TEL AVIV UNIVERSITY אוניברסיטת תל-אביב

The Lester and Sally Entin Faculty of Humanities

Department of Linguistics

THE ROLE OF SIMILARITY IN CO-OCCURRENCE RESTRICTIONS: EVIDENCE FROM THE HEBREW VERBAL SYSTEM

M.A. thesis submitted by

Hadas Yeverechyahu

Prepared under the guidance of:

Prof. Outi Bat-El

Dr. Evan-Gary Cohen

December 2014

ACKNOWLEDGEMENTS

Looking back on the five years I spent at the Linguistics department, as a BA and a graduate student, I feel grateful for the opportunity I have been given to study and explore the field I am passionate about the most. Many people have guided and accompanied me on this fascinating way, and I would like to thank them for that.

First and Foremost, I would like to express my deep gratitude to my advisors, Outi Bat-El and Evan-Gary Cohen. It was in Outi's and Evan's courses during my BA studies that I discovered that my passion has a name - phonology. I would like to thank them for teaching me, guiding me, and always encouraging me to ask the right questions and seek for explanations open-mindedly. I appreciate their being always available for consultation, and their extremely helpful comments, from insights about the analysis to help in translating animals' names. I could not ask for better advisors.

I was fortunate to be a student of Lior Laks in my first semester at the department, and ever since I benefited from his helpful advice and comments in personal correspondence, and at the TAU colloquium. I would also like to thank Orna Peleg, for her helpful comments on my research proposal, and for always showing interest in my study.

I would like to thank other faculty members that I had the privilege to participate in their courses: Galit Adam, Mira Ariel, Irena Botwinik, Julie Fadlon, Julia Horbath, Aya Meltzer-Asscher, Nirit Kadmon, and Tali Siloni. I would also like to thank Moshe Florentin of the Hebrew Culture department, who taught me everything I know about the phonology and morphology of Biblical Hebrew.

The lexical analysis was completed with the help of Shmuel Bolozky who sent me a digital copy of Evan-Shoshan dictionary. I would also like to thank the participants in my experiments, and to Aya Vituri for the invaluable help with the statistical analysis and insights.

I had the opportunity to present my study in several forums, and I would like to thank the audiences of the TAU colloquium, the Israeli Phonology Circle, and IGDAL 2 for their interesting comments.

I would like to thank Tal Oded and Ruti Zusman, the former and current secretaries of the department (respectively), whose doors were always open. Thank you for all the help (and always in no-time) and for always opening my day with a smile.

The department was my second home in the last five years, and I had the honor to share it with other excellent young linguists. I would like to thank my phonologist friends, Daniel Asherov, Si Berrebi, Yael Firer, Noa Karni, Enav Kedar, Stav Klein, Ezer Rasin, and Hadass Zaidenberg, for their helpful insights along the way. I would also like to thank Noa Brandel, Hila Davidovich, Noa Geller, Dolly Goldenberg, Omer Goldman, Anat Grubin, Netanel Haim, Ela Hillel, and Shiran Ofek for their comments on my experiments and for helping distributing them. Finally, special thanks to my dear friend, Tali Arad, for her wise comments, her advice, ideas and support. And to all of you, I would like to thank for the friendship and support along the way.

I would also like to thank my non-linguistic friends, and especially to Naomi and Ohad Feldheim (and their adorable son Lahav), for hours of delightful talks about linguistics in general and on my thesis in particular, and for their interesting insights.

Last but not least, I would like to thank my parents and my brother, for their great support and interest in my research. And finally, I would like to thank Chen Gafni. Thank you for always being there for me both professionally and personally, for your comments about any idea and thought, for reading every draft, and above all, thank you for your endless love and support.

ABSTRACT

In Semitic languages, homorganic consonants tend not to co-occur within the same stem (Greenberg 1950). Previous studies (e.g. McCarthy 1981, 1986; Frisch et al. 2004) suggested that these restrictions are due to similarity effects, that is, the greater the similarity between two (homorganic) consonants, the less likely they are to co-occur. The current study examines the restrictions in the Hebrew verbal system. I ask how similarity between consonants contributes to restrictions, and whether they are due to a universal constraint or influenced by language-specific factors.

The study has three main parts. First, I applied Frisch et al.'s (2004) similarity model to the Hebrew consonant inventory. Second, I analyzed the Hebrew verbal lexicon, focusing on the co-occurrences of C_1-C_2 stem consonants in the verb classes *kal* (*CaCaC*) and *pi'el* (*CiCeC*). The analysis shows a highly significant correlation between the similarity scale and the lexicon, and also suggests that place of articulation has a major role in the restrictions (compared to other features). To strengthen and complement the lexical analysis, I conducted two psycholinguistic experiments: a lexical decision task and a word-likelihood judgment task, both examine the co-occurrence restrictions in the speakers' phonological system. The results of the judgment task were highly correlated with the similarity scale and with the lexical analysis. The experiments also highlight the role of place features in the restrictions.

These findings suggest that there are similarity based co-occurrence restrictions on stem consonants C_1-C_2, both in the lexicon and in the speakers' phonological system. They also suggest that place features have a major role in the restrictions, such that consonants that share the major place feature are less likely to co-occur. However, the experiments cannot suggest whether the influence of similarity on the grammatical system is direct, or indirect through the lexical influences.

TABLE OF CONTENTS

Chapter 1: Introduction

In Semitic languages, homorganic consonants (i.e. consonants that share place of articulation) tend not to co-occur within the same stem. For example, verbs as *datam* or *kagam* are not likely to be found in the lexicon of any Semitic language (Greenberg 1950). Previous studies (as McCarthy 1981, 1986; Frisch et al. 2004) attributed these restrictions to similarity effects, such that the greater the similarity between two (homorganic) consonants, the less likely they are to co-occur.

The current study examines the co-occurrence restrictions in the Hebrew verbal system, focusing on the contribution of similarity between consonants to these restrictions. The study focuses on the co-occurrences of C_1-C_2 stem consonants in the verb classes *kal* (*CaCaC*) and *pi'el* (*CiCeC*), both in the verbal lexicon (lexical analysis) and in the phonological systems of the speakers (psycholinguistic experiments). Similarity between consonants was calculated based on Frisch et al.'s (2004) similarity model (originally proposed for Arabic), adjusted according to the Hebrew consonant inventory.

Three main questions were asked in the study:

a. What are the co-occurrence restrictions in the Hebrew verbal system?
b. What is the role of similarity in the co-occurrence restrictions?
c. Nature vs. Nurture: are co-occurrence restrictions caused by a universal constraint or influenced by language-specific lexical factors?

The study has three main parts. Each part tested a different aspect of the co-occurrence restrictions, and then the correlations between the parts were examined. The parts of the study are as follows:

a. Application of Frisch et al.'s (2004) model to the Hebrew consonant inventory (§4);
b. Lexical analysis of co-occurrence restrictions in the Hebrew verbal lexicon (§6);

c. Two psycholinguistic experiments: a lexical decision task and a judgment task. Both experiments examine the role of co-occurrence restrictions in the speakers' phonological system (§7).

The results show a highly significant correlation between the similarity scale and the lexical analysis, and also between the similarity scale and the results of the judgment experiment. A correlation was found between the lexical analysis and the judgment experiment results as well. These findings suggest that there are similarity based co-occurrence restrictions on stem consonants C_1 and C_2, both in the lexicon and in the speakers' phonological system. However, the experiments cannot suggest whether the influence of similarity on the grammatical system is direct, or indirect through the lexical influences. In addition, the results suggest that place of articulation has a major role in the restrictions (compared to other features), such that consonants that share the major place feature are less likely to co-occur. This finding strengthens previous claims regarding the important role of OCP-Place in co-occurrence restrictions in Semitic languages (McCarthy 1981, 1986; Frisch et al. 2004 among others). However, the highly significant correlation between the results and the similarity scale proposes that not only the major place feature affects co-occurrence restrictions; if so, we would expect to see no effect in non-homorganic pairs.

The study is organized as follows: §2 provides a theoretical background for the study; §3 presents the main issue: the research questions and the data sources; §4 presents the accommodation of Frisch et al.'s model to Hebrew; §5 is dedicated to the different hypotheses of the study; §6 presents the lexical analysis; §7 presents the psycholinguistic experiments: §7.1 describes the lexical decision experiment and §7.2 describes the word-likelihood judgment experiment; §8 discusses the study's results; I conclude in §9.

CHAPTER 2: THEORETICAL BACKGROUND

The study examines the correlation between co-occurrence restrictions and segment similarity. In this chapter, I provide the theoretical background for the study: §2.1 presents co-occurrence restrictions and the OCP, and §2.2 presents previous studies on similarity.

2.1 Co-Occurrence Restrictions

In an extensive cross-linguistic research, Greenberg (1950) observed that in Semitic languages, there are no verbs in which the first two stem consonants are identical (e.g. *didem*), and more generally, that homorganic consonants tend not to co-occur within the same stem. McCarthy (1979, 1981, 1986) provides a theoretical account for this phenomenon based on the Obligatory Contour Principle (OCP; Leben 1973, Goldsmith 1976), which was originally formulated for tonal systems. McCarthy expanded the principle to root consonants in Semitic verbal systems (1979, 1981, 1986) and it was further broadened to segments in general, features, syllables, and even morphemes (see Yip 1998). A common definition of OCP, cited from McCarthy (1986a:208), appears in (1).

(1) *The Obligatory Contour Principle (OCP):*

At the melodic level, adjacent identical elements are prohibited.

McCarthy (1986) suggests that stem consonants and vocalic patterns are independent morphological units, located on different tiers. Since stem consonants are adjacent on their tier, the OCP rules out any representation with adjacent identical element. Note that the second and the third stem consonants (C_2 and C_3) are allowed to be identical (e.g. *dimem* 'to bleed', *kilel* 'to curse', *miʃeʃ* 'to grope').[1] McCarthy suggests that in these verbs, the stem contains only two consonants, and the second consonant C_2 spread into the empty C_3 slot. This type of verb is beyond the scope of this study.

Rose (2000) claims that these restrictions can be explained without referring to tiers. In her view, the OCP is not restricted to adjacent consonants but depends on proximity as well. Thus, identical consonants separated by vowels (i.e. C_iVC_i) also violate the OCP constraint, though to a lesser extent than C_iC_i given the higher proximity. Along this line, the restrictions on C_1 and C_2 will be greater than the restrictions on C_1 and C_3, since C_1-C_3 are farther away from each other. Greenberg (1950) indeed shows this tendency, as does Frisch et al. (2004). The current study examines only C_1-C_2, and leaves proximity for further research.

Hebrew and Arabic supply evidence for these co-occurrence restrictions. Laks (2011) shows blocking due to OCP in Hebrew and Arabic, where some verbs fail to undergo valence-changing operations since such operation would lead to an OCP violation. For example, *dike* 'to make depressed' does not undergo the valence-changing operation to **hitdake* 'to get depressed', although it is semantically possible. If such a derivation had occurred, it would have created an OCP violation (*t*-*d*). In addition, OCP restrictions have empirical support from psycholinguistic experiments: Frisch and Zawaydeh (2001) for Arabic, Berent and colleagues (Berent and Shimron 1997, Berent, Everett and Shimron 2001 among others) for Hebrew.

Bat-El (2003) claims that these restrictions are not unique to Semitic languages, and that co-occurrence restrictions on stem consonants can be found in non-Semitic languages as well. In English, for example, there are no monosyllabic words of the form *sCVC* in which the same non-coronal consonant (i.e. labial or velar) appears in both

[1] The verbs are presented in 3rd person singular past throughout.

sides of the vowel, for example **spep*, **skik* (Fudge 1969, Clements and Keyser 1983, Davis 1984).

In Japanese, co-occurrences restrictions on homorganic consonants are found in Yamato (native-Japanese) stems. In addition, Japanese has blocking effects due to OCP violations. Consider, for example, the phenomenon of Rendaku - voicing of the first consonant of the second member in a compound. For historical reasons, *h* alternates with *b*, as in *nui* 'saw' + *hari* 'needle' → *nui-bari* 'sewing needle'. However, when the stem begins with *h* followed by *m*, Rendaku is blocked in order to avoid two near homorganic consonants, for example *mai* 'dance' + *hime* 'princess' → *mai-hime* 'dancing princess', and not **mai-bime* (Kawahara et al. 2006). Note that when there is a non-labial consonant between the *h* and the *m*, Rendaku does occur (e.g. *ryoori-basami* 'cooking scissors', *naga-bakama* 'long hakama'). This finding suggests that proximity plays a role as well.[2]

McCarthy (1986) suggests that blocking due to an OCP violation is universal. Odden (1988) stipulates that blocking differs cross-linguistically, and language differ in the sets of relevant features for the principle. In Optimality Theory (Prince and Smolensky 1993) this is represented by different constraint rankings in different languages.

2.2 Similarity

Different studies (Pierrehumbert 1993, Frisch et al. 2004, Mielke 2009 among others) have addressed the question of how segment similarity should be measured. The current study focuses on the phonological approach that is based on articulatory phonological features; other approaches, like those based on acoustic parameters (see, for example, Mielke 2009), are beyond the scope of this study.

[2] See Yip (1988) for more examples of OCP as process-trigger or process-blocker.

Pierrehumbert (1993) calculates similarity between two segments by counting the number of feature values the segments share. Frisch et al. (2004) expand this model to a natural-classes-based model, in which similarity value is computed for each pair of segments by the number of natural classes they share. Thus, in Frisch et al.'s model, similarity is computed by dividing the number of shared natural classes of two segments by the sum of the shared and non-shared natural classes of the two segments. The formula appears in (2).

(2) Frisch et al.'s (2004) similarity formula

$$similarity = \frac{shared\ natural\ classes}{shared\ natural\ classes + non\ shared\ natural\ classes}$$

By this procedure, a similarity scale for each language can be computed, based on the contrastive features and natural classes of the language. Frisch et al. (2004) tested the model on Arabic verb stem consonants, looking for OCP restrictions in the verbal paradigms. First, they showed that OCP restrictions do occur in the lexicon, where combinations of consonants with shared features are underrepresented systematically. Next, using the above formula, Frisch et al. constructed a similarity scale based on natural classes defined according to contrastive phonological features found in Arabic.[3] Then, the results of the lexical study were examined in light of the similarity scale. The study showed a strong correlation between them, namely the similarity scale, based on natural classes, successfully explaining the co-occurrence restrictions in the Arabic lexicon. The current study will examine this model in Hebrew, by a lexical study and psycholinguistic experiments.

[3] [±consonantal], [±sonorant], [±continuant], [±acute], [±strident], [±nasal], [±lateral], [labial], [coronal], [dorsal], [pharyngeal], [radical], [±anterior], [±back], [±voice], [±spread glottis], and [±constricted glottis].

Next, the question arises as to which features are relevant to similarity. Rose and Walker (2004) claim that [sonorant], [continuant] and place features are the most important in computing similarity. Kawahara (2007) suggests that manner features (such as palatalization, voicing, nasalization, and continuity) contribute to similarity more than place features. This is compatible with claims that manner features are perceptually more salient, and that speakers tend to rely on acoustic parameters while calculating similarity (see Mielke 2009). Kaisse (1988) claims that the OCP applies to feature groups and not just to single features, and so provides direct evidence for Feature Geometry, which argues for feature hierarchies (Clements 1985, Sagey 1986, Clements and Hume 1995, see also McCarthy 1988). Along this line, Padgett (1995:181) revised the definition of the OCP to take into account feature hierarchies (3):

(3) *The Obligatory Contour Principle (OCP):*

At the melodic level, adjacent identical elements F F are prohibited, iff all subsidiary features stipulated for F are also identical.

