Modeling Acceptability of Variation in Modern Hebrew

MICHAL TEMKIN MARTINEZ
Boise State University
michaltmartinez@boisestate.edu

Abstract: Modern Hebrew spirantization (MHS) is a variable process with many exceptions. This paper reports on an experiment testing the acceptability of variation in alternating and exceptional segments in MHS. The results show that variation is acceptable in both alternating and exceptional segments, but is significantly more acceptable in alternating segments than exceptional ones. This suggests that, despite the ubiquity of variation, speakers still distinguish between alternating and exceptional segments categorically. Consequently, an Optimality Theoretic (OT) analysis, combining stochastic constraint ranking (Hayes & MacEachern 1998, Boersma 1998, Zuraw 2000, Boersma & Hayes 2001) and set-based indexation (Pater 2000), is used to model the results. The combined model presented here accounts for the experimental data, allowing for both variation and exceptionality within a single phenomenon across all participants. Preliminary analysis of within-speaker variation demonstrates that the combined model can also account for individuals' grammars across word position and segment type.

0. Introduction

This paper presents an Optimality Theoretic (OT) model of the acquisition of Modern Hebrew Spirantization, a variable phenomenon. The stop/fricative pairs participating in the alternation are the labials [p]~[f] and [b]~[v], as well as the voiceless dorsals [k]~[χ]. Due to historical sound mergers, there are many lexical exceptions to spirantization in Modern Hebrew (words containing non-alternating instances of the sounds listed above). In addition to these exceptions, there is a high level of variation in segments that do alternate. The analysis presented in this paper is based on the results of an experimental rating task showing that variation is acceptable, in both alternating and exceptional segments. I argue that to model this gradience in the two types of segments, it is necessary to combine two mechanisms: stochastic constraint ranking (Boersma 1998, Hayes & MacEachern 1998, Zuraw 2000, Boersma & Hayes 2001) and set-based indexation (Pater 2000).

1. Modern Hebrew Spirantization

In Modern Hebrew, the stops [p], [b], and [k] alternate with their fricative counterparts [f], [v], and [χ], respectively. The stops occur in word-initial and post-consonantal position, while the fricatives occur post-vocalically. This is illustrated in the verbal paradigms in Table 1.

Table 1: Spirantization distribution in Modern Hebrew

	Root	**Infinitive**	**Uninflected**	**Gloss**
/**p**/ ~ **[f]**	/**p**gʃ/	[li**f**goʃ]	[**p**agaʃ]	‘to meet’
/**b**/ ~ **[v]**	/**b**gd/	[li**v**god]	[**b**agad]	‘to betray’
/**k**/ ~ **[χ]**	/**k**tb/	[li**χ**tov]	[**k**atav]	‘to write’

Exceptions to spirantization are cases of non-alternation of the sounds [p], [f], [b], [v], [k] and [χ], such that they surface in unexpected environments. Such non-alternating segments are seen in the verbal paradigms in Table 2, where stops may surface post-vocalically and fricatives may surface in word-initial position.

Table 2: Exceptions to spirantization in Modern Hebrew (in underlined words)

	Root	**Uninflected**	**Infinitive**	**Gloss**
/k/	/**k**rʔ/	[**k**ara]	[<u>li**k**ro</u>]	‘to read’
/v/	/**v**tr/	[<u>**v**iter</u>]	[le**v**ater]	‘to give up’

In addition to these exceptions, high levels of variation have been reported in the alternating segments (Adam 2002, Temkin Martínez 2010). This variation includes the surfacing of stops and fricatives in contexts not predicted by the spirantization distribution by segments that normally *do* conform to it. This is illustrated in Table 3.

