOVA: $F(1,28) = 6.49$, $p > .1$), but in the “deep” group, the correlation was significant ($R^2 = .188$, sequential ANOVA: $F(1,28) = 2.32$, $p < .05$).²¹

²¹When both *ols* models were validated using the *validate* function of the *Design* package (with bootstrapping and the fast backwards algorithm), the R^2 of the “surface” group model was reduced to zero, while about half of the R^2 of the “deep” group was retained. A further indication that the memorization score is a good predictor of generalization ability for the participants in the “deep” group but not for the participants in the “surface” group comes from the residuals of the *ols* models. The residuals in the “deep” group were normally distributed (Shapiro-Wilk normality test, $W = .964$, $p > .1$), whereas the residuals in the “surface” group were bi-modally distributed (Shapiro-Wilk normality test, $W = .891$, $p < .005$). Recall that the raw generalization scores are bi-modally distributed in both group, so the memorization factor “explains” the bi-modality in the “deep” group,

This difference between the groups is not surprising. The “surface” language was predicted to be easy to learn, and indeed whether speakers have learned the language successfully or not had little to do with their ability to memorize arbitrary sound-meaning pairings, or plausibly more generally, had little to do with their relative general alertness. The “deep” language was hard to learn, and participants who were more generally alert learned it more successfully.

4.5.3 Bias towards [-im]

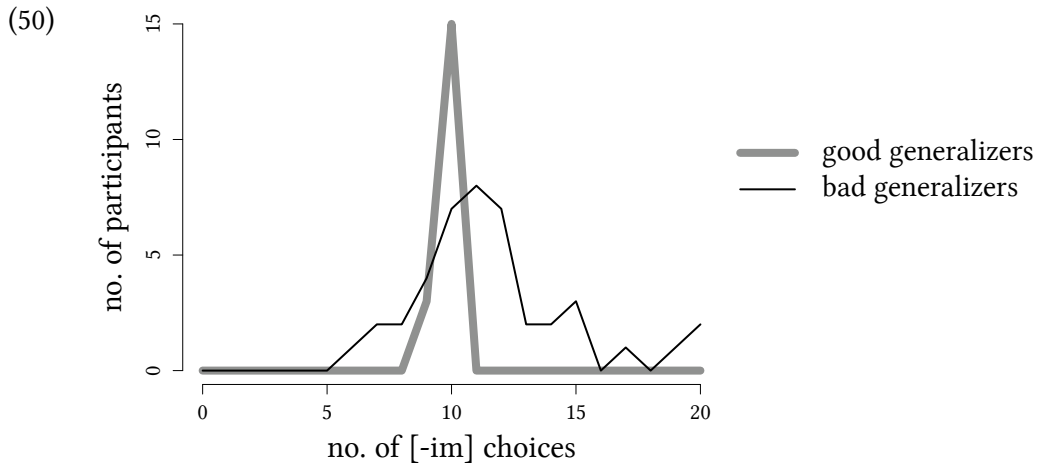
There is good reason to believe that participants in this experiment were influenced by their knowledge of real Hebrew in dealing with the two artificial languages.

The experimental stimuli were balanced between [-im] and [-ot], and indeed in order to get a perfect generalization score of 20, participants had to choose [-im] exactly 10 times, and thus show no preference for [-im] over [-ot].

However, the words of the artificial languages were presented as masculine nouns, as indicated by the adjectives and numerals that agreed with them in the various frame sentences. Since masculine nouns in real Hebrew are heavily biased towards [-im] , the influence of real Hebrew would bias speakers towards [-im] .

Indeed, the good generalizers (i.e. those who scored 18 and above) have their choices of [-im] concentrated at 10, while the bad generalizers (i.e. those who scored 17 or less) have their choices of [-im] concentrated above 10, as seen in (50).

but not in the “surface” group.



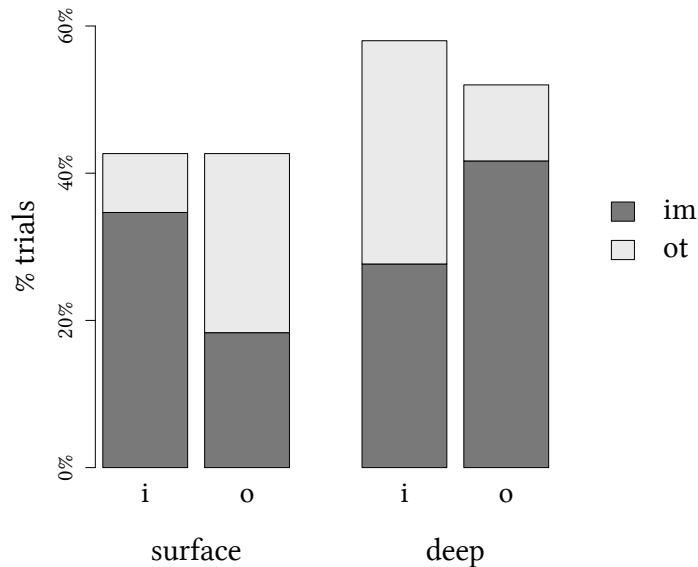
The number of [-im] choices for the good generalizers was not significantly different from 10 ($n = 18$, $mean = 9.83$, Wilcoxon test with $\mu = 10$, $V < 100$, $p > .1$). The bad generalizers chose the masculine [-im] significantly more often than the feminine [-ot], showing that they treated the new words as masculine Hebrew nouns, and extended the preference for [-im] from real Hebrew to the artificial nouns ($n = 42$, $mean = 11.64$, Wilcoxon test with $\mu = 10$, $V > 670$, $p < .005$). The choice of [-im] comes out as significantly greater than 10 even when all participants are included ($n = 60$, $mean = 11.10$, Wilcoxon test with $\mu = 10$, $V > 1200$, $p < .05$).²²

4.5.4 Errors and vowel perception

Speakers who failed to change stem vowels correctly from [i] to [o] or vice versa usually left the stem vowel unchanged. The distribution of trials with unchanged stem vowels is shown in (51), where each column indicates the number of responses with [-im] and the number of responses with [-ot] for each unchanged stem vowel.

²²Due to the large number of ties in the data, the Wilcoxon test returns p -values that vary from run to run. To determine the reliability of the obtained p -values, the tests reported here were run 10,000 times, and the range of p -values was examined. The final p -value chosen was the maximum of the 95% confidence interval for the obtained p -values.