Along the line of Frisch et al.'s (2004) study, the current study examines natural classes and does not test the influence of every feature individually. The only feature that is examined separately is place of articulation, following the importance of OCP-Place as suggested by previous studies (see §2.1). Further research is needed in order to test the influence of other features and the correlations with feature hierarchies.

CHAPTER 3: THE ISSUE

3.1 Research Questions

The goal of this study is to examine the co-occurrence restrictions in the Hebrew verbal system, focusing on the contribution of similarity between consonants to these restrictions. Three main questions are addressed:

a. What are the co-occurrence restrictions in the Hebrew verbal system? While previous studies (McCarthy 1981, 1986; Pierrehumbert 1993; Frisch et al. 2004) focused on OCP-Place, shared place as a necessary feature for the effect, the current study attributes equal weight to all features.

b. What is the role of similarity in the co-occurrence restrictions?

c. Nature vs. Nurture: are co-occurrence restrictions caused by a universal constraint or influenced by language-specific lexical factors?

3.2 Data

The study examines the co-occurrence restrictions in two sources of data:

a. The Hebrew verbal lexicon, focusing on C_1-C_2 stem consonants in verb classes *kal* (*CaCaC*) and *pi'el* (*CiCeC*). I use the list of verbs from the Even-Shoshan dictionary (edition 1970 with completions from 1983).[4]

b. Two psycholinguistic experiments, a lexical decision task and a word-likelihood judgment task, to examine the role of similarity in the speakers' phonological system.

All the data were analyzed with respect to Frisch et al.'s (2004) model. This is, inter alia, since Arabic and Hebrew are historically related (Schwarzwald 2002 among many others), and Frisch et al.'s model successfully explained the OCP effect in Arabic verbs.

[4] Many thanks to Shmuel Bolozky for an electronic version of the verb list.

CHAPTER 4: THE SIMILARITY MODEL

The first part of the study applies Frisch et al.'s (2004) similarity model to the Hebrew consonant inventory. As discussed in §2, the model computes a single similarity value (from 0 to 1) for each pair of consonants, and the computation is based on the natural classes to which the consonants belong. The classes are defined according to the language contrastive features. The Hebrew consonant inventory appears in (4), and the set of contrastive features I used appears in (5).

(4) Hebrew consonants

	Bilabial	Labio-dental	Alveolar	Palato-alveolar	Palatal	Velar	Uvular
Plosive	p b		t d			k g	
Fricative		f v	s z	ʃ		x[5]	
Affricate			ʦ				
Nasal	m		n				
Lateral			l				
Approximant					j		ʁ[6]

(5) Set of contrastive features

	p	b	m	f	v	t	d	s	z	ʦ	ʃ	n	l	j	k	g	x	ʁ
[±sonorant]	-	-	+	-	-	-	-	-	-	-	-	+	+	+	-	-	-	+
[LAB]	√	√	√	√	√													
[COR]						√	√	√	√	√	√	√	√	√				
[strident]								√	√	√	√							
[±anterior]						+	+	+	+	+	-	+	+	-				
[DOR]															√	√	√	√
[±continuant]	-	-	-	+	+	-	-	+	+	-	+	-	+	+	-	-	+	+
[±voice]	-	+		-	+	-	+	-	+	-	-				-	+	-	

[5] Bolozky and Kreitman (2007) consider the Hebrew dorsal fricative to be uvular. Nevertheless, its exact place of articulation has no consequences for the current study, since minor place features for the dorsals are not contrastive in Modern Hebrew.

[6] The Hebrew rhotic is considered a uvular approximant with certain frication (Bolozky and Kreitman 2007), IPA: [ʁ̞]. Hereinafter it will be transcribed as ʁ.

I excluded borrowed consonants (*ʒ, ʤ, ʧ* and *w*) from the analysis due to their rare appearance in the verbal system, and the glottals (*ʔ* and *h*) due to their tendency to be omitted in Modern Hebrew.[7] This was done mainly for the sake of comparison between the lexical analysis and the experiments' results. The feature system I used is based on binary (6a) and unary (6b) values:

(6) Features:

a. Binary: [±sonorant], [±continuant], [±voice], [±anterior]
b. Unary: place features: [LAB] (labials), [COR] (coronals), [DOR] (dorsals); [strident]

Two issues should be noted: First, [±voice] is relevant only for obstruents; it is not contrastive among sonorants, and it has been claimed that the voice feature of the sonorants is inherent in them and therefore differs from the voice feature of the obstruents (Rice 1993). Second, I refer to stridency as a unary feature, such that the value [-strident] is not a part of the system. The stridents in Hebrew show a common phonological behavior – they undergo metathesis in *binyan hitpa'el* (e.g. *hit-sapeʁ* → *histapeʁ* 'to have a haircut'). Therefore, [strident] is relevant for the Hebrew phonological system.[8] The non-strident consonants, on the other hand, do not show any common phonological behavior in Hebrew, and [-strident] is also not necessary for minimal distinction between consonants in the system. For these reasons, I excluded the [-] value of this feature from the analysis.[9] The natural classes were defined based on this feature system, down to the level of singletons.

[7] The question of whether the phonological system of Hebrew represents glottals is beyond the scope of this paper.
[8] Note that [strident] is more of an acoustic rather than articulatory feature. Nonetheless, it is widely used (also in Frisch et al.'s model) and explains different phonological processes in different languages.
[9] Frisch et al. (2004) used [-strident] only for non-strident coronal fricatives. Hebrew has no such consonants in its inventory.

Based on this phonological feature system, I computed the similarity value for each pair of consonants, using Frisch et al.'s (2004) formula. The formula was presented in (2), and is repeated in (7). The natural classes and the similarity values were calculated via a Microsoft Excel macro.[10]

(7) Frisch et al.'s (2004) similarity formula

$$similarity = \frac{shared\ natural\ classes}{shared\ natural\ classes + non\ shared\ natural\ classes}$$

For example, the similarity value for *p* and *b* is calculated as follows: they share 7 natural classes, and do not share 8 classes, namely, there are 8 natural classes in which only one of them is a member (see list in (8)). Therefore, the similarity value for the pair *p,b* is: $\frac{7}{7+8} = 0.467$.[11]

(8) Shared and non-shared natural classes for the pair *p-b*:

a. Shared classes: [-son] = {*p,b,f,v,t,d,s,z,ʦ,ʃ,k,g,x*}, [-son, LAB] = {*p,b,f,v*}, [-son, LAB, -cont] = {*p,b*}, [-son, -cont] = {*p,b,t,d,ʦ,k,g*}, [LAB] = {*p,b,m,f,v*}, [LAB, -cont] = {*p,b,m*}, [-cont] = {*p,b,m,t,d,ʦ,n,k,g*}.

b. Non-shared classes:

   i. *p* and not *b*: [-son, LAB, -cont, -voice] = {*p*}, [-son, LAB, -voice] = {*p,f*}, [-son, -cont, -voice] = {*p,t,ʦ,k*}, [-son, -voice] = {*p,f,t,s,ʦ,ʃ,k,x*};

   ii. *b* and not *p*: [-son, LAB, -cont, +voice] = {*b*}, [-son, LAB, +voice] = {*b,v*}, [-son, -cont, +voice] = {*b,d,g*}, [-son, +voice] = {*b,v,d,z,g*}.

Note that although there is only one feature that distinguishes these two segments, [±voice], they have eight non-shared natural classes. Since the calculation is

---

[10] Many thanks to Chen Gafni for programming the Natural Classes Generator on Microsoft Excel platform. The full list of natural classes appears in Appendix A.

[11] I treat similarity as symmetrical, and (a)symmetry is beyond the scope of this paper. See also §9.

based on natural classes and not on features directly, the distance between them in this model is more notable.

Table (9) presents the similarity values for the consonants in Hebrew, and table (10) the most similar pairs on the scale. A full list appears in Appendix B.

(9) Similarity values

	p	b	m	f	v	t	d	s	z	ʦ	ʃ	n	l	j	k	ɡ	x	ʁ
p	1.000	0.467	0.200	0.313	0.167	0.250	0.143	0.067	0.037	0.192	0.074	0.050	0.000	0.000	0.313	0.176	0.100	0.000
b		1.000	0.200	0.167	0.313	0.136	0.263	0.032	0.077	0.107	0.036	0.050	0.000	0.000	0.167	0.333	0.048	0.000
m			1.000	0.063	0.063	0.050	0.053	0.000	0.000	0.038	0.000	0.214	0.059	0.059	0.063	0.067	0.000	0.077
f				1.000	0.429	0.091	0.045	0.192	0.125	0.071	0.217	0.000	0.050	0.050	0.111	0.056	0.313	0.063
v					1.000	0.043	0.095	0.107	0.227	0.034	0.120	0.000	0.050	0.050	0.053	0.118	0.167	0.063
t						1.000	0.500	0.296	0.192	0.700	0.185	0.200	0.087	0.042	0.263	0.150	0.087	0.000
d							1.000	0.172	0.304	0.375	0.107	0.211	0.091	0.043	0.150	0.294	0.043	0.000
s								1.000	0.520	0.414	0.500	0.069	0.185	0.103	0.069	0.034	0.185	0.037
z									1.000	0.233	0.296	0.080	0.217	0.120	0.038	0.083	0.120	0.043
ʦ										1.000	0.226	0.154	0.069	0.033	0.200	0.115	0.069	0.000
ʃ											1.000	0.037	0.115	0.208	0.077	0.038	0.208	0.042
n												1.000	0.313	0.167	0.053	0.056	0.000	0.063
l													1.000	0.467	0.000	0.000	0.048	0.200
j														1.000	0.000	0.000	0.048	0.200
k															1.000	0.462	0.313	0.063
ɡ																1.000	0.176	0.067
x																	1.000	0.200
ʁ																		1.000

(10) Most similar pairs[12]

	pair	similarity value
1	ʦ-t	0.7
2	s-z	0.52
3	t-d	0.5
3	s-ʃ	0.5
4	p-b	0.467
4	l-j	0.467
5	k-g	0.462
6	f-v	0.429
7	ʦ-s	0.414

[12] Excluding identical consonants, which have a similarity value of 1.

Chapter 5: Hypotheses

After calculating the similarity scale, I examined the correlation between co-occurrence restrictions and the similarity value of the first two stem consonants of the verbs. The correlation was examined on two levels: the lexical level and the phonological level (in the word-likelihood judgment experiment). The lexical analysis and the word-likelihood judgment experiment (as well as their correlation) may have different results, which would lead to different conclusions.

a. *Lexicon 1 - Experiment 1:* In this scenario, the same similarity-based co-occurrence restrictions are found both in the lexicon and in the speakers' judgments. Such results may indicate that similarity plays a role in co-occurrence restrictions in Hebrew. However, it will not suggest whether the influence of similarity on the grammatical system is direct, or indirect through the lexical influences.

b. *Lexicon 0 - Experiment 0:* In this scenario, similarity-based co-occurrence restrictions are not found in Hebrew at all. Based on previous studies on OCP in Arabic (Greenberg 1950; McCarthy 1981, 1986; Frisch et al. 2004), this is the least plausible scenario.

c. *Lexicon 1 - Experiment 0:* In this scenario, co-occurrence restrictions are found in the lexicon but not in the speakers' judgments. Such results may indicate that OCP was active in previous stages of Hebrew (many verbs in the Modern Hebrew lexicon have origins in Biblical Hebrew or in Mishnaic Hebrew), but nowadays the constraint is no longer active.

d. *Lexicon 0 - Experiment 1:* In this scenario, co-occurrence restrictions are not found in the lexicon but are found in the speakers' judgments. Such results may indicate that OCP is not active in the lexicon, but the speakers are sensitive to it, nevertheless. A plausible explanation would be that the OCP is a universal principle, which the sensitivity to it does not come from the segmental distributions in a specific language. This hypothesis would be supported by studies such as Berent (2008), who found that speakers of Korean, which does not have clusters, are nonetheless sensitive to SSG (Sonority Sequencing Generalization) violations.

CHAPTER 6: THE LEXICAL ANALYSIS

The second part of the study analyzes the Hebrew verbal lexicon, based on Frisch et al.'s (2004) model. Unlike Frisch et al.'s analysis of Arabic, the current study examines not only homorganic consonants, but also every other possible combination of consonants. The list of verbs is taken from Even-Shoshan dictionary (edition 1970 with completions from 1983), and I used Barkali (1964) for full paradigms.

6.1 Design

The study focuses on two verb classes (*binyanim*): *kal* (*CaCaC*) and *pi'el* (*CiCeC*). I chose these classes since they show different behaviors throughout the paradigm: while in *pi'el* C_1-C_2 are separated by one vowel throughout the inflectional paradigm, in *kal* the future paradigm gives rise to adjacent C_1-C_2 (see Appendix C for sample paradigms). Thus, it is possible to examine whether this difference in distance has an impact on the results.

The analysis was conducted from a synchronic point of view, with the aim of comparing its results with the psycholinguistic experiments results. Thus, I analyzed only regular verbs (*shlemim*, see Zadok 2012), in which all three-stem consonants appear synchronically throughout the paradigm. Therefore, I excluded from the analysis the glottals (*ʔ*, *h*), *v* (orthographic: *va"v*; historical: *w*, synchronic: *v*) and *j*. For example, the verb *ʃama(ʕ)* 'to hear' historically ended with a *ʕ* and traditionally is considered as part of the regular verbs. However, nowadays final *ʕ* is omitted, so the verb is in a template of *CaCa*. Along the lines of Zadok (2012), this verb is not part of the regular verbs, and therefore was omitted from the analysis. In addition, consonants that have undergone a historical change are considered by their synchronic status. Thus, historical

$t^ʕ$ is considered as *t*; historical *q* as *k*; historical *ħ* as *x*. Overall, 779 verbs in *kal* and 678 verbs in *pi'el* were analyzed.[13]

The analysis takes into account paradigms (and not only stems), such that each verb appears in three forms drawn from the past, present and future paradigms (all forms are in 3rd, singular, masculine). In this way, alternations throughout the paradigm can be considered, including differences in the distance between C_1 and C_2 (e.g. ***ʃam**aʁ* 'he saved' – C_1VC_2, *ji**ʃm**oʁ* 'he will save' C_1C_2), and the spirantization of *b,p,* and *k* to *v,f,* and *x* respectively (e.g. ***k**atav* 'he wrote', *ji**x**tov* 'he will write'), see Appendix C for sample paradigms. For example, consider the pairs *d-ʃ* and *b-ʃ*. Each pair has only one verb in *pi'el*: *diʃen* 'to fertilize' for *d-ʃ* and *biʃel* 'to cook' for *b-ʃ*. However, due to spirantization alternations, *d-ʃ* has three occurrences in the lexicon: *diʃen-medaʃen-jedaʃen* 'to fertilize Past-Present-Future', while *b-ʃ* has only one: *biʃel* 'to cook Past'. The present and future forms, *mevaʃel* and *jevaʃel* respectively, contain *v* instead of *b* due to spirantization, and thus contribute to the pair *v-ʃ*.

After selecting the relevant verbs, I counted how many forms there were for each C_1-C_2 pair. For example, consider the pair *d* and *m*: 24 forms in the tested lexicon begin with this pair: 9 forms in *kal* (6 for *d-m* and 3 for *m-d*), and 15 forms in *pi'el* (6 for *d-m*

[13] A few comments are addressed:

a. In *binyan kal*, future tense, an epenthetic vowel may be inserted after a synchronic *x* that historically originated in *ħ*, for example *jaxaʃov~jaxʃov* 'he will think'. However, synchronically, speakers tend not to epenthesize a vowel in these cases (i.e. stick to the standard form), and evidence for variation between the two forms appears even in the Bible (e.g. *taħbol~taħavol* 'you ms. will take as pledge', Exodus, 22;25, Deuteronomy, 24;17, respectively). Therefore, I included these forms in the analysis.

b. Verbs in *pi'el* with C_2 *ʁ* have (normatively, at least) a vocalic pattern of *CeCeC* (e.g. *seʁek* 'he combed'), and not the standard *CiCeC*, due to historical changes. Since it is plausible to assume that the different vocalic pattern does not influence the similarity between C_1-C_2, I included these forms in the analysis.

c. I included in the analysis verbs in *kal* with C_1 *n*, although in some of these verbs the *n* is deleted in the future form, for example *nafal-jipol* (and not **jinpol*; *p~f* alternation due spirantization) 'he fell\will fall' (respectively). In these verbs, only past and present forms were taken into account.

and 9 for *m-d*). For example, the triplet ***mad**ad-**m**o**d**ed-i**md**od* 'to measure Past-Present-Future' represent three instances.