Table 3: Variation in Modern Hebrew spirantization

	Root	**Expected**	**Acceptable Variant**	**Gloss**
/**p**/ ~ **[f]**	/**p**gʃ/	[**p**agaʃ]	[**f**agaʃ]	‘met’
/**b**/ ~ **[v]**	/k**b**r/	[jik**b**or]	[jik**v**or]	‘will bury’
/**k**/ ~ **[χ]**	/**k**sh/	[je**k**ase]	[je**χ**ase]	‘will cover’

2. Experimental Rating Task

2.1. Methods

The analysis presented in this paper is based on the results of an experiment conducted to quantify the acceptability of variation in Modern Hebrew spirantization. Seventy-four native speakers of Hebrew between the ages of 19 and 40 participated in the study. A total of 42 roots were used to form the experimental stimuli. Twenty-four roots contained an alternating segment, twelve contained an exceptional segment, and six contained two target segments. Each of the roots was presented in two forms: the uninflected form and the infinitive. To determine the acceptability of variation, each of the target words was presented in its expected and variant form for a total of 204 target words. Examples of target words from the experiment are given in Table 4.

Table 4: Expected and variant forms in the spirantization distribution

Pair	**Root**	**Uninflected**		**Infinitive**		**Gloss**
		Expected word-initial stop	**Variant** word-initial fricative	**Expected** post-vocalic fricative	**Variant** post-vocalic stop	
/p/~[f]	/**p**rs/	[**p**aras]	[**f**aras]	[li**f**ros]	[li**p**ros]	'to spread'
/b/~[v]	/**b**nh/	[**b**ana]	[**v**ana]	[li**v**not]	[li**b**not]	'to build'
/k/~[χ]	/**k**tb/	[**k**atav]	[**χ**atav]	[li**χ**tov]	[li**k**tov]	'to write'

Each of the 204 target verbs was inserted into carrier sentences containing the verb in phrase-medial position. All sentences were identical in syllable count, and each ended with a noun which complemented the target verb semantically. A sample sentence is given in (1).

(1)	[amru	li	ʃedaniel	(<u>target word</u>)	le/be/me/et ________]
	told	to-me	that-Daniel		to/in/from/the

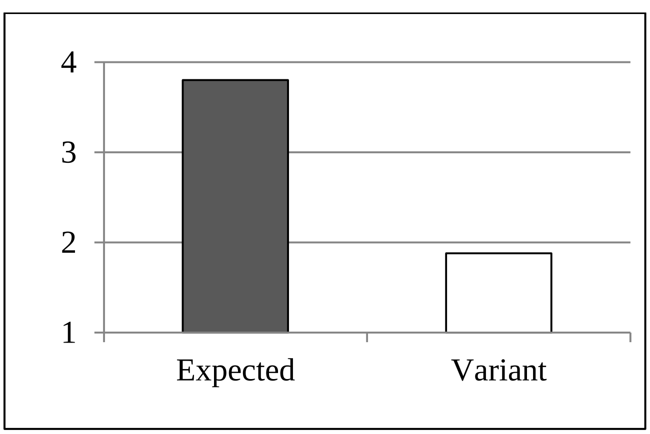
'I've been told that Daniel (target word) to/in/from/the …'
(e.g. 'I've been told that Daniel <u>built</u> the <u>hut</u>.')

The sentences were recorded by a native speaker at a natural speaking rate, and placed in an online experiment using a .php script. Participants were presented with each sentence auditorily and instructed to pay special attention to the target verb, rating the naturalness of its pronunciation on a 4-point scale.

2.2. Results

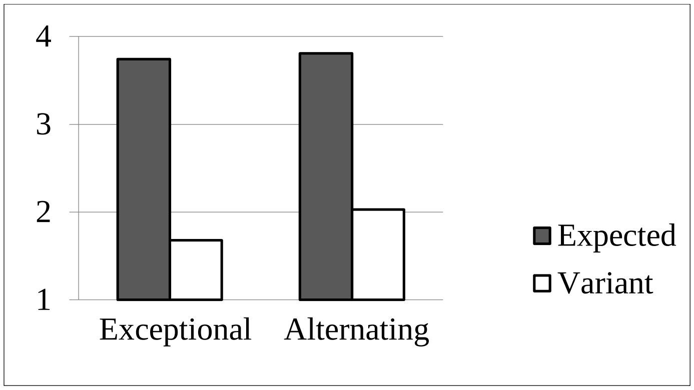
Participant responses were converted to numerical values with the highest value (4) corresponding to 'very natural' and the lowest value (1) corresponding to 'very unnatural.' Although an ANOVA showed that variation was acceptable overall, there was a significant preference for the expected form across conditions, with a main effect of *allophone*, or whether the segment was the expected or variant form ($F(1, 73) = 886.521$, $p < .001$). This is seen in Figure 1.