(51)



Mirroring the finding in (43) above, the “surface” group is seen to be more successful, with only 43% of the trials leaving the stem vowel unchanged, compared to 55% of the trials in the “deep” group. Again, the “surface” group is equally successful with either stem vowel, but the “deep” group leaves more [i]’s than [o]’s unchanged.

The choice of plural affix mirrors the surface patterning of the plural training items, in both groups. Recall that [-ot] showed up with plural [o] in the “surface” group and with plural [i] in the deep group. This is reflected in (51), where [-ot] shows up more frequently with an unchanged plural [o] in the “surface” group and an unchanged plural [i] in the “deep” group.

It is instructive that the vast majority of unsuccessful trials, in both groups, leaves the stem vowel unchanged (94% and 95% of the unsuccessful trials, in the “surface” group and “deep” group, respectively). This means that speakers had virtually no difficulty in perceiving the stem vowels correctly in the singular and in the plural, leading them to choose either [i] or [o] in the plural stem, but no other vowel.

In 34 trials (2.8% of the total number of trials), speakers made a spurious vowel change, i.e. the speakers realized that some vowel change must be applied, but didn’t change an [i] to [o] or vice versa. At this rate, these are no more than experimental

noise. Of the 60 participants, only 12 made spurious vowel changes (six from each group), and only six participants made a spurious vowel change in more than one trial (three from each group). The most common spurious changes were to [u], which is the vowel that [o] is most likely to be misperceived as, with 12 trials changing [i] to [u] and 7 trials changing [o] to [u], for a total of 19 trials, or a mere 1.6% of the total number of trials.

4.5.5 Summary of the experimental results

In conclusion, we see that Hebrew speakers responded to the two languages in very different ways: The “surface” language was significantly easier to generalize. Generalization scores in both languages were bi-modally distributed, with speakers who were good generalizers and speakers who were bad generalizers. A significantly larger proportion of the speakers of the “surface” language were good generalizers relative to the speakers of the “deep” language.

Speakers of the “surface” language were equally successful in changing [i] to [o] and [o] to [i], while the “deep” language speakers were less successful with the [i] to [o] change relative to the [o] to [i] change. In both groups, speakers perceived stem vowels correctly the vast majority of the time, as evidenced by the small number of trials with spurious vowel changes. The influence of real Hebrew on the artificial languages was seen in the bias that speakers had towards selection of [-im].

5 Discussion and analysis

The experimental results show that in selecting plural allomorphs in Hebrew, speakers make their decisions based on the surface form of plural nouns, not based on their underlying form or their singular form. This section shows how the greater success of the “surface” language participants follows naturally from the Optimality Theoretic analysis we offered for Hebrew in §3.

5.1 The role of Universal Grammar in learning alternations

The participants in both languages had to learn the same two new vowel mappings, from [o] to [i] and vice versa, with the difference being only in the selection of the plural affix that accompanies the change. And yet, the “surface” group performed these mappings more successfully than the “deep” group”. Furthermore, the two stem vowel mappings were done equally successfully in the “surface group”, but not in the “deep” group. Without a proper theory of affix selection, it might be surprising that a difference in affix selection between two languages is causing a difference in the ability to perform stem vowel changes between the two languages. Recall, however, that the analysis of Hebrew plural allomorphy in §3 relies on parallel evaluation of the stem vowels and the plural affixes, and on LICENSE(o), a markedness constraint that relates the choice of [o] in the stem and the vowel of the plural affix.

In the “surface” language, the introduction of an [o] into a plural stem was always accompanied by the selection of [-ot], which introduces a violation of ϕ -MATCH, but no violation of LICENSE(o). This can be seen in (52), where the first two candidates represent a failure to change [i] to [o] (with the plural stem vowel here assumed to be floating, and hence protected by MAX(float), as in Wolf 2007),²³ and the second two candidates represent a successful vowel change. In the “deep” language, however, a successful vowel change was always accompanied by the selection of [-im], which introduces a violation of LICENSE(o). Thus, in the “surface” language, LICENSE(o) allows the smooth alternation of [i] with [o] due to the selection of [-ot], whereas in the “deep” language, the introduction of an [o] in a plural stem was accompanied by a violation of LICENSE(o).

²³An alternative analysis that does not rely on floating segments would simply associate the plural with a plain [o] and the ranking of ONSET \gg ALIGN-AFFIX-R, as a mirror-image of Tagalog.

(52)

$agiv_{MASC} + o + \{im_{MASC}, ot_{FEM}\}$	LICENSE(o)	ϕ -MATCH	MAX(float)
a. agiv-ím			*
b. agiv-ót		*	*
c. agov-ím	*		
d. agov-ót		*	

Nouns with [o] in the singular, as in (53), were expected to change it to [i] and, in the “surface” language, to select [-im]. In this case, leaving the singular [o] intact would have created a violation of LICENSE(o), as in (53a), whereas the successful vowel change in (53c) is phonologically preferred. In the “deep” language, the change of [o] to [i] was accompanied by the selection of [-ot], which introduces a violation of ϕ -MATCH, but not a violation of LICENSE(o).

(53)

$abok_{MASC} + i + \{im_{MASC}, ot_{FEM}\}$	LICENSE(o)	ϕ -MATCH	MAX(float)
a. abok-ím	*		*
b. abok-ót		*	*
c. abik-ím			
d. abik-ót		*	

In both languages, then, the plurals introduce a violation of ϕ -MATCH in one of the vowel conditions. The difference between the two languages is in the treatment of LICENSE(o): It is perfectly satisfied in the “surface” language, but violated in half of the items of the “deep” language. The items that created a violation of LICENSE(o), i.e. the [i] \rightarrow [o] items of the “deep” language, are exactly the ones that speakers did the worst on. The effect of ϕ -MATCH was smaller, but also caused a noticeable

disadvantage for the “deep” language in the [o] → [i] items.

This analysis of the artificial languages provides a clear connection between the stem vowel change and the plural affix selection: Once [-im] is chosen, LICENSE(o) requires the change of stem [o] to [i] to avoid violation, and penalizes the change of stem [i] to [o]. Conversely, once [-ot] is chosen, LICENSE(o) allows changing a stem vowel from [i] to [o] with impunity, and offers no motivation to change a stem [o] to [i].