Next, I compared the observed (O) results to the expected (E) ones (O/E), based on consonant frequencies, in order to examine what (if any) the co-occurrence restrictions on C_1-C_2 are. According to previous studies on OCP in Arabic (Greenberg 1950, Frisch et al. 2004 among others), there is a solid basis to assume that some restrictions will be shown in the Hebrew lexicon as well. After calculating the O/E ratio, I compared the results to the similarity model, in order to examine if a correlation can be found between co-occurrences and the similarity values.

Two questions were asked:

a. Observed vs. Expected (O/E): Are there any differences between the observed and the expected occurrences of each consonant pair in the lexicon? In other words, is the number of occurrences of each pair similar to what would be obtained if the lexicon were random?

b. Correlation with the similarity scale: Is there any correlation between the occurrences in the lexicon and the similarity scale?

6.2 Results

6.2.1 Observed vs. Expected (O/E)

First, in order to look for differences between the observed and expected co-occurrences, a *chi-square test* was conducted for each verb class (*binyan*) separately (a full list of the occurrences appears in Appendix D). In one calculation, the order of the consonants was taken into account (e.g. *d-t* was calculated separately from *t-d*) and in the other the order was not inserted as a factor (e.g. *d-t* and *t-d* were calculated together as one item). In addition, the tests took into account the frequency of each tested consonant in the corpus (a full list of the frequencies appears in Appendix E). Thus, the expected results refer to what would be expected if the single consonants were combined to pairs of C_1-C_2 randomly.

The results show highly significant differences between the observed and expected, in all the tested cases: *binyan kal*: with no consideration of order: $\chi^2 = 970.24$, $p < 0.0001$; including consideration of order: $\chi^2 = 1069.47$, $p < 0.0001$; *binyan pi'el*: with no consideration of order: $\chi^2 = 767.8215$, $p < 0.0001$; including consideration of order: $\chi^2 = 912.03$, $p < 0.0001$. These results show that there is a gap in the lexicon between the observed and expected consonant co-occurrences. Based on previous studies (see §2), it is plausible to assume that similarity is one of the factors that causes this gap, and the next sub-sections test this assumption.

6.2.2 Observations

Next, a few interesting observations can be made by looking at the bottom of the occurrences list. Since a significant gap between observed and expected was found also in the list that do not take order and *binyanim* into account, the following sub-section deals with the combining list. Thirty pairs of consonants do not appear in the lexicon at all. Table (11) presents them and their similarity values on the similarity scale.

(11) No occurrences at the lexicon (in brackets: place on the list, out of 66)[14]

Pair	Similarity	Pair	Similarity	Pair	Similarity
b-b	1	z-z	1	d-z	0.304 (11)
d-d	1	ʦ-t	0.7 (1)	m-b	0.2 (23)
f-f	1	s-z	0.52 (2)	m-p	0.2 (23)
g-g	1	t-d	0.5 (3)	ʁ-l	0.2 (23)
p-p	1	b-p	0.467 (4)	b-f	0.167 (28)
ʁ-ʁ	1	k-g	0.462 (5)	p-v	0.167 (28)
s-s	1	f-v	0.429 (6)	m-f	0.063 (51)
ʃ-ʃ	1	ʦ-s	0.414 (7)	m-v	0.063 (51)
t-t	1	b-v	0.313 (10)		
ʦ-ʦ	1	n-l	0.313 (10)		
v-v	1	p-f	0.313 (10)		

A few observations can be made. First, it is notable that the most similar pairs indeed do not appear in the lexicon. Identical consonants tend not to co-occur (except for five pairs, which will be discussed in the next paragraph), and places 1-7 in the most similar pairs (except for one) do not appear in the lexicon as well. Note that the pair *s-ʃ* is also ranked third in the similarity scale with a similarity value of 0.5. It has nine occurrences in the lexicon, three verbs in three tenses each, from which only one verb is in use (but with low frequency) in Modern Hebrew (*ʃisef* 'to slit someone's throat').[15]

Five identical consonant pairs have more than zero occurrences in the tested lexicon, summarized in table (12).

[14] Identical consonants have a similarity value of 1, and are not considered in the numbering.
[15] The other two verbs are *ʃasaf* 'to split (Middle Ages Hebrew)' and *ʃasas* 'to despoil (Biblical Hebrew)'.

(12) Identical consonant pairs with occurrences in the lexicon

Pair	Occurrences	Different Verbs	Details
k-k	1	1	One form – normative: *kikev* 'to star', colloquial: *kixev*.
l-l	3	1	*lilev-melalev-jelalev* 'to strengthen with a palm branch Past-Present-Future' – unused in Modern Hebrew.
n-n	3	1	*nines-menanes-jenanes* 'to make smaller Past-Present-Future' – unused in modern Hebrew.
m-m	6	2	*mimen-memamen-jemamen* 'to finance Past-Present-Future' and *mimeʃ-memameʃ-jemameʃ* 'to realize Past-Present-Future' – both are in used in Modern Hebrew.
x-x	15	9	All of them contained historically one spirantized *k* (*x*) and *ħ* (in Modern Hebrew – *x*); therefore, the identical *x*-*x* do not appear throughout the paradigm. They are used very rarely or not at all in Modern Hebrew (for example *xaxaʁ* 'to lease').[16]

Overall, there are only two verbs with originally identical C_1-C_2 that are in use in Modern Hebrew, both with *m*: *mimen-memamen-jemamen* 'he financed Past-Present-Future' and *mimeʃ-memameʃ-jemameʃ* 'he realized Past-Present-Future'. The others are not part of the commonly used lexicon (at least some of them for semantic reasons; *lilev* is not used since it is not customary to use a palm branch as a strengthening device nowadays).

The largest group of verbs with identical C_1-C_2 contains *x* in both positions. However, all these cases are a combination of a historical *ħ* and spirantized allophone

[16] The Hebrew orthography shows evidence to the historical change: ח stands for the historical *ħ* and כ stands for *k*, or for *x* which is a historical allophone of *k*.

of *k*. Hence, historical reasons are responsible to the identical C_1-C_2 in the synchronic lexicon; originally, C_1 and C_2 were different from each other in these verbs.

Another interesting finding is the verb *kikev* 'he starred'. This verb is widely used in Modern Hebrew, but speakers pronounce it *kixev*. It is plausible to assume that this pronunciation is influenced by several factors: first, this verb is strongly related to the noun *koxav* 'a star', and in fact, the first four discussed forms are denominatives (also *lilev-lulav* 'palm branch', *nines-nanas* 'midget', *mimen-mamon* 'money'). Thus, under word-based morphology, the verb was derived from the word *koxav* itself, and the speakers want to keep maximum faithfulness between the noun and the verb (see Aronoff 1976 and Bat-El 1994). Second, changing the second *k* into *x* reduces the similarity between the two. The change is possible because it would not cause a change in orthography (the letter כ allows both consonants, *k* and *x*), and because *x* is a phoneme in Modern Hebrew and can appear in a non-spirant position (i.e. a position in which there are no conditions for spirantization).

The results so far show a correlation between similarity and co-occurrence restrictions. Nonetheless, it can be noted that all the pairs that do not appear in the lexicon (except for *ʁ-l*, which will be discussed in the next paragraph), also share a major place of articulation. Therefore, the co-occurrence restriction can be connected to OCP-Place violation, as was suggested for Arabic and for Semitic languages in general (McCarthy 1979, 1981, 1986; Frisch et al. 2004 among others). Moreover, the list of zero occurrences also contains homorganic pairs that are rated low on the similarity scale, for example *m-f* and *m-v* (both 0.063, place 51 from 66), a finding that strengthens the claim regarding the major role of place features in the co-occurrence restrictions.

The pair *ʁ-l* acts like homorganic pairs that do not appear in the lexicon. Since the Hebrew rhotic is considered uvular and not coronal (Bolozky and Kreitman 2007), this finding is slightly surprising, and can be related to the special status of the Hebrew rhotic. Note that the pair *ʁ-x,* which share a major place of articulation of dorsals, has 72 occurrences in the lexicon. Both *ʁ-x* and *ʁ-l* have a similarity value of 0.2 in the similarity model, a fact that adds more to the puzzle.[17]

To conclude, the 30 pairs that have zero occurrences in the tested lexicon suggest that similarity plays a role in co-occurrence restrictions. Almost all the identical C_1-C_2 and the most similar pairs (according to the model) do not appear in the lexicon. However, it is notable that all the pairs (except for *l-ʁ*) that do not appear in the lexicon are homorganic. This suggests that place features have a special role in the co-occurrence restrictions.

6.2.3 Correlation between the Lexicon and the Similarity Scale

Next, a *Zero-Inflated Poisson Regression* model was built in order to test statistically the correlation between the lexicon and the similarity scale. The similarity factor in the model is expected to be negative, since the more consonants are similar to each other (their similarity value is closer to 1), the less they are expected to co-occur in the lexicon. The model shows that the similarity factor is significantly negative in all cases: *binyan kal*: Estimate = -0.112, SD = 0.023, $p < 0.0001$; *binyan pi'el*: Estimate = -0.147, SD = 0.021, $p < 0.0001$; *kal* and *pi'el* together: Estimate = -0.092, SD = 0.014, $p < 0.0001$.[18]

---

[17] There is evidence that the Biblical Hebrew rhotic is dorsal (it behaves like pharyngeals and glottals, e.g. cannot undergo gemmination and cause vowel lowering). However, some studies suggest that the Biblical rhotic has two variants – coronal and dorsal – and in some Mizrahi Jews communities the rhotic is pronounced as coronal (Blau 2010).

[18] For simplicity of the statistical calculations, in the calculations for *pi'el* and for the sum of *kal* and *pi'el* only the second consonant and the similarity value were taken into account.

The similarity factor in these models indicates the correlation between the appearance in the lexicon and the similarity scale: the more two consonants are similar to each other, the less their chances to co-occur as C_1-C_2 in a Hebrew verb.

6.2.4 Conclusion

The lexical analysis tested pairs of stem consonants C_1-C_2 in the Hebrew verbal lexicon, and shows a significant difference between observed and expected (O/E) co-occurrences. Assuming that languages are systematic, this result suggests that there are factors that shape the lexicon and impose restrictions. Statistical models suggest that similarity (based on Frisch et al.'s 2004 model) is one of these factors. The next chapter tests the role of similarity in co-occurrence restrictions from a synchronic point of view, focusing on the phonological system of Hebrew speakers.

CHAPTER 7: THE EXPERIMENTS

In chapter 6, I described the lexical analysis, which looked for similarity based co-occurrence restrictions. The results show that the gaps in the lexicon are found in correlation to the similarity scale. However, a lexical analysis cannot supply a complete account for co-occurrence restrictions. First, this kind of analysis can only highlight what exists, and not ask directly about what is absent. Second, the tested lexicon represents an upper bound of the Hebrew speaker's vocabulary (not all speakers are familiar with all forms in the lexicon), thus in any case it cannot directly represent the phonological knowledge of a particular Hebrew speaker. Third, lexicons have lots of exceptions for example due to historical residue.

In order to complete the picture, an experiment on nonce-verbs in Hebrew was conducted. In the lexical decision experiment the participants were asked to make a lexical decision about verbs and non-verbs, and in the word-likelihood judgment experiment the participants were asked to give word-likelihood ratings for nonce-verbs in Hebrew.

7.1 The Lexical Decision Experiment

Following the observation that C_1 and C_2 obey the OCP-Place in Semitic languages (Greenberg 1950; McCarthy 1981, 1986), I conducted a psycholinguistic experiment that examined OCP-Place effects in the Hebrew verbal system (Yeverechyahu 2012). The results indicate that Hebrew speakers are sensitive to the OCP-Place in the verbal system, and that sharing place features (i.e. homorganic consonants) is a sufficient condition for the violation.

7.1.1 *Participants*

33 participants took part in the experiment (18 females, 15 males). All of them were native Hebrew speakers who were born in Israel, between ages 21 and 29 (mean age 25, SD=2.21). None of them had studied Linguistics academically. Two additional participants whose mean RTs were greater than the total participants' mean in more than two standard deviations were discarded from the analysis (M_{total}=1206ms, SD=325).

7.1.2 *Stimuli*

The stimuli were 30 Hebrew verbs and 30 nonce-verbs in a legal Hebrew verb template. All items were in *binyan pi'el*, 3rd person, masculine, singular, in past tense, namely in the template *CiCeC* (e.g. *gisem* nonce-verb). In all verbs, C_2 and C_3, as well as C_1 and C_3, did not share place of articulation or manner of articulation, in order to focus on the OCP effect in C_1 and C_2 alone. In addition, the stimuli included only regular verbs (*shlemim*, see Zadok 2012 and §6.1) in which all three stem consonants appear synchronically throughout the paradigm, and nonce-verbs that look like regular verbs. The stimuli were selected with the aid of Barkali's (1964) Hebrew verbs dictionary.

The 30 nonce-verbs were divided to five groups, as follows (a full stimuli list appears in Appendix F):

a. Non-shared features: C_1 and C_2 were different in place, manner and voice (e.g. *gisem*);[19]

b. Place: C_1 and C_2 shared the place of articulation (coronals) and differed in manner and voice (e.g. *disem*);

c. Place and Manner: C_1 and C_2 shared place and manner of articulation (coronal fricatives-stridents) and differed in voice (e.g. *zisem*);

[19] Sonority was not taken into account.

d. Place and Voice: C_1 and C_2 shared place of articulation and voice (voiceless coronals) and differed in manner (e.g. *tisem*);
e. Identical C_1-C_2: C_1 and C_2 are identical (e.g. *sisem*).

In all the verbs that shared one or more features, the shared features were constant (coronal for place, stridency for manner, and voiceless for voice).[20] Referring to Frisch et al.'s (2004) model, this was done in order to avoid differences in the similarity between different groups (i.e. two coronals may be less similar to each other than two labials, since there are more natural classes of coronals in the language than of labials).

Since the study focuses on auditory similarity and not visual similarity, all the stimuli were presented auditorily. This was done in order to focus on the auditory channel, and avoid orthographic or visual influence. The stimuli were recorded by a 30-year-old male native speaker of Hebrew.

7.1.3 Procedure

The experiment was designed using the E-prime software (2.0). The participants sat in front of an Asus *Eee* mini-laptop equipped with earphones, and heard different stimuli. They were asked to determine whether each stimulus was an existing Hebrew verb. "Existing verb" responses were given by pressing *1* and "nonce-verb" responses by pressing *0*, such that opposing responses were made using different fingers and hands.

A short training block was passed at the beginning of the experiment to ensure that the participants understood the task. The training block contained ten stimuli (five existing verbs, five nonce-verbs), and the participants were given feedback (a smiley or a disappointed face icon) following each response.