Figure 1: Ratings of expected and variant forms



While variation was also acceptable in alternating segments, it was deemed significantly less natural than in exceptional segments, with a main effect of *type* ($F(1, 73) = 80.073$, $p < .001$), and an interaction between *type* and *allophone* ($F(1, 73) = 18.707$, $p < .001$). This is seen in Figure 2.

Figure 2: Ratings of exceptional and alternating segments



For the analysis, results for each variant were entered as input frequencies. To calculate the input frequency for each variant of a given verb, all 'natural' ratings (a selection of '3' or '4' on the 4-point scale) were counted for all variants of the word, and each variant's rating was calculated as a percentage of the 'natural' ratings for the word. This is seen in Table 5 for /bikeʃ/ 'he asked for.'

Table 5: Calculation for input frequency based on experimental results

Token	Number of 'natural' ratings	Input frequency (percentage)
bikeʃ	37	67%
biχeʃ	3	5%
vikeʃ	13	24%
viχeʃ	2	4%
Total	55	100%

Since each target word had a different number of 'natural' ratings, and each variant could have a different input frequency, the model used in the OT analysis must be able to handle gradience in its analysis of variation.

3. OT Analysis

3.1. Alternating Segments

In order to account for the complementary distribution found in alternating segments participating in spirantization (prior to considering variation), I propose using a contextual markedness constraint banning post-vocalic stops, the context-free markedness constraints banning fricatives and stops, and a faithfulness constraint for [continuant], the feature distinguishing stops and fricatives. The ranking is such that the constraint banning post-vocalic stops dominates the faithfulness constraint, resulting in the alternation between stops and fricatives. Ranking the faithfulness constraint below markedness ensures that non-alternation only occurs in positions other than post-vocalic. The constraints, their definitions, and the proposed ranking for alternating segments are given in (2).

(2) Constraints for the analysis of alternation

***V-STOP**	Post-vocalic stops are prohibited.
***[+cont, -sib]**	Non-sibilant fricatives are prohibited.
***STOP**	Stops are prohibited.
IDENT-IO[cont]	Input-output correspondents are identical in [±cont].

Constraint ranking: *V-STOP » *[+cont, -sib] » IDENT-IO[cont], *STOP

Applying the constraints above to a root containing alternating segments, we are able to account for post-vocalic fricatives as well as word-initial and post-vocalic stops. In the tableau in (3), we see this in the root /bkh/ 'to cry,' in which both /b/ and /k/ are alternating segments. In the infinitive, the /b/ occurs post-vocalically, resulting in a [v], and the /k/ in post-consonantal position, surfacing as a [k]. In the uninflected form, the /b/ occurs word-initially, surfacing as a [b], while the /k/ occurs post-vocalically, resulting in a [χ].

(3) Alternation in the root /bkh/ (both segments alternate)

Input	Output	*V-STOP	*[+cont, -sib]	IDENT-IO[cont]	*STOP
/bkh/ + (inf.) 'to cry'	☞ a. livkot		*	*	**
	b. libkot	*!			***
	c. livχot		**!	**	*
	d. libχot	*!	*	*	**
/bkh/ + (3p.sg.m.past) 'he cried'	☞ a. baχa		*	*	*
	b. baka	*!			**
	c. vaχa		**!	**	
	d. vaka	*!	*	*	*

3.2 Exceptional Segments

Since exceptional segments (those in Table 2) are phonetically indistinguishable from those that normally alternate according to the spirantization distribution, it is crucial to distinguish the two in the analysis. An OT analysis that allows for this distinction is set-indexation (Pater 2000). Using set-indexation, exceptional segments are indexed to a special set for which additional constraints are designated (in addition to those already being used), with the indexed constraints applying only to segments indexed to the same set. The additional constraint in this case is a clone of the faithfulness constraint IDENT-IO[cont]. Once cloned and indexed to the exceptional set (Set 1), it is ranked above the markedness constraints relevant to spirantization in the constraint hierarchy. This ranking makes the segments in Set 1 essentially immune to the alternation driven by the markedness constraints, since any alternation would violate this highly ranked faithfulness constraint. This is illustrated in Figure 3.