As for finding a constraint ranking for the two languages, it again emerges that the “surface” language is easier to analyze, and is thus expected to be easier to learn: In the “surface” language, nouns that have an [o] in their plural stem always select [-ot], so LICENSE(o) can be uniformly ranked over ϕ -MATCH. Nouns that have [i] in their plural stem always select [-im], which is compatible with a uniform ranking of ϕ -MATCH over * \acute{o} /HIGH. Under this view, the “surface” language is just a simpler, more extreme expression of actual Hebrew, modulo the unfamiliar stem vowel changes. The ranking arguments in (54) give rise to the simple constraint ranking in (55), which can be successfully used to provide the correct choice of plural affix for the “surface” language.

(54)

	LICENSE(o)	ϕ -MATCH	* \acute{o} /HIGH
a. agov- \acute{o} t \succ *agov- \acute{i} m	W	L	W
b. abik- \acute{i} m \succ *abik- \acute{o} t		W	L

(55) LICENSE(o) \gg ϕ -MATCH \gg * \acute{o} /HIGH

In the “deep” language, speakers cannot find a single constraint ranking for the language that uses the markedness constraints that are active in the plural allomorph selection of actual Hebrew. Since nouns with [i] in their plural stems always take [-ot], a speaker could rank * \acute{o} /HIGH over ϕ -MATCH, but that would entail selection of [-ot] for *all* nouns, contrary to overt evidence. Nouns with [o] in their plural stems always take [-im] in the “deep” language, which would imply ranking ϕ -MATCH

over LICENSE(o). This ranking leaves LICENSE(o) completely inactive in the artificial language, and attributes all of the *ot*-selection of the language to *ó/HIGH, contrary to the situation in real Hebrew, where most *ot*-selection is due to LICENSE(o). The ranking arguments for the “deep” language in (56) are inconsistent, but can give rise to a grammar via constraint cloning, as shown in (57). The nouns that have a known plural will be divided between two clones of *ó/HIGH (or equivalently, two clones of ϕ -MATCH).

(56)

	*ó/HIGH	ϕ -MATCH	LICENSE(o)
a. agov-ím \succ *agov-ót	L	W	L
b. abik-ót \succ *abik-ím	W	L	

(57) *ó/HIGH_{afiv, axis, amig, azix, adits} \gg ϕ -MATCH
 \gg *ó/HIGH_{agof, apoz, atsok, abof, alod}, LICENSE(o)

While the grammar in (57) allows the participant to correctly select a plural affix once they have heard the correct plural form, it does not allow them to generalize correctly to forms that were only given in the singular. While the nouns with [i] and the nouns with [o] are neatly divided between the clones of *ó/HIGH, they are listed under a constraint that is indifferent to the vowel of the stem, and hence this neat division cannot be reliably extended to novel items: The only constraint in this grammar that refers to the vowel of the stem, LICENSE(o), is at the bottom of the hierarchy. The only way to access information about vowel quality in (57) is to look inside the lists of items associated with clones of *ó/HIGH, which is either costly or impossible, as discussed in §3.2, and does not predict the “deep” group’s lower performance on the [i] to [o] mapping.

It is worth mentioning a different approach to stem changes, where both allomorphs of the stem are listed, as in (Kager 2009). While speakers can simply memorize the different shapes of the stems in the training stage, when they hear both the singular

and the plural, this strategy cannot extend to the plural generation stage. A speaker who heard *agív* in the singular only, and has generated the plural *agov-ót*, must have generated the stem *agov-* from the singular *agív* by changing the stem vowel; this is a logical necessity.

It is important to note that a purely serial derivation, in which the stem vowel is changed first, and then independently a plural suffix is chosen, cannot be maintained. Recall that in both groups, speakers had to apply the exact same vowel mappings, from [o] to [i] and from [i] to [o]. And yet, the speakers in the “surface” group were more successful than the speakers in the “deep” group in applying the stem vowel changes. A serial derivation that generates the change of stem vowel independently from the choice of plural allomorph makes the wrong prediction that the speakers in the two groups will do equally well on the stem vowel changes (since those were the same changes in both groups), and that the groups will only differ in the ability to select the plural suffix allomorphs. To avoid making this prediction, a serial model will have to have the ability to look back at the stem vowel change later in the derivation, when it’s time to select a plural affix, or have the ability to assess the harmony of the steps in the entire derivation, as e.g. in [Wolf \(2008\)](#).

This same objection applies to the devil’s advocate idea that the stem vowel changes were somehow done “outside” the phonology, or even “outside” the grammar. Any cognitive module that has the ability to change the stem vowels outside the phonology is left without an explanation for the differences in the groups’ ability to apply the same stem vowel changes, unless this external module can interface with the module that selects plural affixes. If both the stem vowel changes and the plural affix selection are shunted off “outside” the phonology, the account is even worse, since there are two reasons to think that the plural affixes were selected inside the native Hebrew phonology: In trials that had no vowel change, the affix selection was product-oriented in both groups (see §4.5.4), as predicted by the OT account, and the affix selection was biased towards [-im], a Hebrew-internal effect (see §4.5.3, §5.5).

A different analytic possibility that might be available to the participants in the “deep” language is to use the OCP (Obligatory Contour Principle, [Goldsmith 1976](#)) to

choose the plural allomorph that has a vowel that is not identical to the last vowel of the root. An OCP effect on vowels is observed in actual Hebrew inside roots. The data in (58), from Bolozky & Becker (2006), shows Observed/Expected values for vowels in native disyllabic roots, with actual counts in parentheses. The combination of two [o]’s and the combination of two [i]’s (in boldface) are severely underrepresented.

(58)

V1 \ V2	a	e	i	o	u
a	(607) 1.77	(126) 0.48	(292) 2.55	(169) 1.57	(101) 0.31
e	(129) 0.80	(353) 2.85	(30) 0.55	(39) 0.77	(49) 0.32
i	(171) 0.43	(39) 0.13	(5) 0.04	(110) 0.89	(865) 2.28
o	(103) 0.89	(297) 3.35	(22) 0.57	(10) 0.28	(1) 0.01
u	(58) 1.14	(2) 0.05	(8) 0.47	(7) 0.44	(5) 0.10

Extending the effect of the OCP from roots to whole words would give the participant a single grammar to derive the “deep” language. Using the OCP this way still makes the “deep” language more different from actual Hebrew than the “surface” language: In the “surface” language, the OCP is only active inside roots, like real Hebrew, while the in “deep” language, the OCP needs to apply across morpheme boundaries, unlike real Hebrew. Even with the OCP, then, the “deep” language is predicted to be harder to learn than the “surface” language.