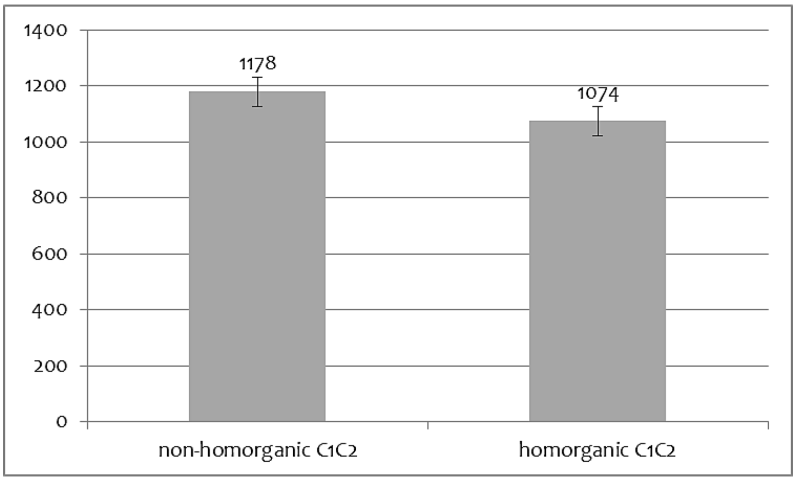
[20] Except for two nonce-verbs with identical segments, which were affricates and not fricatives (ʦ), in order to reach six different nonce-verbs.

The order of the items in the experiment and in the training was randomized across subjects. Accuracy and reaction times (hereinafter: RTs) measured from stimulus onset were collected.[21] Each subject was tested individually, and each session lasted approximately five minutes.

7.1.4 Results

The results suggest that the subjects are sensitive to the OCP-Place constraint. Accuracy for all nonce-verbs was extremely high (99.19% correct answers), errors were excluded from the RT analysis. RTs for the non-homorganic C_1-C_2 among the nonce-verbs were significantly greater than the RTs for the homorganic C_1-C_2 ($t(34)=5.99$, $p<0.0001$), as can be seen in Figure (13). Thus, the subjects needed more time to decide that non-homorganic C_1-C_2 nonce-verbs were not part of their lexicon. This finding suggests that when C_1 and C_2 are homorganic, the gap in the lexicon is systematic and predicted by the OCP-Place constraint. The OCP-Place provides the subjects with a cue that these verbs are less likely to be Hebrew verbs.

(13) Mean RT: homorganic\non-homorganic C_1-C_2 (ms)



[21] An ANOVA test reveals that the sound samples were not statistically different in duration ($F(5,24)=2.36$, $p=0.08$).

However, the differences among the five stimulus groups which violated OCP-Place to different degrees were not significant ($F(3,128)=0.61$, $p=0.61$). It seems that sharing the place feature, namely homorganic C_1 and C_2, is a sufficient condition to determine quickly that the stimulus is a nonce-verb. This finding is compatible with McCarthy's (1979, 1981, 1986) claim that the shared place of C_1 and C_2 causes OCP-Place violation in Semitic stems.

7.1.5 Discussion

The lexical decision experiment sheds light on the role of the OCP in the Hebrew verbal system, and opens the door for further research on the topic. Particularly, the experiment focuses on homorganic consonants (as in Frisch et al. 2004), and it raises the question of whether co-occurrence restrictions will also be observed among non-homorganic consonants that do share some features (voice or manner). For example, in the nonce-verb *dibem*, C_1 *d* and C_2 *b* share voice ([+voice]) and manner (stops), but not place (coronal and labial, respectively). What will be the effect of the similarity between C_1 and C_2 in this case? Second, the experiment focused on division to place, manner and voice, without looking into phonological features (as [±son], [±cont] etc.) and natural classes. Since co-occurrence restrictions may be phonological by nature (and not purely phonetic), it is interesting to address the issue from a more phonological point of view, which takes into account phonological features and natural classes. Frisch et al.'s (2004) model is based on such properties, and therefore will be suitable for analysis the results.

7.2 Word-Likelihood Judgment Experiment

In light of the results of the lexical decision experiment, a word-likelihood judgment experiment was conducted. The aim of this experiment is to broaden the lexical decision experiment by considering any C_1-C_2 combination, and comparing the results to the similarity scale.

7.2.1 *Participants*

138 participants participated in the experiment (79 females, 59 males). All of them were Hebrew native speakers who were born in Israel, between ages 20 and 40 (mean age 27, SD=3.84). 14 of them were BA Linguistics students, but none of them had taken advanced courses in morphology or phonology.

7.2.2 *Stimuli*

The stimuli in the experiment were 331 Hebrew nonce-verbs in *binyan kal*, 3sg past, in the template of *CaCaC* (e.g. *dadam*). All of them are non-existing verbs in Hebrew. However, they were put in a verb template in order to cue the participants to consider them as potential Hebrew verbs, so I expected that phonological factors (and not morphological) would affect the participants' judgments.

In order to make the experiment in a reasonable length, I focused only on one verb class, *binyan kal*. Testing all the consonant combinations in both *kal* and *pi'el* would have made the experiment too long, and would have made it harder for the participants to be focused during the entire session. Since both *kal* and *pi'el* showed sensitivity to similarity effects in the lexical analysis, I decided to focus in the current study on *kal* alone. I preferred *kal* over *pi'el* since the spirantization in the *kal*'s paradigm (see Appendix C) allows us to test more consonants, as there are no *f* or *v* in the base forms of binyan *pi'el*.[22]

As with the lexical research, the experiment focuses on the similarity between C_1 and C_2, while ignoring C_3. The stimuli's stem consonants were selected in the following way:

[22] Except for several nominatives, for example *fiʃel* 'to screw up' from the word *faʃla* 'mistake' (borrowed from Arabic).

C_1 and C_2: As in the lexical analysis, the stimuli included all Hebrew consonants in the *shlemim* paradigm (Zadok 2012), with the correct spirantization restrictions (see §6.1). Thus, the consonants *b* and *p* appeared as stops in C_1 and as fricatives *v* and *f* in C_2 (respectively). For example, stimuli contained nonce-verbs such as ***b**a**f**at*, but not ***f**a**b**at*. *k* and *x* could appear in both positions: in C_1 (non-spirant position) *k* represents alternating *k* or non-alternation *k* (historical *q*), and *x* represents only non-alternating *x* (historical *ħ*). In C_2 (spirant position), *k* represents only non-alternating *k*, and *x* represents alternating *x* or non-alternation *x*. Overall, 17 different consonants were examined, the same as in the similarity scale and the lexicon analysis.[23] In this way, the relations between the similarity scale, the lexical research, and the experiment results can be tested.

In order to make the experiment shorter, the stimuli contained each consonant pair in only one order. For example, the pair *k* and *l* appeared as *k-l* and not as *l-k*. The order of each pair was chosen to be the one in which there are more verbs, as found in the lexical research. For example, there are 17 occurrences of *k-l* and 10 occurrences of *l-k*, and therefore the stimuli in the experiment contained *k-l* and not *l-k*. In many cases, the more frequent order in the lexicon was the order in which C_1 is less sonorous than C_2. This is an interesting point that may indicate the role of sonority in co-occurrence restrictions, and it opens a door for further research. Based on this observation, for pairs that have the same number of occurrences in the lexicon in each order (usually 0), I chose the order in which C_1 was less sonorous than C_2.

C_3: The role of C_3 and its similarity to C_1 and C_2 was not tested in this study. Nevertheless, it is plausible that C_3 influences co-occurrences as well, and therefore it

[23] Except for *j*, which was considered in the similarity scale since it is a Hebrew consonant, but was eliminated from the lexical analysis and the experiment since it causes changes in the paradigm.

should be neutralized. In order to control for the effect of C_3, each C_1-C_2 pair was combined with three different C_3s, such that C_3 would be the least similar to C_1-C_2.

Based on previous studies, the control on C_3 was based on place of articulation (see McCarthy 1979, 1981, 1986; Frisch et al. 2004 among others for OCP-Place violations in Semitic languages). C_3s were selected in the following way:

a. If C_1 and C_2 did not share a major place of articulation (labial, coronal or dorsal), C_3 was in the third place of articulation. Note that C_3 could not be *b* or *p* since it is a spirant position. For example, C_3s for the pair *b* (labial) and *d* (coronal) were dorsals.

b. If C_1 and C_2 shared a major place of articulation, C_3 was chosen in the following way: if C_1 and C_2 were labials, the chosen C_3s were dorsals; if C_1 and C_2 were coronals, the chosen C_3s were labials; if C_1 and C_2 were dorsals, the chosen C_3s were coronals.

c. When this strategy did not result in enough suitable candidates for C_3 (see next paragraph for reasons for candidate elimination), I chose a different place of articulation, such that C_2 and C_3 would not be at the same place.

No stimuli were part of the Hebrew verbal lexicon. Moreover, they did not share stem consonants with existing verbs in different classes (*binyanim*), for example nonce-verbs such as *baʁaʃ* (shared stem consonants with *hivʁiʃ* 'to brush X') and *bakaʃ* (shared stem consonants with *bikeʃ* 'to request') were not included in the experiment. Stimuli that obeyed these conditions but formed existing words in Hebrew (e.g. *pagaz* 'shell (projectile)') were also excluded.

In cases in which three different options for C_3 could not be found, one form was repeated in order to obtain three stimuli for each C_1-C_2 pair. Overall, the stimuli contained 147 tested consonant pairs C_1-C_2, forming 331 different nonce-verbs, and

441 stimuli including repetitions. In order to make the experiment shorter, each participant was exposed to a random list of 49 stimuli (in random order as well), in which each tested pair of segments was presented at most once.

The nonce-verbs were inserted into frame sentences in the template of *male proper name + verb + an animal*. For example, *xen* ***batsag*** *et ha-tanin* 'Chen *batsag*-ed (nonce-verb) the crocodile'. Inserting the nonce-verbs into frames had two main reasons: first, the template of *CaCaC* is used both for verbs and nouns, for example *pagaz* 'shell (projectile)', *zamaʁ* 'singer'. Combining the nonce-words in sentences ensures that the participants would refer to the nonce-word as a verb and not as a noun. Second, the sentences were intended to make the experiment more interesting than presenting verbs in isolation, and I hoped that it would make participants more attentive to the task. The fixed template was aimed at controlling for semantic effects on the judgments. The stimuli were recorded by a 26-year-old female native speaker of Hebrew (the author), a full stimuli list appears in Appendix G.

7.2.3 Procedure

The experiment was conducted online via the *Qualtrics* website (*www.qualtrics.com*). Each participant heard a random list of 49 sentences, and was asked to rate the word-likelihood of each nonce-verb. The ratings were on a scale of 1 to 7, which was defined as follows (here translated to English), based on Frisch and Zawaydeh (2001):

(14) The rating scale

1 – No. The verb sounds terrible; it cannot be a valid verb in Hebrew.
3 – Not likely. The verb sounds strange; I doubt it can be a valid verb in Hebrew.
5 – Maybe. The verb sounds a bit strange, but it can possibly be a valid verb in Hebrew.
7 – Yes. The verb sounds good; it can be a valid verb in Hebrew.
2, 4 and 6 are found between these guidelines.

The instructions were shown after each question. Since I hypothesized that word-likelihood would be gradient rather than dichotomous, I chose to use a scale and not a yes\no decision in order to allow participants to express small differences in judgments (see Kawahara 2011 for a different approach).

As in the lexical decision experiment, this experiment also focuses on auditory similarity and not visual similarity. Therefore, all the sentences were presented auditorily, in order to focus on the auditory channel, and avoid orthographic or visual influence. The experiment took approximately ten minutes.

7.2.4 Results

The results were calculated as follows: first, a pre-analysis was done in order to check whether the effect of C_3 was neutralized. Second, the correlation between the results and the previous parts of the study (the lexicon analysis and the similarity scale) was tested.

7.2.4.1 Pre-analysis: The Effect of C_3

Since the study focuses on C_1 and C_2, C_3 was carefully chosen in order to reduce its influence on the results. Hence, before analyzing the effects of C_1 and C_2, I checked whether C_3 in each triplet of stimuli influenced the participants' judgments. For example, I checked whether the ratings for *bafat*, *bafad* and *bafan* were significantly different.

In order to test this possibility, a *Kruskal-Wallis* test was run on the 119 cases in which different C_3s were combined to the tested C_1-C_2 pairs. Out of the 119 cases, only 19 were found significant, namely there were 19 triplets in which combining different C_3s affected the results (see Appendix H for a full list). Note that statistically, even if the assumption that C_3 did not affect the results was true, we would expect an error of 5%, namely six cases in which C_3 affected the results.

Many of the anomalous stimuli are nonce-verbs that are very similar to real Hebrew verb, usually different only in the voicing of one consonant. For example, the nonce-verb *baʁas* is similar to the Hebrew verb *paʁas* 'to spread/slice', the nonce-verb *batsax* is similar to *patsax* 'to start' and the nonce-verb *padak* is similar to *badak* 'to check'. This finding implies another type of similarity that affects judgments: Not only similarity between consonants is relevant to grammar, but also similarity on a higher level, between words. This finding is compatible with Frisch and Zawaydeh (2001) findings in Arabic.

Next, I converted the average rating for each triplet into a single value. The 19 cases in which a significant difference was found inside the triplets were eliminated from the calculations.

7.2.4.2 Observations

A few interesting observations can be made by looking at the bottom of the results list, namely on the items that were rated the lowest on the word-likelihood scale (mean ratings for all tested consonant pairs appear in Appendix I). Table (15) presents the consonant pairs that received the ten lowest ratings, their similarity value on the similarity scale and their number of occurrences in the lexicon.

(15) the ten lowest ranked consonant pairs (in brackets: ranking on the similarity list, out of 66)[24]

	Pair	Rating	Similarity	Occurrences
1	s-s	2.05	1	0
2	ʦ-s	2.21	0.414 (7)	0
3	k-g	2.27	0.462 (5)	0
4	ʃ-ʃ	2.29	1	0
5	z-z	2.51	1	0
6	t-d	2.53	0.5 (3)	0
7	t-ʦ	2.60	0.7 (1)	6 (2 verbs)
8	b-v	2.65	0.313 (10)	0
8	ʦ-ʁ	2.65	0 (66)	51
9	s-z	2.66	0.52 (2)	0
9	k-k	2.66	1	1 (*kikev*)
10	g-g	2.67	1	0

A few observations can be made. All the items (except one) are rated among the ten highest similar pairs on the similarity scale, share the major place feature, and have zero, or almost zero, occurrences in the lexicon. The correlation between the parameters is salient. The only pair among the most similar tested pairs that does not appear among the ten lowest pairs in the experiment is *ʃ-s* (similarity value: 0.5), with a mean result of 3.15 (placed 22).

The only pair that does not follow this generalization is *ʦ-ʁ*, which has a similarity value of 0 and 51 occurrences in the lexicon. An explanation for this exception can be made by examining the chosen C_3s for this pair: the three items were *ʦaʁaʃ, ʦaʁas* and *ʦaʁaz,* all C_3s are stridents as C_1 *ʦ*. A key condition in this experiment was that all stimuli would be nonce-verbs, and since the pair *ʦ-ʁ* has a large number of occurrences in the lexicon, there are not many options left for nonce-verb stimuli. Along

[24] Identical consonants had similarity value of 1, and were not considered in the ranking.

the lines of the scheme presented in §7.2.2, all the chosen C_3s were stridents. Therefore, the low ratings for this pair could have been influenced by the similarity between C_1 and C_3, and not only by the relations between C_1 and C_2.