Figure 3: Schema for exceptionality and alternation using a set-based approach

IDENT-IO[cont]$_1$ »	**Markedness constraints** » *V-STOP, *[+cont, -sib]	**IDENT-IO[cont]**
Prohibits alternation in exceptional segments		*Determines the distribution of stops and fricatives in alternating segments*

By indexing exceptional segments to IDENT-IO[cont]$_1$, words containing non-alternating segments can be accounted for successfully. In (4), the root /bχr/ contains both an alternating segment (/b/) and an exceptional segment (/χ/). Indexing the /χ/ to Set 1 accounts for the lack of alternation in this segment and its surfacing as a fricative [χ] in post-consonantal position. The alternating /b/, not indexed to Set 1, is free to alternate between [b] in word-initial position and [v] post-vocalically.

(4) Words containing an exceptional segment (indexed to Set 1)

Input	Output	IDENT-IO[cont]$_1$	*V-STOP	*[+cont, -sib]	IDENT-IO[cont]	*STOP
/bχ$_1$r/ + (inf.) ‘to choose’	☞ a. livχ$_1$or			**	*	
	b. libχ$_1$or		*!	*		*
	c. livk$_1$or	*!		*	**	*
	d. libk$_1$or	*!	*		*	**
/bχ$_1$r/ + (3p.sg.m.past) ‘he chose’	☞ a. baχ$_1$ar			*		*
	b. bak$_1$ar	*!	*		*	**
	c. vaχ$_1$ar			**!	*	
	d. vak$_1$ar	*!	*	*	**	*

3.3 Variation

With the results of the experiment described in Section 2 showing gradience, the analysis for variation must reflect the lack of free variation. Gradience in variation is accounted for by implementing Stochastic OT (Boersma 1998; Boersma & Hayes 2001; Hayes & Londe 2006; Hayes & MacEachern 1998; Zuraw 2000). Utilizing Stochastic OT, the input for each token is the percentage of times the token was rated as ‘natural’ by the participants. Once frequencies are entered into the algorithm, it cycles through the grammar (input/output pairs, candidate frequencies, constraint violations) and assigns constraints values, determining their ranking. The stochastic grammar then attempts to match the candidate frequencies (acceptability ratings from the experiment) by determining the probability of different constraint rankings and assigning ranking values to each constraint, given all inputs. These ranking values affect the frequency with which each constraint in the hierarchy outranks other constraints.

In the tableaux in (5), we see the two rankings that allow for the variants of the form /bitel/ 'he cancelled.' In the experiment, the expected form [bitel] was rated as 'natural' 77.1% of the time, while the variant [vitel] was rated as 'natural' 22.9% of the time. The ranking of *[+cont, -sib] and *STOP is the determining factor in selecting between these variants. In Stochastic OT, then, the algorithm must allow for the rankings of these two constraints to alternate in order for each variant to surface. In fact, given the input frequencies entered for all tokens used in the experiment, the algorithm allows for the expected form to surface 74.7% of the time, and the variant form 25.3% in final grammar.

(5) Rankings for [bitel] (expected, 77.1%) ~ [vitel] (variant, 22.9%):

[bitel] = *[+cont, -sib] » *STOP (occurs 74.7% in grammar)

/btl/ + 3p.sg.m.past 'he cancelled'	*V-STOP	*[+cont, -sib]	IDENT-IO[cont]	*STOP
☞ a. bitel				*
b. vitel		*!	*	

[vitel] = *STOP » *[+cont, -sib] (occurs 25.3% in grammar)

/btl/ + 3p.sg.m.past 'he cancelled'	*V-STOP	*STOP	*[+cont, -sib]	IDENT-IO[cont]
a. bitel		*!		
☞ b. vitel			*	*

Similarly, the tableaux in (6) show that in order for both variants of /mevatel/ 'cancels' to surface, the ranking between *V-STOP and *[+cont, -sib] must allow each to outrank the other some of the time. In the experiment, the expected form [mevatel] was rated as 'natural' 72% of the time, while the variant [mebatel] was 28% of the time. In this case, the ranking of *[+cont, -sib] and *V-STOP is the determining factor in selecting between these variants.