5.2 Stem changes and allomorph selection

A question remains about the mechanism(s) that participants have used to apply vowel changes to the noun stems (other than the options suggested in §5.1). Vowel changes in paradigmatic relations are ubiquitous in Hebrew. In making verbs and deverbal nouns, speakers of Hebrew are able to impose vowel mappings on words regardless of the words’ input vowels. For example, the loanword *lúp* ‘loop’ can give rise to the verb *liplép* ‘to loop’, with nothing left of the input’s [u]. For an OT-based account of Hebrew vowel changes in verbs, see Ussishkin (2000). In nouns, however, it’s less clear that Hebrew allows arbitrary vowel changes.

The most common vowel change in nouns involves an alternation between [e] and [a], as in *mélex* ~ *melaxím* ‘king’. Other vowel alternations are much less common, such as the change from [o] to [u] or from [o] to [a], as in (14) above. All vowel changes, then, are limited to plausible phonologically-driven changes, with mid vowels either rising to their corresponding high vowels or lowering to [a], both of which can be construed as vowel reduction. Excluding the changes that go from various vowels to [a], no nouns involve a change of vowel backness or vowel rounding.

In the artificial languages, vowel changes involve backness and rounding that don’t map onto [a], and thus represent a qualitative departure from real Hebrew. Since seemingly arbitrary vowel mappings are allowed in verbs, however, there is reason to believe that speakers did not go outside their grammatical system to learn the mappings, but only outside their nominal system.

It should be noted that the stem vowel changes created what amounts to an “exchange rule” inside the stem. Exchange rules are impossible as pure phonological operations in Optimality Theory (Moreton 2004, see also Wolf 2007). Classic examples of change rules were contested, such as the Taiwanese tone circle (Myers & Tsay 2002; Myers 2006; Zhang et al. 2006) and DhoLuo voicing (Bye 2006; Pulleyblank 2006; Baerman 2007; Trommer 2007), though see Wolfe (1970); Vaux et al. (2004). However, since it is widely accepted that exchange rules are possible morphological operations, this should not be a concern for the interpretation of the artificial languages, where the stem vowel changes are part of the plural morphology.

Another perspective on the difference between the two artificial languages is offered by the phonological cycle (Chomsky & Halle 1968; Kiparsky 2000). If the theory allows the vowel change to apply independently of the addition of the plural affix, then the “surface” language applies the vowel change first and then chooses the plural affix to go with the changed vowel, while the “deep” language selects the plural affix first, and then changes the stem vowel. The “deep” language, under this view, renders the effect of LICENSE(o) opaque, since the vowels it operates over are no longer in the surface representation. In a version of Optimality Theory where morphological

and phonological operations apply one at a time, as in Wolf (2008), both languages respect LICENSE(o), but the “deep” language does so opaquely. Are opaque languages inherently more difficult to learn than transparent languages? The answer to that is not entirely clear; some interesting directions were explored in Ettliger (2009). Most known cases of opacity in the world languages, if not all, are historically innovative, suggesting that even if speakers might be biased against opacity, this bias can certainly be overcome. Additionally, children innovate opaque interactions that don’t exist in the adult language they’re learning (Smith 1973; Jesney 2007). On the other hand, transparent patterns seem to dominate in the world languages, which could be due to their ease of learning (Kiparsky 1973; Baković 2007). Even if the only difference between the two artificial languages is the transparency of the pattern, it’s not clear that the difference in difficulty that participants had is predicted.

There is reason to believe, however, that Hebrew speakers would not allow the vowel change to apply independently of the affix selection. Semantically, the vowel changes and plural affixes were associated with a single unit of meaning, namely, plurality. Even if a single morpheme is expressed in two different ways, it’s hard to see how the two changes could apply in two different levels of the cycle. Furthermore, vowel changes alone never mark plurality in actual Hebrew. Each and every plural noun in real Hebrew is marked with either [-im] or [-ot], regardless of any vowel change. This is different from the situation in Arabic, where vowel changes in the stem and concatenated plural suffixes are in complementary distribution, and each mark plurality separately.²⁴

If it is agreed that both the vowel change and the plural affix selection must happen at the same level in the cycle, then the theory of allomorph selection in Paster (2006) makes the surprising prediction that it’s the “deep” language that would be the more natural one for speakers. In this theory, allomorph selection is only allowed to refer to the shape that a stem has in the input to the current level in the cycle. In the

²⁴In paradigms of the Arabic broken plural such as *wazir* ~ *wuzara:ʔ* ‘minister’, it is plausible that *-a:ʔ* is a suffix, but it never marks the plural on its own; it always accompanies a vowel change that marks the plural. In contrast, the plural suffixes *-u:na* and *-a:t*, as in *katib* ~ *katib-u:na* ‘writer’, always mark the plural on their own, and are never accompanied by a vowel change.

“deep” language, then, the plural allomorphs harmonize with the vowel of the singular, while in the “surface” language, the plural allomorphs are chosen to go against the phonologically preferred pattern.

5.3 The limited role of phonotactics

The present analysis of the experimental results relies on the activity of two markedness constraints that are quite specific and typologically-supported: LICENSE(o), which penalizes unstressed [o]’s unless followed by a stressed [o], and * \acute{o} /HIGH, which penalizes stressed high vowels. Our analysis predicts that the “surface” language would be easier to learn than the “deep” language. One could argue, however, that the preference for the “surface” language should be stated in much more general terms, as a simple reflection of Hebrew phonotactics. In this section we show that a simple projection of Hebrew phonotactics predicts that the “surface” language is actually harder than the “deep” language.

Looking at the attested vowel combinations in the singular forms of Hebrew shows a preference for non-identical vowels. The table in (59) shows counts from Bolozky & Becker (2006) for all singular native nouns that contain the relevant vowel sequences and counts for native masculine disyllabic nouns only. Both counts show that disharmonic vowel sequences are more frequent than harmonic ones (see also 58 above).

(59)

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

Perhaps counts of vowel combinations in plural nouns are more relevant for comparing preferences that speakers make in the plurals of the artificial languages.

The table in (60) gives the counts for plurals by the final vowel of their stem, broken down by gender.