Nevertheless, not all the most similar pairs nor pairs with zero occurrences in the lexicon appear at the top of the experiment results. Looking at the other results reveals that all the pairs with identical C_1-C_2 are rated among the lowest 20. For example, *l-l*, *m-m* and *n-n* are rated 11th, 12th and 13th respectively, and *x-x* is rated 19th.[25] Note that these are exactly the identical pairs that do have occurrences in the lexicon (together with *k-k*), and *x-x* had the highest number of occurrences, 15, affected by historical reasons (see §6.2.2 for discussion). This finding can show the correlation between the lexicon and the experiment, but not to point to the source of the influence. It could be that the small-but-not-zero occurrences in the lexicon affect the speakers' phonological system, or that these identical pairs show (for some reasons) fewer restrictions, and therefore are more flexible both in the lexicon and in the speakers' judgments. Nonetheless, the differences between the pairs of identical C_1-C_2 are small, and it can be concluded that all of them show co-occurrence restrictions in the speakers' judgments.

Next, I examined the other pairs that had zero occurrences in the lexicon. Table (16) summarizes the comparison.

[25] The other pairs with identical C_1-C_2 were eliminated from the results according to the pre-analysis of the effect of C_3 (§7.2.4.1).

(16) Other pairs with zero occurrences in the lexicon (in brackets: the rank on the relevant scale)[26]

Pair	Rating
p-f	2.81 (16)
b-f	2.93 (18)
d-z	3.00 (20)
b-m	3.02 (21)
p-v	3.21 (24)
n-l	3.42 (27)
m-f	3.43 (28)
m-v	3.55 (34)
ʁ-l	3.70 (38)

It can be seen that not all these pairs were rated low in the judgment task. There are pairs that appear among the second or third group of ten, but also pairs that rated 34 and 38. Note that *m-f* and *m-v*, that rated low on the similarity scale (0.063 (51)), do not show strong co-occurrence restrictions in the judgments task, as the pair *ʁ-l*, in which C_1 and C_2 do not share a place of articulation. Hence, in these cases the similarity factors are stronger than the occurrences in the lexicon.

7.2.4.3 Comparisons to the Scales

Next, a statistical analysis was done in order to examine the correlation between the experiment results on the one hand, and the lexicon analysis and the similarity scale on the other hand.

Comparison to the similarity scale: An *ordered logistic regression model* was built in order to test the influence of the similarity factor on the results. This factor is assumed to be negative, since the greater the similarity between two consonant is, the participants are assumed to give lower ratings to the nonce-verbs. Indeed, the similarity

[26] Not including the pairs that were eliminated due to spirantization (*b-p, f-v*) or during the pre-analysis of the effect of C_3.

factor is negative and its influence is significant (Estimate = -1.79, SD = 0.1, $p < 0.0001$). This result shows that the more similar the consonants are (their similarity value is closer to 1), the lower the ratings the participants will give to the word-likelihood of the nonce-verb.

Comparison to the lexical analysis: An *ordered logistic regression model* was built in order to test the influence of the frequency factor on the results. This factor was assumed to be positive, since the more frequent the consonant pair in the lexicon is, the participants are assumed to give higher ratings to the nonce-verb. Indeed, the frequency factor is positive and its influence is significant (Estimate = 0.04, SD = 0.004, $p < 0.0001$). This result shows that nonce-verb containing frequent pairs of consonants receive higher ratings of word-likelihood.

These results show a strong correlation between the word-likelihood ratings and the similarity scale and with the lexical analysis. Since a correlation was found between the lexical analysis and the similarity scale as well, it is not surprising that the participants' ratings correlate with both of them.

CHAPTER 8: GENERAL DISCUSSION

The study examines the co-occurrence restrictions in the Hebrew verbal system, focusing on the contribution of similarity between consonants to these restrictions. The phenomenon was tested on two levels: lexical analysis and psycholinguistic experiments. The results suggest that there are co-occurrence restrictions on stem consonants C_1 and C_2, both in the lexicon and in the speakers' phonological system. The similarity factor was found to be significant in all the tested cases.

The similarity model used in this study is based on Frisch et al.'s (2004) model, originally proposed for Arabic. The model is built with a phonological orientation: it is based on phonological natural classes and phonological features. Note that the model itself is not unique for linguistic similarity and can be adapted to similarity in any domain; it is the phonological features that make the similarity model language oriented. Since a correlation was found between the co-occurrence restrictions and the similarity model, the study supports the idea that phonological features can constitute a proper base for similarity calculations. However, the fact that there are articulatory-based similarity effects does not mean that acoustic factors do not play a role as well (see Mielke 2009). It would be interesting to examine the interaction between acoustic factors and phonological factors, and this is a window for further studies in the field.

Previous studies on Semitic languages (Greenberg 1950; McCarthy 1981, 1986; Frisch et al. 2004 among others) demonstrated the important role of OCP-Place in co-occurrence restrictions in those languages. The current study strengthens this claim; both the lexical analysis and the experiments suggest that consonants that share the major place feature are less likely to co-occur. However, the results and the correlation with the similarity scale show that not only place feature has a role in co-occurrence restrictions, and correlations were found also between the similarity scale and occurrences or judgments of pairs that do not share place of articulation. It is likely that the major place feature has a great weight in similarity, inter alia due to its high position in the feature geometry hierarchy (Clements 1985, Sagey 1986, Clements and Hume 1995, and see also Kaisse 1988 and Padgett 1995). Frisch et al.'s model does not suggest which phonological features are more important for similarity, and the current findings open a window for further research in the field.

The results of the lexical analysis and the experiment suggest a correlation between the co-occurrence restrictions and the similarity scale. These correlations are statistical, and do not entail causal influence of similarity on the restrictions.

The lexical analysis shows correlation to the similarity scale. Nonetheless, it is plausible to assume that other factors have influenced the lexicon as well; some of them are historical influences that do not have transparent evidence nowadays. For example, the 15 occurrences of *x*-*x* in the lexicon: historically, the origin of *x* is double, one is a result of historical *ħ* (which is assumed to have been pronounced farther backward than the synchronic *x*) and the other is a spirantized *k*. Indeed, all the 15 cases of co-occurrences are results of the historical reasons. The naïve Hebrew speakers may not know this detail, but the frequency of *x* in their lexicon and the cases of verbs with *x* in C_1 and C_2 are likely to affect their phonological system. Thus, the correlation between the lexical analysis and the similarity scale may suggest that similarity is one of the factors that influence the lexicon. However, it is not the only one.

The speakers' word-likelihood judgments in the experiment were correlated both with the similarity scale and the lexical analysis results. It is not surprising that the experiment's results are correlated with both of them, because there is also a correlation between the similarity scale and the lexical analysis results. Therefore, the experiment cannot suggest whether the influence of similarity is direct, or indirect through the lexical influences on the grammatical system (figure 17).

(17)	*lexicon* ← ***similarity*** → *grammatical system*
	or:
	similarity → *lexicon* → *grammatical system*

CHAPTER 9: CONCLUDING REMARKS

The study shows that there are co-occurrence restrictions both in the Hebrew lexicon and in the grammatical system of the speakers. These restrictions have a strong correlation with similarity between consonants, where the tendency is to avoid similar, close consonants. The results suggest that similarity affects the speakers' word-likelihood judgments, but they cannot tell whether the effect is direct or indirect through the lexicon.

The study opens the door to further research on similarity effects in Hebrew and in other languages. First, the current study examines the issue of similarity from a phonological point of view, and successfully shows the relevance of phonological properties to the subject. Future studies should examine the influence of each phonological feature individually, and examine which features are more important to similarity. Under the observed importance of the place of articulation, and previous studies on OCP-Place in Semitic languages (McCarthy 1979, 1981, 1986; Frisch et al. 2004 among others), there is a reason to assume that major place features have a large role in similarity.

Second, the study took into account phonological features that are mostly based on articulatory factors, but not acoustic parameters per-se. However, it has been claimed that acoustic factors play a significant role in similarity as well (Kawahara 2007, Mielke 2009). Models that combine acoustic parameters (exclusively or with articulatory parameters) and test their predictions regarding similarity in Hebrew should be taken into consideration in further studies.

Third, the lexical analysis does not show statistical differences between orders of consonants in the pairs (e.g. *k*-*d* vs. *d*-*k*), but a closer look reveals large differences in some pairs (e.g. *k*-*d*: 34 occurrences, *d*-*k*: 18 occurrences). It would be interesting to examine possible causes to asymmetry in similarity pairs. Previous studies (Johnson 2012 for example) suggest that similarity is not symmetric, such that it is not necessary that *k* is similar to *d* in the same degree that *d* is similar to *k*. Frisch et al.'s (2004) model calculates similarity based on shared and non-shared natural classes, and thus cannot take symmetry into account. However, it is also plausible to assume that other factors cause these differences, such sonority or place of articulation.

Furthermore, the current study does not show differences between the two tested verb classes (*binyanim*) *kal* and *pi'el*: They both showed the same tendency to avoid similar consonant pairs in the lexicon. However, there is a solid basis to assume that there are differences in co-occurrence restrictions between them. For one, the proximity between C_1 and C_2 is different between the templates throughout the paradigms: in *pi'el* they are separated with a vowel through the different tenses and persons, while in *kal* there are forms in which they are adjacent (e.g. future tense: *ji**xt**ov* 'he will write'). Since previous studies suggest that proximity plays a role in co-occurrence restrictions (Rose 2000 among others), and under the assumption that speakers have access to the paradigms and not only to the base forms, differences in restrictions are expected. Since the lexicon did not show significant differences, examination of this issue from a psycholinguistic point of view in needed.

In addition, the current study focuses only on some verb classes and does not deal with nouns at all. It would be interesting to compare co-occurrence restrictions between nouns and verbs, and between two types of Hebrew nouns: Semitic nouns that have similar structure to verbs, with stems and templates (e.g. *ʃmiʁa* 'saving', template: *CCiCa*) and mono-morphemic, non-Semitic nouns (e.g. *ʃulxan* 'table'). It will also be interesting to expand the study to other verb classes (*binyanim*), and look for differences among them.

Finally, the study shows a correlation between the lexical analysis and the speakers' judgments. It would be interesting to further examine the influence of the lexicon on one's grammatical knowledge, and investigate the tension between universal constraints and language specific grammar. Since the current study found correlation between the similarity scale and the lexicon analysis, it cannot determine the source of the influence; a language in which there is no correlation between the lexicon and the universal constraints could shed more light on the influence of the lexicon.

APPENDIX

Appendix A: Natural classes of Hebrew consonants

[+son]	m	n	l	j	ʁ								
[+son, LAB]	m												
[+son, LAB, -cont]	m												
[+son, COR]	n	l	j										
[+son, COR, +ant]	n	l											
[+son, COR, +ant, +cont]	l												
[+son, COR, +ant, -cont]	n												
[+son, COR, -ant]	j												
[+son, COR, -ant, +cont]	j												
[+son, COR, +cont]	l	j											
[+son, COR, -cont]	n												
[+son, DOR]	ʁ												
[+son, DOR, +cont]	ʁ												
[+son, +cont]	l	j	ʁ										
[+son, -cont]	m	n											
[-son]	p	b	f	v	t	d	s	z	ʦ	ʃ	k	g	x
[-son, LAB]	p	b	f	v									
[-son, LAB, +cont]	f	v											
[-son, LAB, +cont, +voice]	v												
[-son, LAB, +cont, -voice]	f												
[-son, LAB, -cont]	p	b											
[-son, LAB, -cont, +voice]	b												
[-son, LAB, -cont, -voice]	p												
[-son, LAB, +voice]	b	v											
[-son, LAB, -voice]	p	f											
[-son, COR]	t	d	s	z	ʦ	ʃ							
[-son, COR, +strident]	s	z	ʦ	ʃ									
[-son, COR, +strident, +ant]	s	z	ʦ										
[-son, COR, +strident, +ant, +cont]	s	z											
[-son, COR, +strident, +ant, +cont, +voice]	z												
[-son, COR, +strident, +ant, +cont, -voice]	s												
[-son, COR, +strident, +ant, -cont]	ʦ												
[-son, COR, +strident, +ant, -cont, -voice]	ʦ												
[-son, COR, +strident, +ant, -voice]	s	ʦ											
[-son, COR, +strident, -ant]	ʃ												
[-son, COR, +strident, -ant, +cont]	ʃ												
[-son, COR, +strident, -ant, +cont, -voice]	ʃ												
[-son, COR, +strident, +cont]	s	z	ʃ										
[-son, COR, +strident, +cont, -voice]	s	ʃ											
[-son, COR, +strident, -voice]	s	ʦ	ʃ										

[-son, COR, +ant]	t	d	s	z	ʦ				
[-son, COR, +ant, -cont]	t	d	ʦ						
[-son, COR, +ant, -cont, +voice]	d								
[-son, COR, +ant, -cont, -voice]	t	ʦ							
[-son, COR, +ant, +voice]	d	z							
[-son, COR, +ant, -voice]	t	s	ʦ						
[-son, COR, -voice]	t	s	ʦ	ʃ					
[-son, DOR]	k	g	x						
[-son, DOR, +cont]	x								
[-son, DOR, +cont, -voice]	x								
[-son, DOR, -cont]	k	g							
[-son, DOR, -cont, +voice]	g								
[-son, DOR, -cont, -voice]	k								
[-son, DOR, -voice]	k	x							
[-son, +cont]	f	v	s	z	ʃ	x			
[-son, +cont, +voice]	v	z							
[-son, +cont, -voice]	f	s	ʃ	x					
[-son, -cont]	p	b	t	d	ʦ	k	g		
[-son, -cont, +voice]	b	d	g						
[-son, -cont, -voice]	p	t	ʦ	k					
[-son, +voice]	b	v	d	z	g				
[-son, -voice]	p	f	t	s	ʦ	ʃ	k	x	
[LAB]	p	b	m	f	v				
[LAB, -cont]	p	b	m						
[COR]	t	d	s	z	ʦ	ʃ	n	l	j
[COR, +ant]	t	d	s	z	ʦ	n	l		
[COR, +ant, +cont]	s	z	l						
[COR, +ant, -cont]	t	d	ʦ	n					
[COR, -ant]	ʃ	j							
[COR, -ant, +cont]	ʃ	j							
[COR, +cont]	s	z	ʃ	l	j				
[DOR]	k	g	x	ʁ					
[DOR, +cont]	x	ʁ							
[+cont]	f	v	s	z	ʃ	l	j	x	ʁ
[-cont]	p	b	m	t	d	ʦ	n	k	g