(6) Ranking for [mevatel] (expected, 72%) ~ [mebatel] (variant, 28%) 'cancels:'

[mevatel] = *V-STOP » *[+cont, -sib] (occurs 74.7% in grammar)

/btl/ + sg.m.pres. 'cancels'	*V-STOP	*[+cont, -sib]	IDENT-IO[cont]	*STOP
☞ a. mevatel		*	*	
b. mebatel	*!			*

[mebatel] = *[+cont, -sib] » *V-STOP (occurs 25.3% in grammar)

/btl/ + sg.m.pres. 'cancels'	*[+cont, -sib]	*V-STOP	IDENT-IO[cont]	*STOP
a. mevatel	*!		*	
☞ b. mebatel		*		*

Note that, as was the case with the variants in (5), the expected form in (6) surfaces 74.7% of the time (though the input frequency was 72%), and the variant form 25.3% in final grammar (though the input frequency was 28%). This, again, is a function of the algorithm and its consideration for the frequencies of all tokens included in the experiment.

3.4 The Combined Model

A need for a combined model arises when we examine 'hybrid' words, which contain both an alternating and an exceptional segment. When taking into consideration the frequency of variation deemed acceptable by participants in hybrid words, we see that neither set-indexation nor stochastic constraint ranking can account for both variation and exceptionality on their own.

Using only set-indexation, we are able to account for the alternation of only one segment in a word, but variation can only be accounted for through unranked constraints, resulting in free variation. Since, in most variant pairs, there was a significant preference for one variant over the other, using only set-indexation results in disproportionate frequencies of variation. An example of this is in (7) where the candidates a. and b. for /kafa/ 'he froze' can be accounted for through the unranking of *V-Stop and *[+cont, -sib], resulting in the selection of either. However, this fails to account for participants' preferring candidate (7a) 90% of the time.

(7) [kafa] (expected, 90%) ~ [kapa] (variant, 10%) 'he froze'

/$k_1$ph/ + 3p.sg.m.past 'he froze'	IDENT-IO[cont]$_1$	*V-STOP	*[+cont, -sib]	*STOP	IDENT-IO[cont]
☞ a. $k_1$afa (0.9)			*	*	*
☞ b. $k_1$apa (0.1)		*		**	
c. $\chi_1$afa	*!		**		**
d. $\chi_1$apa	*!		*	*	*

Similarly, using only stochastic constraint ranking, we are unable to account for the distinction between alternating and exceptional segments. Rather than considering alternating and exceptional segments as different types of segments, both are treated as alternating segments. Recall that, in the experiment, participants found variation in exceptional segments significantly less acceptable that variation in alternating segments. Using stochastic OT without set-indexation, the algorithm averages the rate of acceptability across all segments, increasing the discrepancy between input and generated frequencies. This is seen in Table 5, where the algorithm allows for [meχajem], a variant that participants never rated as natural, to be generated 59% of the time, as well as blocking [mevaker], a variant of a hybrid root rated natural 72% of the time, from ever being generated, the latter being blocked because of the presence of both a stop and a fricative in the same position.

Table 5: Mismatched input and generated frequencies with absence of sets

	Target	**Input Frequency**	**Generated Frequency**
Alternating	[ba**k**a] 'he cried'	0.07	0.40
Exceptional	[me**χ**ajem] 'fulfills'	0.00	0.59
Hybrid	[me**v**a**k**er] 'visits'	0.72	0.00

Additionally, in the absence of sets, hybrid roots and roots containing two alternating segments are treated as equals. This is seen in Table 6, where variants of the hybrid root /kfʔ/ 'to freeze' and the root /kfh/ 'to force,' containing two alternating segments, generate the same frequencies despite significantly different input frequencies. This discrepancy is caused by the status of /k/ in each of the roots: in the hybrid root /kfʔ/ the /k/ is exceptional, whereas it is alternating in /kph/. This means that the variants containing [χ] are rated as natural less frequently in the hybrid root. However, since set-indexation does not distinguish the two instances of /k/, the two variants with the lowest input frequencies for the hybrid are generated 25% and 40% of the time according to the algorithm.