(60)	Stem-affix combination	Masculine	Feminine	Total
	...i-ot	6	1070	1076
	...o-im	527	5	532
	...i-im	437	7	444
	...o-ot	147	178	325

The totals in (60) again show a preference for disharmonic vowel sequences over harmonic ones, so if speakers are thought to select plural suffixes based on phonotactic considerations, the “deep” language is predicted to be easier than the “surface” language, contrary to fact. Even considering the masculine nouns alone makes the same wrong prediction: Since [-im] is the most frequently used affix with either stem vowel, participants would be predicted to prefer the selection of [-im] after any stem vowel, whereas in fact, speaker preferred [-im] only with a stem [i].

The experimental results cannot be reduced, then, to a mere preference for vowel patterns that are frequent in Hebrew, since speakers actively prefer the artificial language patterns that are less frequent phonotactically. Yet frequency data from real Hebrew is not ignored in the task, as it is manifested in the bias towards [-im] (see §4.5.3, §5.5). In our interpretation of the results, speakers analyze the artificial languages in terms of constraints that are active in real Hebrew. A simple projection of the phonotactics of real Hebrew onto the artificial languages, without the mediation of a grammar, makes the wrong prediction.

5.4 Learning without product-oriented generalizations

The two languages taught in this experiment were formally equally complex. The singulars and the plural stems were identical in both, and the choice of plural suffix was completely predictable from the shape of either the singular stem or the plural

stem. A learner who uses a simple information-theoretic approach should find the two languages equally hard to learn, unlike humans, who find the “surface” language easier.

The results are challenging for a source-oriented model of phonology, such as the Minimal Generalization Learner (MGL, Albright & Hayes 2002, 2003, 2006). In the MGL, the selection of the affixes is relativized to observed changes between paradigmatically related forms. In the case of Hebrew, the MGL would identify two changes: going from nothing to [im] and going from nothing to [ot]. These changes compete for the real words of Hebrew, so the addition of [im] would mis-fire with an *ot*-taker, and vice versa. This is why each change is associated with a success rate, which is the number of words it derives correctly divided by the number of words it can apply to. Simplifying the MGL results greatly, its analysis of Hebrew is seen in (61).²⁵ The addition of [im] at the end of the word has a high success rate, since most masculine nouns are *im*-takers. The addition of [ot] at the end of just any word would have a low success rate, but the addition of [ot] to a word that ends in [o] followed by a consonant would have a reasonably high success rate.

(61)	change	environment	success rate
	∅ → [im]	/ ___ #	~97%
	∅ → [ot]	/ ___ #	~3%
	∅ → [ot]	/ o C ___ #	~30%

²⁵The actual output of the MGL contains hundreds of rules, and requires some interpretation. For instance, the MGL rules don’t abstract over the root-final consonants directly, as shown simplistically in (61). Rather, the MGL creates rules that refer to each individual segment, and then gradually abstracts from them using natural classes. The picture in (61) also abstracts away from cases of vowel deletion, which cause the MGL to identify a change that is wider than the simple addition of [im] or [ot]: For example, in *zanáv* ~ *znavót* ‘tail’, the change is from [anáv] to [navót], and the suffix [ot] is not analyzed separately from the deletion of the root vowel.

The MGL simulation reported here was given the masculine native nouns from Bolozky & Becker (2006) as input, with a simple SPE-style segmental feature table. The results were then applied to a list of disyllabic novel words that varied their last vowel and consonant.

The MGL result is impressive in that it manages to extract a set of generalizations from the rather complex raw data: It identifies the suffixes, and it identifies the kind of nouns that take them. In this model, however, the similarity between the suffixes and their environment is accidental: It learns nothing about vowel harmony, and could equally well learn a language, Hebrew', where choosing [-ot] is correlated with any other phonological property of the root.

When the MGL is applied to the two artificial languages, it identifies two changes in each language, as shown in (62). The two changes have a success rate of 100% in the two languages, since the plural allomorph selection is completely regular. Crucially, these four changes are not attested in real Hebrew at all, so the two languages are equally different from real Hebrew, and are thus predicted to be equally easy or equally hard for native speakers. Due to the vowel change in the stem, the MGL can no longer separate the suffixes [im] and [ot] from the stem.

(62)	“surface” language	“deep” language
	o C → [i C im]	o C → [i C ot]
	i C → [o C ot]	i C → [o C im]

Albright & Hayes (2003) recognized this aspect of the MGL in its treatment of the vowel changes in the English past tense. English speakers use the vowel [oʊ] (as in *drove*, *rode*) to form the past tense of novel verbs, regardless of the vowel in the present tense. In real English, only the four vowels [aɪ, eɪ, i:, u:] change to [oʊ] in the past,²⁶ but speakers identify [oʊ] as a good marker of the past tense with little regard for what the present tense vowel is, and extend the use of [oʊ] to unattested vowel mappings (while still preferring mappings that resemble existing mappings). Albright & Hayes (2003) point out that a model of human behavior must include the ability to state generalizations about derived forms separately from the bases they are derived from. We claim that the use of markedness constraints, as proposed here, is suitable for doing just that.

²⁶Examples: *drive* ~ *drove*, *break* ~ *broke*, *freeze* ~ *froze*, and *choose* ~ *chose*.

While the MGL is probably the best concrete, working source-oriented learner on the market, one could imagine a more abstract view of source-oriented learning. The devil’s advocate could say that the participants in the artificial languages learned differently complex pluralization processes: In the “surface” language, the processes are /i/ → /o ot/ and /o/ → /i im/, and in the “deep” language, they are /i/ → /o im/ and /o/ → /i ot/. If the repeating vowels in the affixes of the “surface” language are easier to learn (perhaps by collapsing them to one vowel in the underlying representation), that could explain the differences between the languages in a source-oriented way. This account, while viable in principle, can hardly account for the specifics of the experimental results, where the /i/ → /o im/ change was harder to generalize than the /o/ → /i ot/ change. Furthermore, it is not clear that any existing theory on the market uses source-oriented mechanisms to account for the preference for phonologically simple affixes; rather, phonological simplicity is driven by markedness pressures on outputs.

5.5 The role of the grammar of real Hebrew

The participants’ responses in the experiment make it clear that they identified the plural affixes of the artificial language with the plural affixes of actual Hebrew. All the plural forms that participants produced contained a well-formed plural affix, either [-im] or [-ot].²⁷ Furthermore, speakers were quite successful at recognizing that the choice of affix depends on the vowels of the root, but except for one speaker, they never selected the vowels of the plural suffix independently of its consonants, but rather treated them as two whole units, [-im] and [-ot], just like in real Hebrew.