Appendix B: The similarity scale (based on Frisch et al.'s 2004 model)

bl	0.000	bx	0.048	tl	0.087	kʦ	0.200	tʦ	0.700
bʁ	0.000	xl	0.048	xt	0.087	nt	0.200	bb	1.000
bj	0.000	xj	0.048	dl	0.091	pm	0.200	dd	1.000
dʁ	0.000	bn	0.050	tf	0.091	ʁl	0.200	ff	1.000
gl	0.000	fj	0.050	dv	0.095	ʁj	0.200	gg	1.000
gj	0.000	lf	0.050	px	0.100	xʁ	0.200	kk	1.000
kl	0.000	lv	0.050	sj	0.103	ʃx	0.208	ll	1.000
kj	0.000	pn	0.050	sv	0.107	ʃj	0.208	mm	1.000
mʃ	0.000	tm	0.050	bʦ	0.107	nd	0.211	nn	1.000
mz	0.000	vj	0.050	ʃd	0.107	nm	0.214	pp	1.000
nf	0.000	dm	0.053	kf	0.111	ʃf	0.217	ʁʁ	1.000
nv	0.000	nk	0.053	gʦ	0.115	zl	0.217	ss	1.000
pl	0.000	kv	0.053	ʃl	0.115	ʃʦ	0.226	ʃʃ	1.000
pʁ	0.000	gf	0.056	gv	0.118	zv	0.227	tt	1.000
pj	0.000	ng	0.056	ʃv	0.120	ʦz	0.233	ʦʦ	1.000
sm	0.000	ml	0.059	xz	0.120	pt	0.250	vv	1.000
tʁ	0.000	mj	0.059	zj	0.120	kt	0.263	xx	1.000
ʦʁ	0.000	kʁ	0.063	zf	0.125	bd	0.263	jj	1.000
xm	0.000	mf	0.063	bt	0.136	gd	0.294	zz	1.000
xn	0.000	mv	0.063	pd	0.143	st	0.296		
bs	0.032	ʁf	0.063	kd	0.150	ʃz	0.296		
ʦj	0.033	ʁn	0.063	tg	0.150	dz	0.304		
sg	0.034	km	0.063	ʦn	0.154	bv	0.313		
ʦv	0.034	ʁv	0.063	bf	0.167	nl	0.313		
bʃ	0.036	gm	0.067	bk	0.167	pf	0.313		
nʃ	0.037	gʁ	0.067	nj	0.167	pk	0.313		
pz	0.037	ps	0.067	pv	0.167	xf	0.313		
sʁ	0.037	sn	0.069	xv	0.167	xk	0.313		
ʃg	0.038	ʦl	0.069	sd	0.172	bg	0.333		
ʦm	0.038	ʦx	0.069	gx	0.176	ʦd	0.375		
zk	0.038	ks	0.069	pg	0.176	ʦs	0.414		
ʃʁ	0.042	ʦf	0.071	ʃt	0.185	fv	0.429		
tj	0.042	pʃ	0.074	sl	0.185	kg	0.462		
tv	0.043	bz	0.077	sx	0.185	bp	0.467		
dx	0.043	mʁ	0.077	pʦ	0.192	lj	0.467		
dj	0.043	kʃ	0.077	sf	0.192	ʃs	0.500		
zʁ	0.043	nz	0.080	tz	0.192	td	0.500		
df	0.045	gz	0.083	bm	0.200	sz	0.520		

Appendix C: Sample derivations

	kal (3rd ms. sg.)				*pi'el* (3rd ms. sg.)			
	Past	Present	Future		Past	Present	Future	
Regular Verbs	*ʃamaʁ*	*ʃomeʁ*	*jiʃmoʁ*	'save'	*ʃimeʁ*	*meʃameʁ*	*jeʃameʁ*	'preserve'
C_1: spirantization	***k**atav*	***k**otev*	*ji**x**tov*	'write'	***k**itev*	*me**x**atev*	*je**x**atev*	'subscribe'
C_2: spirantization	*sa**f**aʁ*	*so**f**eʁ*	*jis**p**oʁ*	'count'	*si**p**eʁ*	*mesa**p**eʁ*	*jesa**p**eʁ*	'tell'

Appendix D: Lexical analysis results

kal

bb	0	bb	0	sum	0
bd	8	db	4	sum	12
bf	0	fb	0	sum	0
bg	4	gb	4	sum	8
bk	2	kb	4	sum	6
bl	12	lb	4	sum	16
bm	0	mb	0	sum	0
bn	0	nb	3	sum	3
bp	0	pb	0	sum	0
bʁ	16	ʁb	5	sum	21
bs	6	sb	5	sum	11
bʃ	2	ʃb	8	sum	10
bt	6	tb	2	sum	8
bʦ	6	ʦb	3	sum	9
bv	0	vb	0	sum	0
bx	8	xb	11	sum	19
bz	6	zb	3	sum	9
dd	0	dd	0	sum	0
df	6	fd	0	sum	6
dg	6	gd	15	sum	21
dk	6	kd	19	sum	25
dl	15	ld	0	sum	15
dm	6	md	3	sum	9
dn	0	nd	10	sum	10
dp	3	pd	0	sum	3
dʁ	12	ʁd	9	sum	21
ds	0	sd	9	sum	9
dʃ	3	ʃd	9	sum	12

pi'el

bb	0	bb	0	sum	0
bd	4	db	9	sum	13
bf	0	fb	0	sum	0
bg	0	gb	18	sum	18
bk	3	kb	21	sum	24
bl	4	lb	18	sum	22
bm	0	mb	0	sum	0
bn	0	nb	9	sum	9
bp	0	pb	0	sum	0
bʁ	6	ʁb	15	sum	21
bs	5	sb	15	sum	20
bʃ	1	ʃb	15	sum	16
bt	5	tb	9	sum	14
bʦ	2	ʦb	6	sum	8
bv	0	vb	0	sum	0
bx	5	xb	36	sum	41
bz	0	zb	9	sum	9
dd	0	dd	0	sum	0
df	0	fd	2	sum	2
dg	12	gd	15	sum	27
dk	12	kd	15	sum	27
dl	9	ld	0	sum	9
dm	6	md	9	sum	15
dn	3	nd	9	sum	12
dp	6	pd	1	sum	7
dʁ	6	ʁd	6	sum	12
ds	3	sd	9	sum	12
dʃ	3	ʃd	9	sum	12

kal and pi'el

bb	0	bb	0	sum	0
bd	12	db	13	sum	25
bf	0	fb	0	sum	0
bg	4	gb	22	sum	26
bk	5	kb	25	sum	30
bl	16	lb	22	sum	38
bm	0	mb	0	sum	0
bn	0	nb	12	sum	12
bp	0	pb	0	sum	0
bʁ	22	ʁb	20	sum	42
bs	11	sb	20	sum	31
bʃ	3	ʃb	23	sum	26
bt	11	tb	11	sum	22
bʦ	8	ʦb	9	sum	17
bv	0	vb	0	sum	0
bx	13	xb	47	sum	60
bz	6	zb	12	sum	18
dd	0	dd	0	sum	0
df	6	fd	2	sum	8
dg	18	gd	30	sum	48
dk	18	kd	34	sum	52
dl	24	ld	0	sum	24
dm	12	md	12	sum	24
dn	3	nd	19	sum	22
dp	9	pd	1	sum	10
dʁ	18	ʁd	15	sum	33
ds	3	sd	18	sum	21
dʃ	6	ʃd	18	sum	24

kal

dt	0	td	0	sum	0
dʦ	0	ʦd	3	sum	3
dv	8	vd	4	sum	12
dx	15	xd	11	sum	26
dz	0	zd	0	sum	0
ff	0	ff	0	sum	0
fg	4	gf	0	sum	4
fk	7	kf	24	sum	31
fl	4	lf	6	sum	10
fm	0	mf	0	sum	0
fn	1	nf	6	sum	7
fp	0	pf	0	sum	0
fʁ	11	ʁf	12	sum	23
fs	6	sf	16	sum	22
fʃ	4	ʃf	12	sum	16
ft	8	tf	22	sum	30
fʦ	3	ʦf	8	sum	11
fv	0	vf	0	sum	0
fx	6	xf	14	sum	20
fz	4	zf	2	sum	6
gg	0	gg	0	sum	0
gk	0	kg	0	sum	0
gl	18	lg	6	sum	24
gm	15	mg	6	sum	21
gn	12	ng	13	sum	25
gp	0	pg	8	sum	8
gʁ	27	ʁg	18	sum	45
gs	3	sg	9	sum	12
gʃ	9	ʃg	12	sum	21
gt	0	tg	0	sum	0
gʦ	0	ʦg	0	sum	0
gv	8	vg	2	sum	10
gx	9	xg	6	sum	15
gz	12	zg	0	sum	12
kk	0	kk	0	sum	0
kl	17	lk	10	sum	27
km	26	mk	3	sum	29
kn	10	nk	21	sum	31
kp	6	pk	12	sum	18

pi'el

dt	0	td	0	sum	0
dʦ	0	ʦd	6	sum	6
dv	0	vd	8	sum	8
dx	9	xd	18	sum	27
dz	0	zd	0	sum	0
ff	0	ff	0	sum	0
fg	6	gf	0	sum	6
fk	10	kf	0	sum	10
fl	14	lf	0	sum	14
fm	0	mf	0	sum	0
fn	6	nf	0	sum	6
fp	0	pf	0	sum	0
fʁ	4	ʁf	0	sum	4
fs	8	sf	0	sum	8
fʃ	6	ʃf	0	sum	6
ft	12	tf	0	sum	12
fʦ	4	ʦf	0	sum	4
fv	0	vf	0	sum	0
fx	12	xf	0	sum	12
fz	6	zf	0	sum	6
gg	0	gg	0	sum	0
gk	0	kg	0	sum	0
gl	9	lg	0	sum	9
gm	15	mg	9	sum	24
gn	6	ng	18	sum	24
gp	12	pg	3	sum	15
gʁ	15	ʁg	18	sum	33
gs	0	sg	15	sum	15
gʃ	9	ʃg	15	sum	24
gt	3	tg	6	sum	9
gʦ	3	ʦg	0	sum	3
gv	0	vg	0	sum	0
gx	6	xg	3	sum	9
gz	6	zg	3	sum	9
kk	1	kk	0	sum	1
kl	14	lk	12	sum	26
km	12	mk	15	sum	27
kn	9	nk	33	sum	42
kp	12	pk	5	sum	17

kal and pi'el

dt	0	td	0	sum	0
dʦ	0	ʦd	9	sum	9
dv	8	vd	12	sum	20
dx	24	xd	29	sum	53
dz	0	zd	0	sum	0
ff	0	ff	0	sum	0
fg	10	gf	0	sum	10
fk	17	kf	24	sum	41
fl	18	lf	6	sum	24
fm	0	mf	0	sum	0
fn	7	nf	6	sum	13
fp	0	pf	0	sum	0
fʁ	15	ʁf	12	sum	27
fs	14	sf	16	sum	30
fʃ	10	ʃf	12	sum	22
ft	20	tf	22	sum	42
fʦ	7	ʦf	8	sum	15
fv	0	vf	0	sum	0
fx	18	xf	14	sum	32
fz	10	zf	2	sum	12
gg	0	gg	0	sum	0
gk	0	kg	0	sum	0
gl	27	lg	6	sum	33
gm	30	mg	15	sum	45
gn	18	ng	31	sum	49
gp	12	pg	11	sum	23
gʁ	42	ʁg	36	sum	78
gs	3	sg	24	sum	27
gʃ	18	ʃg	27	sum	45
gt	3	tg	6	sum	9
gʦ	3	ʦg	0	sum	3
gv	8	vg	2	sum	10
gx	15	xg	9	sum	24
gz	18	zg	3	sum	21
kk	1	kk	0	sum	1
kl	31	lk	22	sum	53
km	38	mk	18	sum	56
kn	19	nk	54	sum	73
kp	18	pk	17	sum	35

kal					
kʁ	34	ʁk	20	sum	54
ks	16	sk	16	sum	32
kʃ	25	ʃk	24	sum	49
kt	21	tk	10	sum	31
kʦ	12	ʦk	0	sum	12
kv	16	vk	1	sum	17
kx	8	xk	13	sum	21
kz	2	zk	13	sum	15
ll	0	ll	0	sum	0
lm	3	ml	18	sum	21
ln	0	nl	0	sum	0
lp	3	pl	8	sum	11
lʁ	0	ʁl	0	sum	0
ls	0	sl	12	sum	12
lʃ	3	ʃl	24	sum	27
lt	9	tl	12	sum	21
lʦ	3	ʦl	18	sum	21
lv	8	vl	6	sum	14
lx	17	xl	28	sum	45
lz	0	zl	12	sum	12
mm	0	mm	0	sum	0
mn	0	nm	0	sum	0
mp	0	pm	0	sum	0
mʁ	18	ʁm	12	sum	30
ms	12	sm	12	sum	24
mʃ	12	ʃm	9	sum	21
mt	9	tm	21	sum	30
mʦ	9	ʦm	15	sum	24
mv	0	vm	0	sum	0
mx	18	xm	28	sum	46
mz	6	zm	6	sum	12
nn	0	nn	0	sum	0
np	1	pn	2	sum	3
nʁ	0	ʁn	3	sum	3
ns	7	sn	9	sum	16
nʃ	14	ʃn	9	sum	23
nt	18	tn	6	sum	24
nʦ	8	ʦn	12	sum	20
nv	8	vn	0	sum	8

pi'el					
kʁ	0	ʁk	33	sum	33
ks	6	sk	33	sum	39
kʃ	14	ʃk	33	sum	47
kt	26	tk	21	sum	47
kʦ	9	ʦk	0	sum	9
kv	0	vk	6	sum	6
kx	3	xk	14	sum	17
kz	4	zk	15	sum	19
ll	3	ll	0	sum	3
lm	6	ml	6	sum	12
ln	0	nl	0	sum	0
lp	6	pl	7	sum	13
lʁ	0	ʁl	0	sum	0
ls	0	sl	21	sum	21
lʃ	3	ʃl	21	sum	24
lt	9	tl	18	sum	27
lʦ	0	ʦl	15	sum	15
lv	0	vl	8	sum	8
lx	6	xl	25	sum	31
lz	0	zl	6	sum	6
mm	6	mm	0	sum	6
mn	3	nm	15	sum	18
mp	0	pm	0	sum	0
mʁ	12	ʁm	6	sum	18
ms	9	sm	18	sum	27
mʃ	9	ʃm	18	sum	27
mt	18	tm	21	sum	39
mʦ	6	ʦm	18	sum	24
mv	0	vm	0	sum	0
mx	6	xm	24	sum	30
mz	6	zm	9	sum	15
nn	3	nn	0	sum	3
np	15	pn	3	sum	18
nʁ	0	ʁn	3	sum	3
ns	9	sn	15	sum	24
nʃ	21	ʃn	15	sum	36
nt	24	tn	6	sum	30
nʦ	6	ʦn	6	sum	12
nv	0	vn	0	sum	0

kal and pi'el					
kʁ	34	ʁk	53	sum	87
ks	22	sk	49	sum	71
kʃ	39	ʃk	57	sum	96
kt	47	tk	31	sum	78
kʦ	21	ʦk	0	sum	21
kv	16	vk	7	sum	23
kx	11	xk	27	sum	38
kz	6	zk	28	sum	34
ll	3	ll	0	sum	3
lm	9	ml	24	sum	33
ln	0	nl	0	sum	0
lp	9	pl	15	sum	24
lʁ	0	ʁl	0	sum	0
ls	0	sl	33	sum	33
lʃ	6	ʃl	45	sum	51
lt	18	tl	30	sum	48
lʦ	3	ʦl	33	sum	36
lv	8	vl	14	sum	22
lx	23	xl	53	sum	76
lz	0	zl	18	sum	18
mm	6	mm	0	sum	6
mn	3	nm	15	sum	18
mp	0	pm	0	sum	0
mʁ	30	ʁm	18	sum	48
ms	21	sm	30	sum	51
mʃ	21	ʃm	27	sum	48
mt	27	tm	42	sum	69
mʦ	15	ʦm	33	sum	48
mv	0	vm	0	sum	0
mx	24	xm	52	sum	76
mz	12	zm	15	sum	27
nn	3	nn	0	sum	3
np	16	pn	5	sum	21
nʁ	0	ʁn	6	sum	6
ns	16	sn	24	sum	40
nʃ	35	ʃn	24	sum	59
nt	42	tn	12	sum	54
nʦ	14	ʦn	18	sum	32
nv	8	vn	0	sum	8