Table 6: /kfʔ/ ‘to freeze’ and /kfh/ ‘to force’ -- same generated frequencies for hybrids and for roots containing two alternating segments

Target (hybrid root)	**Input Frequency**	**Target (2 alternating)**	**Input Frequency**	**Generated Frequency**
[likfo]	0.452	[likfot]	0.194	**0.000**
[likpo]	0.435	[likpot]	0.104	**0.339**
[liχfo]	0.048	[liχfot]	0.194	**0.257**
[liχpo]	0.065	[liχpot]	0.507	**0.404**

By combining set-indexation and stochastic constraint rankings, we are better able to account for the gradience in variation identified in the rating task. With the addition of Set 1 and the set-indexed constraint IDENT-IO[cont]$_1$, the ranking value of the non-indexed IDENT-IO[cont] drops, placing it below the relevant markedness constraints and generating a distinction between alternating and exceptional segments. This re-ranking improves matches for the problematic cases in Table 5, as seen in Table 7.

Table 7: Improvement for problematic forms for different root types with sets

	Target	**Input Frequency**	**Generated Frequency (no sets)**	**Generated Frequency (with sets)**
Alternating	[ba**k**a] ‘he cried’	0.07	0.40	**0.34**
Exceptional	[me**χ**ajem] ‘fulfills’	0.00	0.59	**0.24**
2 segments	[me**v**a**k**er] ‘visits’	0.72	0.00	**0.42**

In addition to this improvement, the combined model has a higher rate of matching input and generated frequencies for hybrid words, which were completely blocked without set-indexation. This is seen in (8).

(8) Hybrid root /bkʃ/ 'to ask for' using the combined model

/**bk₁**ʃ/ + sg.m.pres 'asks for'	IDENT-IO[cont]₁ (.723)	*V-STOP (.637)	*[+cont, -sib] (.592)	*STOP (1)	IDENT-IO[cont]
☞ a. me**v**a**k₁**eʃ Input (57.4%) Generated **(42.3%)**		*	*	*	*
b. me**b**a**k₁**eʃ Input (39.3%) Generated **(33.8%)**		**!		**	
c. me**v**a**χ₁**eʃ Input (0%) Generated **(23.9%)**	*!		**		**
d. me**b**a**χ₁**eʃ Input (3.3%) Generated **(0%)**	*!	*	*	*	*

Though the frequency matches are still not exact, the difference between the generated frequencies with or without the sets for exceptional segments is the closer match for input and generated frequencies of the highest rated variants [mevakeʃ] and [mebakeʃ], which the algorithm was unable to generate without sets.

4. Conclusion

Looking at Modern Hebrew spirantization, the presence of variation and exceptionality in the same phenomenon requires a combination of two mechanisms within the OT analysis. While the combination of stochastic constraint ranking and set-indexation helps bridge the gap between input and generated frequencies and account for the surfacing of hybrid words, more work is necessary to improve the frequency matches across all tokens.

References

Adam, Galit. 2002. From Variable to Optimal Grammar: Evidence from Language Acquisition and Language Change. PhD Dissertation. Tel-Aviv University.

Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. The Hague: Holland Academic Graphics.

Boersma, Paul, and Bruce Hayes. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45–86.

Hayes, Bruce, and Zsuzsa Londe. 2006. Stochastic Phonological Knowledge: The Case of Hungarian Vowel Harmony. *Phonology* 23: 59–104.

Hayes, Bruce, and Margaret MacEachern. 1998. Quatrain Form in English Folk Verse. *Language* 74: 473–507.

Pater, Joe. 2000. Nonuniformity in English Stress: The Role of Ranked and Lexically Specific Constraints. *Phonology* 12: 237–74.

Temkin Martinez, Michal. 2010. Sources of Non-Conformity in Phonology: Variation and Exceptionality in Modern Hebrew Spirantization. PhD Dissertation. University of Southern California.

Zuraw, Kie. 2000. Patterned Exceptions in Phonology. PhD Dissertation. UCLA.