Whenever the participants produced plural forms, either repeating forms they have heard or generating plurals that they haven’t heard, they pronounced them all with final stress without fail. This indicates that the nouns of the artificial languages were not accepted as just any nouns of Hebrew, but more specifically as native nouns of

²⁷A single participant offered the following four paradigms: *amóv* ~ *amivít*, *agív* ~ *agivít*, *atóx* ~ *atixít*, and *afóts* ~ *axifóts*. The rest of this participant’s responses were unremarkable, with either [-im] or [-ot] in them.

Hebrew. With loanwords, plurals are formed without moving the stress away from the root, so a pluralized loanword will never surface with final stress.²⁸ Another aspect of the stimuli that helped make them acceptable as native nouns was their disyllabicity. In a pilot study that had mono-syllabic stimuli (e.g. *góf* instead of *agóf*), some participants generated plurals with stress on the stem, marking their treatment as loanword. In other words, participants engaged their Hebrew phonology in assigning stress to the plurals they generated; they did not simply “repeat” the stress they heard on the training items.

Finally, the preference for [-im] over [-ot] in the experiment, as discussed in §4.5.3, is the clearest indication that participants accepted the artificial nouns as nouns of Hebrew, and that frequency effects from real Hebrew influenced the participants’ behavior. In the artificial languages, [-im] and [-ot] were equally represented, so the higher frequency of [-im] responses must be attributed to the influence of real Hebrew. It is very likely that speakers accepted the artificial nouns as masculine, especially given the numerals and adjectives that agreed in gender with those nouns in the various frame sentences. However, [-im] is more frequent than [-ot] in real Hebrew overall (since masculine nouns are more than twice as common as feminine nouns), so speakers can show a bias for [-im] even if they ignore the cues for masculine gender in the experiment.

The bias for [-im] is left unexplained in “task-specific” interpretations of the experimental results. If participants were using some Hebrew-external, or ever language-external strategy in learning the artificial language they were faced with, they would have no access to the bias for [-im] in Hebrew.

²⁸Some nouns that are etymologically borrowed were fully nativized and now get final stress in the plural, e.g. *balon-im* ‘balloon’. These nouns are all di-syllabic, just like the majority of native Hebrew nouns (Becker 2003).

5.6 Source-oriented generalizations?

The aim of this work is to highlight the importance of product-oriented generalizations in phonology, yet it is obviously still the case the languages have source-oriented generalizations. Even the Hebrew plural affix, which we have shown to be subject to a product-oriented generalization, is also subject to a source-oriented generalization: Loanwords that end in [a] in the singular invariably take the plural [-ot], regardless of their gender, as noted in (3-4) and (18) above. In other words, the choice of plural affix must also be sensitive to some aspect of the input to the derivation.

In Optimality Theory, there are two ways in which an output can be sensitive to the input: (a) The activity of faithfulness constraints can force identity between an input and an output, and (b) some mechanism of opacity can give rise to structure that depends phonologically on some aspect of the input, e.g. in the Tiberian Hebrew /defʔ/ → [defe], the second [e] in the output is not present due to faithfulness, but its presence depends on the presence of the glottal stop in the input (McCarthy 2007).

The source-oriented selection of the plural affix in loanwords can be due to faithfulness, if one accepts Faust's (2008) analysis in which the plural affix [-ot] phonologically contains the feminine suffix [-a]. The alternative proposed in (18) is that an unstressed final [a] is more aggressively parsed as a feminine suffix than a stressed final [a], in which case it is the morphological structure imposed on the input that is determining the selection of [-ot].

6 Conclusions

This paper examined the distribution of the two plural suffixes [-im] and [-ot] on Hebrew nouns. The lexicon study showed a connection between having [o] in the root and a preference for selecting [-ot], a preference that speakers extend productively to novel nouns. We offered an OT-based analysis of plural allomorph selection in Hebrew, which relied on a mechanism of constraint cloning to build

lexical trends into the grammar, and project those trends onto novel nouns. In the analysis, allomorph selection was understood to be without faithfulness cost, and therefore only markedness constraints were involved.

Since markedness constraints only assess output forms, they have no access to underlying representations or to paradigmatically related forms. In deriving Hebrew plurals, the selection of [-ot] is predicted to correlate with the presence of [o] in the plural stem, regardless of the vowels of the singular. Since in real Hebrew, the presence of [o] in a plural stem always corresponds to the presence of [o] in the singular, the prediction cannot be tested on the real words of the language.

To test whether the selection of the plural affix is sensitive to the vowels of the input or the vowels of the output, we created a pair of artificial languages, where a singular [i] alternates with a plural [o] and vice versa. In one language, the selection of [-ot] correlated with the presence of [o] in the plural stem, and in the other language, the selection of [-ot] correlated with the presence of [o] in the singular stem. As predicted, speakers were significantly more successful at generalizing the language where the selection of [-ot] correlated with the presence of [o] in the plural stem.

The artificial languages were designed and presented as languages that are just like real Hebrew, with the only difference being the vowel changes from [o] to [i] and vice versa, which don't occur in real Hebrew. To insure that singulars and plurals are correctly paired, participants never heard or produced a plural form without hearing or producing its singular in the same trial. Indeed, the experimental results show that the participants accepted the artificial nouns as native masculine nouns of Hebrew, evidenced by their generation of plural forms with final stress and a bias towards [-im].

The prediction of the markedness-based analysis, which favors the language that pairs [-ot] with plural [o]'s, was contrasted with an MGL-based analysis (Albright & Hayes 2002, 2003, 2006), which predicts that the two languages would be equally different from Hebrew, and thus equally difficult for Hebrew speakers. The point is applicable more generally to any analysis that relies on source-oriented mechanisms: since the two artificial languages are formally equally complex, with the same

amount of information in them, there is no a priori reason to prefer generalizations about output forms over generalizations about input forms. Additionally, we have shown that the experimental results cannot be reduced to a mere phonotactic preference, since the phonotactics of real Hebrew prefer the pairing of non-identical vowels over identical vowels.