kal						*pi'el*						*kal and pi'el*					
nx	19	xn	23	sum	42	nx	18	xn	21	sum	39	nx	37	xn	44	sum	81
nz	7	zn	6	sum	13	nz	6	zn	9	sum	15	nz	13	zn	15	sum	28
pp	0	pp	0	sum	0	pp	0	pp	0	sum	0	pp	0	pp	0	sum	0
pʁ	22	ʁp	6	sum	28	pʁ	2	ʁp	15	sum	17	pʁ	24	ʁp	21	sum	45
ps	12	sp	8	sum	20	ps	4	sp	9	sum	13	ps	16	sp	17	sum	33
pʃ	8	ʃp	6	sum	14	pʃ	3	ʃp	9	sum	12	pʃ	11	ʃp	15	sum	26
pt	16	tp	11	sum	27	pt	6	tp	15	sum	21	pt	22	tp	26	sum	48
pʦ	6	ʦp	4	sum	10	pʦ	2	ʦp	9	sum	11	pʦ	8	ʦp	13	sum	21
pv	0	vp	0	sum	0	pv	0	vp	0	sum	0	pv	0	vp	0	sum	0
px	14	xp	13	sum	27	px	6	xp	24	sum	30	px	20	xp	37	sum	57
pz	8	zp	1	sum	9	pz	3	zp	6	sum	9	pz	11	zp	7	sum	18
ʁʁ	0	ʁʁ	0	sum	0	ʁʁ	0	ʁʁ	0	sum	0	ʁʁ	0	ʁʁ	0	sum	0
ʁs	6	sʁ	39	sum	45	ʁs	9	sʁ	3	sum	12	ʁs	15	sʁ	42	sum	57
ʁʃ	9	ʃʁ	9	sum	18	ʁʃ	15	ʃʁ	3	sum	18	ʁʃ	24	ʃʁ	12	sum	36
ʁt	24	tʁ	24	sum	48	ʁt	21	tʁ	18	sum	39	ʁt	45	tʁ	42	sum	87
ʁʦ	12	ʦʁ	21	sum	33	ʁʦ	9	ʦʁ	9	sum	18	ʁʦ	21	ʦʁ	30	sum	51
ʁv	10	vʁ	8	sum	18	ʁv	0	vʁ	12	sum	12	ʁv	10	vʁ	20	sum	30
ʁx	28	xʁ	44	sum	72	ʁx	15	xʁ	21	sum	36	ʁx	43	xʁ	65	sum	108
ʁz	6	zʁ	15	sum	21	ʁz	0	zʁ	15	sum	15	ʁz	6	zʁ	30	sum	36
ss	0	ss	0	sum	0	ss	0	ss	0	sum	0	ss	0	ss	0	sum	0
sʃ	0	ʃs	6	sum	6	sʃ	0	ʃs	3	sum	3	sʃ	0	ʃs	9	sum	9
st	18	ts	3	sum	21	st	15	ts	3	sum	18	st	33	ts	6	sum	39
sʦ	0	ʦs	0	sum	0	sʦ	0	ʦs	0	sum	0	sʦ	0	ʦs	0	sum	0
sv	10	vs	3	sum	13	sv	0	vs	10	sum	10	sv	10	vs	13	sum	23
sx	38	xs	26	sum	64	sx	9	xs	21	sum	30	sx	47	xs	47	sum	94
sz	0	zs	0	sum	0	sz	0	zs	0	sum	0	sz	0	zs	0	sum	0
ʃʃ	0	ʃʃ	0	sum	0	ʃʃ	0	ʃʃ	0	sum	0	ʃʃ	0	ʃʃ	0	sum	0
ʃt	21	tʃ	6	sum	27	ʃt	15	tʃ	0	sum	15	ʃt	36	tʃ	6	sum	42
ʃʦ	3	ʦʃ	0	sum	3	ʃʦ	0	ʦʃ	0	sum	0	ʃʦ	3	ʦʃ	0	sum	3
ʃv	16	vʃ	1	sum	17	ʃv	0	vʃ	2	sum	2	ʃv	16	vʃ	3	sum	19
ʃx	42	xʃ	23	sum	65	ʃx	18	xʃ	19	sum	37	ʃx	60	xʃ	42	sum	102
ʃz	6	zʃ	0	sum	6	ʃz	6	zʃ	0	sum	6	ʃz	12	zʃ	0	sum	12
tt	0	tt	0	sum	0	tt	0	tt	0	sum	0	tt	0	tt	0	sum	0
tʦ	0	ʦt	0	sum	0	tʦ	0	ʦt	6	sum	6	tʦ	0	ʦt	6	sum	6
tv	4	vt	3	sum	7	tv	0	vt	10	sum	10	tv	4	vt	13	sum	17
tx	26	xt	36	sum	62	tx	9	xt	40	sum	49	tx	35	xt	76	sum	111
tz	3	zt	0	sum	3	tz	3	zt	3	sum	6	tz	6	zt	3	sum	9
ʦʦ	0	ʦʦ	0	sum	0	ʦʦ	0	ʦʦ	0	sum	0	ʦʦ	0	ʦʦ	0	sum	0
ʦv	6	vʦ	3	sum	9	ʦv	0	vʦ	4	sum	4	ʦv	6	vʦ	7	sum	13

kal					
tsx	12	xts	12	sum	24
tsz	0	zts	0	sum	0
vv	0	vv	0	sum	0
vx	4	xv	14	sum	18
vz	3	zv	6	sum	9
xx	9	xx	0	sum	9
xz	13	zx	5	sum	18
zz	0	zz	0	sum	0

pi'el					
tsx	6	xts	12	sum	18
tsz	0	zts	0	sum	0
vv	0	vv	0	sum	0
vx	10	xv	0	sum	10
vz	0	zv	0	sum	0
xx	6	xx	0	sum	6
xz	14	zx	0	sum	14
zz	0	zz	0	sum	0

kal and pi'el					
tsx	18	xts	24	sum	42
tsz	0	zts	0	sum	0
vv	0	vv	0	sum	0
vx	14	xv	14	sum	28
vz	3	zv	6	sum	9
xx	15	xx	0	sum	15
xz	27	zx	5	sum	32
zz	0	zz	0	sum	0

Appendix E: Frequencies in the lexicon

kal:

b	C_1	76	3.29%	C_2	57	2.44%	sum	133	2.86%
d	C_1	84	3.64%	C_2	102	4.36%	sum	186	4.01%
f	C_1	58	2.51%	C_2	128	5.48%	sum	186	4.01%
g	C_1	132	5.72%	C_2	96	4.11%	sum	228	4.91%
k	C_1	240	10.40%	C_2	162	6.93%	sum	402	8.66%
l	C_1	72	3.12%	C_2	204	8.73%	sum	276	5.94%
m	C_1	114	4.94%	C_2	153	6.55%	sum	267	5.75%
n	C_1	135	5.85%	C_2	93	3.98%	sum	228	4.91%
p	C_1	116	5.03%	C_2	64	2.74%	sum	180	3.88%
ʁ	C_1	180	7.80%	C_2	300	12.84%	sum	480	10.34%
s	C_1	201	8.71%	C_2	108	4.62%	sum	309	6.65%
ʃ	C_1	216	9.36%	C_2	123	5.26%	sum	339	7.30%
t	C_1	150	6.50%	C_2	195	8.34%	sum	345	7.43%
ʦ	C_1	102	4.42%	C_2	78	3.34%	sum	180	3.88%
v	C_1	38	1.65%	C_2	114	4.88%	sum	152	3.27%
x	C_1	324	14.04%	C_2	282	12.07%	sum	606	13.05%
z	C_1	69	2.99%	C_2	78	3.34%	sum	147	3.17%

pi'el:

b	C_1	35	1.72%	C_2	180	8.85%	sum	215	5.29%
d	C_1	78	3.83%	C_2	117	5.75%	sum	195	4.79%
f	C_1	90	4.42%	C_2	0	0.00%	sum	90	2.21%
g	C_1	117	5.75%	C_2	102	5.01%	sum	219	5.38%
k	C_1	146	7.18%	C_2	240	11.80%	sum	386	9.49%
l	C_1	63	3.10%	C_2	177	8.70%	sum	240	5.90%
m	C_1	114	5.60%	C_2	180	8.85%	sum	294	7.23%
n	C_1	186	9.14%	C_2	108	5.31%	sum	294	7.23%
p	C_1	45	2.21%	C_2	141	6.93%	sum	186	4.57%
ʁ	C_1	165	8.11%	C_2	126	6.19%	sum	291	7.15%
s	C_1	162	7.96%	C_2	90	4.42%	sum	252	6.19%
ʃ	C_1	180	8.85%	C_2	105	5.16%	sum	285	7.01%
t	C_1	129	6.34%	C_2	213	10.47%	sum	342	8.41%
ʦ	C_1	81	3.98%	C_2	57	2.80%	sum	138	3.39%
v	C_1	70	3.44%	C_2	0	0.00%	sum	70	1.72%
x	C_1	298	14.65%	C_2	144	7.08%	sum	442	10.87%
z	C_1	75	3.69%	C_2	54	2.65%	sum	129	3.17%

kal and *pi'el*:

b	C_1	111	2.56%	C_2	237	5.42%	sum	348	3.99%
d	C_1	162	3.73%	C_2	219	5.01%	sum	381	4.37%
f	C_1	148	3.41%	C_2	128	2.93%	sum	276	3.17%
g	C_1	249	5.74%	C_2	198	4.53%	sum	447	5.13%
k	C_1	386	8.89%	C_2	402	9.20%	sum	788	9.04%
l	C_1	135	3.11%	C_2	381	8.72%	sum	516	5.92%
m	C_1	228	5.25%	C_2	333	7.62%	sum	561	6.44%
n	C_1	321	7.39%	C_2	201	4.60%	sum	522	5.99%
p	C_1	161	3.71%	C_2	205	4.69%	sum	366	4.20%
ʁ	C_1	345	7.95%	C_2	426	9.75%	sum	771	8.85%
s	C_1	363	8.36%	C_2	198	4.53%	sum	561	6.44%
ʃ	C_1	396	9.12%	C_2	228	5.22%	sum	624	7.16%
t	C_1	279	6.43%	C_2	408	9.33%	sum	687	7.89%
ʦ	C_1	183	4.22%	C_2	135	3.09%	sum	318	3.65%
v	C_1	108	2.49%	C_2	114	2.61%	sum	222	2.55%
x	C_1	622	14.33%	C_2	426	9.75%	sum	1048	12.03%
z	C_1	144	3.32%	C_2	132	3.02%	sum	276	3.17%

Appendix F: Lexical decision experiment: stimuli

Nonce-Verbs				
0 shared features	1 shared feature (place)	2 shared features (place and manner)	2 shared features (place and voice)	Identical C_1-C_2
gisem	disem	ʃizek	titsem	tsitseg
gitsem	liseg	sizek	tiʃem	ʃiʃeg
likem	litsem	ziseg	tsiʃem	tsitsem
miʃeg	tsizek	zisem	tisem	ʃiʃem
ʃimek	zitem	ziʃeg	sitseg	sisem
zikem	zitseg	ziʃem	ʃitseg	siseg

Real Verbs				
bikeʃ	kibel	kipets	pinek	ʃilem
biʃel	kibes	kiʃef	sibex	ʃilev
dileg	kidem	litef	sikem	ʃitef
gibeʃ	kilef	mizeg	silek	ʃitek
gilem	kimet	piked	sipek	tinef
kibed	kipel	pileg	ʃikem	tsilem

Trial Session	
Real Verbs	Nonce-Verbs
dibeʁ	bideg
limed	digev
sipeʁ	gidev
sixek	piget
xibek	tipeg

Appendix G: Word-likelihood judgment experiment: stimuli

All stimuli are in the form of: *proper name + nonce-verb + the animal*

Each participant got 49 random sentences, with a random verb from the bold triplet.

Yaron ***bafad\bafan\bafat*** *et ha-ʃablul* (snail)
Ron ***bakaz\bakat\bakas*** *et ha-dolfin* (dolphin)
Yonatan ***bagaz\bagat\bagal*** *et ha-ʃoʁ* (bull)
Dror ***baxat\baxats\baxas*** *et ha-dvoʁa* (bee)
Tomer ***baʃak\baʃag\baʃax*** *et ha-tolaat* (worm)
Itamar ***bamax\bamaʁ\bamag*** *et ha-tigʁis* (tiger)
Elad ***banak\banax\banaʁ*** *et ha-gdi* (young goat)
Uri ***bavak\bavax\bavag*** *et ha-dʒiʁafa* (giraffe)
Elad ***baʁal\baʁan\baʁas*** *et ha-kof* (monkey)
Yossi ***balaʁ\balag\balax*** *et ha-tinʃemet* (barn owl)
Idan ***basak\basax\basag*** *et ha-axbaʁ* (mouse)
Chen ***batsag\batsax\batsag*** *et ha-tanin* (crocodile)
Eyal ***bazag\bazax\bazax*** *et ha-paʁoʃ* (flea)
Nadav ***badag\badag\badag*** *et ha-oʁev* (crow)
Yoav ***batag\batag\batag*** *et ha-tsipoʁ* (bird)
Yuval ***dafax\dafaʁ\dafaʁ*** *et ha-jona* (pigeon)
Gal ***damaʁ\damak\damag*** *et ha-dov* (bear)
Doron ***dadav\dadaf\dadam*** *et ha-pingwin* (penguin)
Michael ***dazam\dazav\dazaf*** *et ha-tsav* (turtle)
Ran ***davag\davax\davag*** *et ha-tsav* (turtle)
Nadav ***daxam\daxav\daxav*** *et ha-paʁpaʁ* (butterfly)
Ram ***dalam\dalam\dalav*** *et ha-taʁnegol* (rooster)
Amit ***daʁaf\daʁaf\daʁav*** *et ha-zeev* (wolf)
Israel ***gafad\gafaz\gafal*** *et ha-livjatan* (whale)
Netanel ***gamas\gamat\gamats*** *et ha-meduza* (jellyfish)
Erez ***gagaf\gagam\gagav*** *et ha-dinozauʁ* (dinosaur)
Barak ***gatsav\gatsaf\gatsam*** *et ha-axbeʁoʃ* (rat)
Tal ***gavad\gavats\gavaz*** *et ha-nemala* (ant)
Alon ***gazav\gazav\gazaf*** *et ha-akaviʃ* (spider)
Gil ***gaʁan\gaʁats\gaʁan*** *et ha-dov* (bear)
Roee ***gaxaf\gaxaf\gaxav*** *et ha-xipuʃit* (ladybug)
Oren ***galak\galaʁ\galaʁ*** *et ha-janʃuf* (owl)
Amir ***gadav\gadav\gadav*** *et ha-akʁav* (scorpion)
Shachar ***kagam\kagaf\kagav*** *et ha-boeʃ* (polecat)
Ariel ***kakaf\kakav\kakam*** *et ha-kelev* (dog)
Uri ***paxan\paxal\paxaʃ*** *et ha-snai* (squirrel)
Dror ***pafak\pafag\pafaʁ*** *et ha-atalef* (bat)
Dan ***pamag\pamak\pamax*** *et ha-kaʁiʃ* (shark)
Chen ***pavag\pavaʁ\pavax*** *et ha-taiʃ* (male goat)
Ori ***pazak\pazax\pazag*** *et ha-janʃuf* (owl)
Oz ***panak\panaʁ\panax*** *et ha-zvuv* (fly)
Nitsan ***padak\padaʁ\padag*** *et ha-saknai* (pelican)
Dor ***palaʁ\palaʁ\palak*** *et ha-letaa* (lizard)
Omer ***pasag\pasag\pasaʁ*** *et ha-tsfaʁdea* (frog)
Doron ***paʃak\paʃag\paʃak*** *et ha-tuki* (parrot)
Netanel ***patsag\patsag\patsak*** *et ha-akʁav* (scorpion)
Michael ***paʁan\paʁal\paʁan*** *et ha-kof* (monkey)
Yair ***patag\patag\patag*** *et ha-tolaat* (worm)
Tomer ***ʁaval\ʁavan\ʁavat*** *et ha-aʁje* (lion)
Eran ***ʁafal\ʁafaz\ʁafan*** *et ha-baʁvaz* (duck)
Nir ***ʁalaf\ʁalav\ʁalam*** *et ha-ʃual* (fox)
Shay ***ʁanaf\ʁanam\ʁanav*** *et ha-egel* (calf)
Saar ***ʁaʁam\ʁaʁaf\ʁaʁav*** *et ha-xipuʃit* (ladybug)
Adam ***sagaf\sagats\sagat*** *et ha-dvoʁa* (bee)
Yotam ***saxaz\saxaf\saxats*** *et ha-kipod* (hedgehog)
Aviv ***sazam\sazav\sazaf*** *et ha-kivsa* (sheep)
Boaz ***sadam\sadaf\sadav*** *et ha-nameʁ* (leopard)
Guy ***sasam\sasav\sasaf*** *et ha-xatul* (cat)
Matan ***safaz\safats\safaz*** *et ha-tuki* (parrot)
Itamar ***savag\savak\savak*** *et ha-ʃoʁ* (bull)
Yoav ***sanav\sanam\sanav*** *et ha-nemala* (ant)
Ran ***sataf\satav\sataf*** *et ha-ʃafan* (rabbit)
Yuval ***samag\samag\samag*** *et ha-naxaʃ* (snake)
Or ***saʁam\saʁam\saʁam*** *et ha-avaz* (goose)
Oded ***salav\salav\salav*** *et ha-ʃimpanza* (chimp)
Oren ***ʃagaf\ʃagav\ʃagam*** *et ha-gdi* (young goat)
Omri ***ʃaʃaf\ʃaʃam\ʃaʃav*** *et ha-letaa* (lizard)
Danny ***ʃalaʁ\ʃalag\ʃalaʁ*** *et ha-ez* (nanny goat)
Amit ***ʃafag\ʃafak\ʃafak*** *et ha-hipopotam* (hippo)
Omri ***ʃaʁaf\ʃaʁaf\ʃaʁam*** *et ha-saknai* (pelican)