In real Hebrew, the connection between [o] in the stem and the selection of [-ot] is equally reliable when stated over singulars or over plurals: One can say that singulars with [o] often choose [-ot], or one can say that plural stems with [o] often choose [-ot]. And yet, the results of the artificial language experiment show that speakers are biased to choose the plural-based interpretation over the singular-based interpretation, even in the absence of evidence for this bias in the language. This bias follows naturally from the analysis we offer, which attributes allomorph selection to the activity of universal markedness constraints, as is standardly assumed in the OT literature.

References

- Aharoni, Moti (2007). Different plural forms (tsurot rabim fonot). http://tora.us.fm/tnk1/messages/ljon_cwrt_yxyd_wcwr_rbym_0.html.
- 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 computational experimental 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.

- Anttila, Arto (1997). *Variation in Finnish phonology and morphology*. Ph.D. dissertation, Stanford University, Stanford, CA.
- Aronoff, Mark (1994). *Morphology by itself*. MIT Press.
- Baayen, R. Harald (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press.
- Baerman, Matthew (2007). Morphological reversals. *Journal of Linguistics* 43. 33–61.
- Baković, Eric (2007). A revised typology of opaque generalizations. *Phonology* 24. 217–259.
- Barnes, Jonathan (2006). *Strength and Weakness at the Interface: Positional Neutralization in Phonetics and Phonology*. Berlin/New York: Mouton de Gruyter.
- Bat-El, Outi (1993). Parasitic metrification in the modern hebrew stress system. *The Linguistic Review* 10. 189–210.
- Bat-El, Outi (2008). A gap in the feminine paradigm of Hebrew: A consequence of identity avoidance in the suffix domain. In Curt Rice (ed.) *When Nothing Wins: Modelling Ungrammaticality in OT*, London: Equinox.
- Becker, Michael (2003). Lexical statification of Hebrew: The disyllabic maximum. In Yehuda N. Falk (ed.) *Proceedings of the Israel Association for Theoretical Linguistics* 19.
- Becker, Michael (2009). *Phonological Trends in the Lexicon: The Role of Constraints*. Ph.D. dissertation, UMass Amherst.
- Becker, Michael, Nihan Ketrez & Andrew Nevins (2008). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. ROA-1001.
- Beckman, Jill (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14. 1–46.
- Beckman, Jill (1998). *Positional Faithfulness*. Ph.D. dissertation, UMass, Amherst.

- 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.
- Boersma, Paul & David Weenink (2008). Praat: Doing phonetics by computer (version 5.0.24). [Computer program].
- Bolozky, Shmuel & Michael Becker (2006). Living Lexicon of Hebrew Nouns. Ms. UMass Amherst.
- Bybee, Joan (2001). *Phonology and language use*. Cambridge University Press.
- Bybee, Joan & Carol Lynn Moder (1983). Morphological classes as natural categories. *Language* . 251–270.
- Bybee, Joan & Jean Newman (1995). Are stem changes as natural as affixes? *Linguistics* 33. 633–654.
- Bybee, Joan & Dan Slobin (1982). Rules and schemas in the development and use of the english past tense. *Language* 58. 265–289.
- Bye, Patrik (2006). Eliminating exchange rules in dholuo. Ms. TromsøUniversity.
- Cassimjee, Farida & Charles Kisseberth (1998). Optimal domains theory and bantu tonology: A case study from isixhosa and shingazidja. In Larry Hyman & Charles Kisseberth (eds.) *Theoretical Aspects of Bantu Tone*, Stanford: CSLI. 33–132.
- Chomsky, Noam & Morris Halle (1968). *The Sound Pattern of English*. the MIT press.
- Coetzee, Andries W. (2008). Grammaticality and ungrammaticality in phonology. *Language* 84. 218–257.
- Coetzee, Andries W. & Joe Pater (2008). Weighted constraints and gradient restrictions on place co-occurrence in muna and arabic. *Natural Language and Linguistic Theory* 26. 289–337.

- Crosswhite, Kathryn (2001). *Vowel reduction in Optimality Theory*. New York: Routledge.
- de Lacy, Paul (1997). Dissimilation as local conjunction. In Kiyomi Kusumoto (ed.) *Proceedings of NELS 27*. Amherst, MA: GLSA, 17–31.
- de Lacy, Paul (2004). Markedness conflation in optimality theory. *Phonology* 21. 145–199.
- Ernestus, Miriam & R. Harald Baayen (2003). Predicting the Unpredictable: Interpreting Neutralized Segments in Dutch. *Language* 79. 5–38.
- Esper, Erwin Allen (1925). *A Technique for the Experimental Investigation of Associative Interference in Artificial Language Material*. Linguistic Society of America.
- Ettlinger, Marc (2009). *Input driven opacity*. Ph.D. dissertation, UC Berkeley.
- Even-Shoshan, Avraham (1966). *The New Dictionary (hamilón hehadáf)*. Jerusalem: Kiryat Sefer.
- Faust, Noam (2008). In a position to spell out: The syntactic structure of vocalic patterns in Hebrew. Ms. University of Paris VII.
- Finley, Sara & William Badecker (2008). Substantive biases for vowel harmony languages. In J. Bishop (ed.) *Proceedings of WCCFL 27*. 168–176.
- Finley, Sara & William Badecker (in press). Artificial language learning and feature-based generalization. *Journal of Memory and Language* .
- Glinert, Lewis (1989). *The grammar of Modern Hebrew*. Cambridge University Press.
- Goldsmith, John (1976). *Autosegmental Phonology*. Ph.D. dissertation, MIT.
- Hargus, Sharon (1997). Augmentation as affixation in athabaskan languages. *Phonology* 14. 177–220.
- Hayes, Bruce & Zsuzsa Londe (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology* 23. 59–104.

- Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.
- Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Londe (to appear). Natural and unnatural constraints in hungarian vowel harmony. *Language* .
- Inkelas, Sharon, Aylin Kuntay, John Lowe, Orhan Orgun & Ronald Sprouse (2000). Turkish electronic living lexicon (TELL). Website, <http://socrates.berkeley.edu:7037/>.
- Jesney, Karen (2007). Child chain shifts as faithfulness to input prominence. In Alyona Belikova, Luisa Meroni & Mari Umeda (eds.) *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition – North America*. Somerville, MA: Cascadilla Proceedings Project, 188–199.
- Kager, René (1996). On affix allomorphy and syllable counting. In U. Kleinhenz (ed.) *interfaces in phonology*, Berlin: Akademie Verlag. 155–171.
- Kager, René (2009). Lexical irregularity and the typology of contrast. In Kristin Hanson & Sharon Inkelas (eds.) *The Nature of the Word: Essays in Honor of Paul Kiparsky*, Cambridge, MA: MIT Press. 397–432.
- Kenstowicz, Michael (1997). Quality-Sensitive Stress. *Rivista di Linguistica* 9.1. 157–187.
- Kiparsky, Paul (1973). ‘Elsewhere’ in Phonology. In *A Festschrift for Morris Halle*, Holt, Rinehart and Winston. 93–106.
- Kiparsky, Paul (2000). Opacity and cyclicity. *The Linguistic Review* 17. 351–367.
- Mahanta, Shakuntala (2007). *Directionality and Locality in Vowel Harmony*. Ph.D. dissertation, Utrecht University.
- Mascaró, Joan (1996). External allomorphy as emergence of the unmarked. In Jacques Durand & Bernard Laks (eds.) *Current Trends in Phonology: Models and Methods*, European Studies Research Institute, University of Salford. 473–483.

- Maye, Jessica, Richard N. Aslin & Michael K. Tanenhaus (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science* 32. 543–562.
- McCarthy, John J. (2004). Headed spans and autosegmental spreading. Ms., University of Massachusetts, Amherst. [Available as ROA-685].
- McCarthy, John J. (2007). *Hidden Generalizations: Phonological Opacity in Optimality Theory*. London: Equinox Publishing Company.
- Mester, Armin (1994). The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12. 1–61.
- Moreton, Elliott (2004). Non-computable functions in optimality theory. In John J. McCarthy (ed.) *Optimality Theory in Phonology*, Blackwell. 141–163.
- Myers, James (2006). Tone circles and chance. Ms. National Chung Cheng University, available as ROA-843.
- Myers, James & Jane Tsay (2002). Neutralization in taiwanese tone sandhi. Ms. National Chung Cheng University.
- Paster, Mary (2006). *Phonological Conditions on Affixation*. Ph.D. dissertation, University of California, Berkeley.
- 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.
- Pater, Joe & Anne-Michelle Tessier (2005). Phonotactics and alternations: Testing the connection with artificial language learning. In Kathryn Flack & Shigeto Kawahara (eds.) *University of Massachusetts Occasional Papers in Linguistics 31*, Amherst, MA: GLSA. 1–16.

- Peperkamp, Sharon, Katrin Skoruppa & Emmanuel Dupoux (2006). The role of phonetic naturalness in phonological rule acquisition. In D. Bamman, T. Magnitskaia & C Zaller (eds.) *Proceedings of the 30 Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 464–475.
- Prince, Alan (2002). Arguing optimality. ROA 562-1102.
- Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. [ROA-537].
- Pulleyblank, Douglas (2006). Minimizing ug: Constraints upon constraints. In Donald Baumer, David Montero & Michael Scanlon (eds.) *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 15–39.
- Pycha, Anne, Pawel Nowak, Eurie Shin & Ryan Shosted (2003). Phonological Rule-Learning and Its Implications for a Theory of Vowel Harmony. In Gina Garding & Mimura Tsujimura (eds.) *WCCFL 22*, Cascadilla Press. 423–435.
- Pycha, Anne, Eurie Shin & Ryan Shosted (2006). Directionality of assimilation in consonant clusters: An experimental approach. In *UC Berkeley Phonology Lab Annual Report*. 152–166.
- R Development Core Team (2007). R: A language and environment for statistical computing. Vienna, Austria. (<http://www.R-project.org>).
- Schane, Sanford A., Bernard H. Tranel & Harlan Lane (1974). On the psychological reality of a natural rule of syllable structure. *Cognition* 3. 351–358.
- Skoruppa, Katrin, Anna Lambrechts & Sharon Peperkamp (in press). The role of phonetic distance in the acquisition of phonological alternations. In *Proceedings of the 39th North Eastern Linguistics Conference*.
- Skoruppa, Katrin & Sharon Peperkamp (2006). Mécanismes d’acquisition phonologique précoce. Slides.

- Smith, Neilson V. (1973). *The Acquisition of Phonology: A Case Study*. Cambridge University Press.
- 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 (1998). Learnability in optimality theory. *Linguistic Inquiry* 29. 229–268.
- Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Tessier, Anne-Michelle (2008). Children's exceptional words: Lexical challenges from the time-course of L1 phonological acquisition. Talk given at the University of Western Ontario.
- Trommer, Jochen (2007). Voicing and polarity in luo. Ms. Leipzig University.
- Trommer, Jochen (2008). Syllable-counting allomorphy by indexed constraints. Talk given at OCP 5.
- Tucker, Matthew A. (2009). The root-and-prosody approach to arabic verbs. ROA-1030.
- Ussishkin, Adam (2000). *The Emergence of Fixed Prosody*. Ph.D. dissertation, UC Santa Cruz.
- Vaux, Bert, Andrew Nevins & Justin Fitzpatrick (2004). Exchange rules and Feature-Value Variables. Paper presented at the 3rd North American Phonology Conference, Concordia University.
- Walker, Rachel (2006). Long-distance metaphony: A generalized licensing proposal. Talk given at the PhonologyFest Workshop, Indiana University.
- Wilson, Colin (2003). Experimental investigation of phonological naturalness. In G. Garding & M. Tsujimura (eds.) *WCCFL 22 Proceedings*. Somerville, MA: Cascadilla Press, 101–114.

- Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* 30. 945–982.
- Wolf, Matthew (2007). For an autosegmental theory of mutation. In Leah Bateman, Michael O’Keefe, Ehren Reilly & Adam Werle (eds.) *UMOP 32: Papers in Optimality Theory III*, Amherst, MA: GLSA. 315–404.
- Wolf, Matthew (2008). *Optimal Interleaving: Serial Phonology-Morphology Interaction in a Constraint-Based Model*. Ph.D. dissertation, UMass Amherst.
- Wolfe, Patricia (1970). Some theoretical questions on the historical english vowel shift. *Papers in Linguistics* 3. 221–235.
- Zhang, Jie, Yuwen Lai & Craig Turnbull-Sailor (2006). Wug-testing the ”tone circle” in taiwanese. In David Montero Donald Baumer & Michael Scanlon (eds.) *Proceedings of the 25th West Coast Conference on Formal Linguistics*. 453–461.
- Zuraw, Kie (2000). *Patterned Exceptions in Phonology*. Ph.D. dissertation, UCLA.