Ori **kamad\kamad\kamaz** *et ha-dag* (fish)

Gal **kadav\kadaf\kadaf** *et ha-pingwin* (penguin)

Matan **katag\katax\katax** *et ha-nameʁ* (leopard)

Saar **kalag\kalaʁ\kalag** *et ha-pil* (elephant)

Moti **kavaz\kavaz\kavat** *et ha-hipopotam* (hippo)

Nimrod **kafas\kafas\kafas** *et ha-paʁpaʁ* (butterfly)

Moti **kaʁaf\kaʁaf\kaʁaf** *et ha-xamoʁ* (donkey)

Yaniv **kasav\kasav\kasav** *et ha-zebʁa* (zebra)

Daniel **katsam\katsam\katsam** *et ha-boeʃ* (polecat)

Danny **kaʃam\kaʃam\kaʃam** *et ha-zvuv* (fly)

Assaf **lavax\lavak\lavag** *et ha-baʁvaz* (duck)

Idan **lafaʁ\lafag\lafak** *et ha-xamoʁ* (donkey)

Aviad **lalam\lalaf\lalav** *et ha-xatul* (cat)

Erez **mafax\mafag\mafaʁ** *et ha-tsipoʁ* (bird)

Ofer **mavaʁ\mavag\mavak** *et ha-axbeʁoʃ* (rat)

Or **maʁas\maʁats\maʁaz** *et ha-akaviʃ* (spider)

Alon **mamax\mamag\mamaʁ** *et ha-tavas* (peacock)

Amir **mazax\mazak\mazak** *et ha-snai* (squirrel)

Daniel **maʃaʁ\maʃaʁ\maʃag** *et ha-tsvi* (deer)

Tsachi **malaʁ\malaʁ\malaʁ** *et ha-paʁa* (cow)

Yakir **nadak\nadag\nadax** *et ha-oʁev* (crow)

Ben **nafak\nafaʁ\nafag** *et ha-ʃual* (fox)

Noam **nalav\nalam\nalaf** *et ha-jona* (pigeon)

Oded **nanam\nanaf\nanav** *et ha-baʁbuʁ* (swan)

Ohad **nakas\nakas\nakats** *et ha-zeev* (wolf)

Tom **navag\navag\navak** *et ha-paʁa* (cow)

Ariel **nazav\nazam\nazav** *et ha-tanin* (crocodile)

Ido **naʃag\naʃag\naʃag** *et ha-kengeʁu* (kangaroo)

Shay **nagam\nagam\nagam** *et ha-tsfaʁdea* (frog)

Guy **natam\natam\natam** *et ha-taiʃ* (male goat)

Yaron **namaʁ\namaʁ\namaʁ** *et ha-axbaʁ* (mouse)

Ofer **pakaz\pakat\pakaʃ** *et ha-ʃablul* (snail)

Yaniv **pagan\pagas\pagats** *et ha-gamal* (camel)

Evyatar **zafak\zafax\zafaʁ** *et ha-ʃafan* (rabbit)

Tsachi **zazav\zazaf\zazam** *et ha-xaziʁ* (pig)

Lior **zakav\zakam\zakam** *et ha-ogeʁ* (hamster)

Yoni **zavaʁ\zavaʁ\zavak** *et ha-aʁje* (lion)

Yonatan **zalav\zalav\zalam** *et ha-atalef* (bat)

Aviad **zaʁaf\zaʁaf\zaʁaf** *et ha-tigʁis* (tiger)

Nir **ʃadam\ʃadav\ʃadam** *et ha-sus* (horse)

Eyal **ʃasam\ʃasav\ʃasav** *et ha-aʁnav* (hare)

Yakir **ʃatsam\ʃatsav\ʃatsam** *et ha-jatuʃ* (mosquito)

Itay **ʃaxam\ʃaxam\ʃaxam** *et ha-kaʁnaf* (rhino)

Dan **ʃavag\ʃavag\ʃavag** *et ha-kaʁnaf* (rhino)

Noam **ʃatav\ʃatav\ʃatav** *et ha-gamal* (camel)

Dor **ʃazav\ʃazav\ʃazav** *et ha-xaziʁ* (pig)

Tal **tagaf\tagam\tagav** *et ha-ez* (nanny goat)

Ido **tavak\tavaʁ\tavag** *et ha-naxaʃ* (snake)

Israel **tatav\tatam\tataf** *et ha-zebʁa* (zebra)

Ofir **tatsav\tatsam\tatsaf** *et ha-ʤiʁafa* (giraffe)

Yair **tafak\tafak\tafag** *et ha-dolfin* (dolphin)

Ohad **tamak\tamag\tamag** *et ha-kelev* (dog)

Tom **tadav\tadav\tadaf** *et ha-tsaʁtsaʁ* (cricket)

Yoni **tazav\tazaf\tazaf** *et ha-kivsa* (sheep)

Itay **taʁav\taʁav\taʁav** *et ha-aʁnav* (hare)

Boaz **talav\talav\talav** *et ha-tavas* (peacock)

Ron **tsaʁaf\tsaʁas\tsaʁaz** *et ha-paʁoʃ* (flea)

Eran **tsafax\tsafak\tsafag** *et ha-livjatan* (whale)

Barak **tsaxam\tsaxaf\tsaxav** *et ha-kipod* (hedgehog)

Assaf **tsasaf\tsasam\tsasav** *et ha-avaz* (goose)

Adam **tsatsam\tsatsaf\tsatsav** *et ha-sus* (horse)

Yossi **tsazav\tsazam\tsazaf** *et ha-baʁbuʁ* (swan)

Nitsan **tsalag\tsalag\tsalaʁ** *et ha-kaʁiʃ* (shark)

Yotam **tsavak\tsavak\tsavag** *et ha-tsaʁtsaʁ* (cricket)

Evyatar **tsadav\tsadam\tsadam** *et ha-ogeʁ* (hamster)

Nimrod **tsamag\tsamag\tsamag** *et ha-ʃimpanza* (chimp)

Oz **tsanav\tsanav\tsanav** *et ha-dinozauʁ* (dinosaur)

Ram **xalag\xalaʁ\xalaʁ** *et ha-jatuʃ* (mosquito)

Ofir **xatak\xatak\xatag** *et ha-xasida* (stork)

Aviv **xazav\xazaf\xazaf** *et ha-pil* (elephant)

Roee **xaxav\xaxaf\xaxav** *et ha-kengeʁu* (kangaroo)

Gil **xaʁan\xaʁal\xaʁan** *et ha-egel* (calf)

Avi **xafal\xafal\xafad** *et ha-dag* (fish)

Omer **xakav\xakav\xakaf** *et ha-xasida* (stork)

Shachar **xavan\xavaz\xavan** *et ha-meduza* (jellyfish)

Lior **xanaʁ\xanaʁ\xanaʁ** *et ha-tsvi* (deer)

Ben **xamaz\xamaz\xamaz** *et ha-taʁnegol* (rooster)

Appendix H: Word-likelihood judgment experiment: The triplets with significant difference in ratings

		Average	SD	P-value
	baʁal	3.43	1.79	
bʁ	baʁan	4.25	2.02	0.015
	baʁas	5.39	1.75	
	batsag	3.45	1.69	
bts	batsag	2.36	1.22	0.032
	batsax	4.07	1.73	
	dadaf	3.08	2.22	
dd	dadam	3.41	2.00	0.037
	dadav	1.82	1.29	
	dafaʁ	4.69	1.35	
df	dafaʁ	4.83	1.79	0.015
	dafax	3.43	1.79	
	daʁaf	5.18	1.88	
dʁ	daʁaf	5.00	1.53	0.024
	daʁav	3.86	1.75	
	daxam	3.14	1.79	
dx	daxav	5.06	1.85	0.002
	daxav	4.94	1.30	
	padag	3.12	1.36	
pd	padak	5.12	1.80	0.004
	padaʁ	3.93	1.58	
	pakaʃ	4.50	1.70	
pk	pakat	4.67	1.67	0.022
	pakaz	3.00	1.81	
	palak	5.23	1.36	
pl	palaʁ	4.15	1.68	0.038
	palaʁ	4.07	1.58	
	pamag	1.92	0.64	
pm	pamak	3.50	1.15	0.000
	pamax	4.24	1.92	
	paʃag	3.27	1.94	
pʃ	paʃak	4.43	0.94	0.011
	paʃak	4.83	1.72	

		Average	SD	P-value
	patsag	3.80	1.32	
pts	patsag	3.07	1.59	0.012
	patsak	4.56	1.03	
	ʁaʁaf	3.29	1.82	
ʁʁ	ʁaʁam	2.07	1.44	0.001
	ʁaʁav	1.31	0.70	
	ʃafag	3.43	1.83	
ʃf	ʃafak	4.41	1.46	0.029
	ʃafak	4.94	1.92	
	sagaʃ	3.25	2.14	
sg	sagat	3.71	1.69	0.047
	sagats	2.25	1.57	
	ʃalag	5.06	1.61	
ʃl	ʃalaʁ	3.78	1.59	0.043
	ʃalaʁ	4.06	2.10	
	savag	5.14	1.41	
sv	savak	3.58	1.73	0.005
	savak	3.67	1.50	
	tataf	3.18	1.78	
tt	tatam	3.73	1.75	0.017
	tatav	2.25	1.81	
	tsatsaf	2.89	1.75	
tsts	tsatsam	2.21	1.63	0.020
	tsatsav	1.70	1.49	

Appendix I: Word-likelihood judgment experiment: Results

	Average Rating		Average Rating		Average Rating		Average Rating
ss	2.05	gf	3.49	gd	4.05	nt	4.63
ʦs	2.21	sx	3.50	bn	4.06	ʃv	4.64
kg	2.27	pg	3.52	dl	4.08	gʁ	4.64
ʃʃ	2.29	sm	3.55	bd	4.09	ng	4.64
zz	2.51	mv	3.55	ʦx	4.09	ʦf	4.64
td	2.53	bs	3.56	ps	4.10	xv	4.64
tʦ	2.60	zk	3.60	gv	4.13	ʃʁ	4.65
bv	2.65	xt	3.68	xk	4.16	mʃ	4.73
ʦʁ	2.65	ʁl	3.70	ʦn	4.20	kd	4.74
kk	2.66	nʃ	3.70	st	4.20	xm	4.74
sz	2.66	xl	3.74	ʦv	4.21	gx	4.79
gg	2.67	pn	3.76	gz	4.22	zʁ	4.79
ll	2.68	gl	3.77	tg	4.23	nz	4.83
mm	2.69	bʃ	3.79	nv	4.24	bx	4.84
bt	2.69	zf	3.79	kt	4.25	bl	4.87
nn	2.71	dm	3.80	xz	4.27	kf	5.00
ʦz	2.73	ks	3.80	kʦ	4.28	kʁ	5.10
tz	2.73	zv	3.83	ʃg	4.28	xf	5.22
sf	2.80	ʁf	3.84	ʃd	4.29	px	5.39
pf	2.81	lf	3.85	zl	4.30	ʃx	5.41
tm	2.91	sd	3.85	mʁ	4.30		
bf	2.93	ml	3.86	kʃ	4.31		
ʃʦ	2.96	kl	3.87	tf	4.31		
xx	2.96	nd	3.88	sʁ	4.33		
dz	3.00	ʃt	3.90	bz	4.33		
bm	3.02	pt	3.91	gm	4.34		
ʃs	3.15	ʁv	3.93	nm	4.38		
pz	3.16	kv	3.93	nk	4.39		
pv	3.21	km	3.94	nf	4.39		
gʦ	3.27	dv	3.96	mz	4.42		
ʁn	3.31	tl	3.96	ʃz	4.46		
bk	3.42	xn	3.98	tv	4.48		
nl	3.42	lv	3.98	xʁ	4.48		
mf	3.43	sn	4.02	sl	4.49		
ʦd	3.44	ʦl	4.02	pʁ	4.52		
ʦm	3.45	bg	4.04	tʁ	4.60		

REFERENCES

Aronoff, Mark. 1976. *Word formation in Generative Grammar*. Cambridge: MIT Press.

Barkali, S. 1964. *Luach hape'alim hashalem* (18th ed.). Jerusalem: Rubin Mass. [in Hebrew]

Bat-El, Outi. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory 12*: 572-596.

Bat-El, Outi. 2003a. Semitic verb structure within a universal perspective. In Joseph Shimron (ed.) *Language Processing and Acquisition in Languages of Semitic, Root-based, Morphology*. pp. 29-59. Amsterdam: John Benjamins.

Berent, Iris, Everett, Daniel. L., and Joseph Shimron. 2001. Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology*, *42*(1), 1-60.

Berent, Iris, Lennertz, Tracy, Jun, Jongho, Moreno, Miguel, A., and Paul Smolensky. 2008. Language universals in human brains. Proceedings of the National Academy of Sciences, 105, 5321-5325.

Berent, Iris, and Joseph Shimron. 1997. The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition*, *64*(1), 39-72.

Blau, Joshua. 2010. *The Phonology and Morphology of Biblical Hebrew.* The Academy of Hebrew Language. [In Hebrew]

Bolozky, Shmuel and Rina Kreitman. 2007. Uvulars in Israeli Hebrew – their phonetic and phonological status. *The National Association of Professors of Hebrew's International Conference on Hebrew Language, Literature and Culture*. Sydney, Australia July 2-4.

Clements, George N. 1985. The Geometry of Phonological Features. *Phonology Yearbook* 2:225-252.

Clements, George N. and Elizabeth Hume. 1995. The Internal Organization of Speech Sounds. In John Goldsmith, ed., *Handbook of Phonological Theory*. Oxford: Basil Blackwell, Oxford, pp. 245-306

Clements, George N. and Samuel J. Keyser. 1983. *CV Phonology*. MIT Press, Cambridge, MA.

Davis. Stuart M. 1984. Some Implications of Onset-Coda Constraints for Syllable Phonology. *Chicago Linguistic Society 20*: 46-51

Even-Shoshan, Avraham. 1970. The New Hebrew Dictionary. *Jerusalem: Kiryat Sefer*.

Even-Shoshan, Avraham. 1983. The Condensed Hebrew Dictionary. *Jerusalem: Kiryat Sefer*

Frisch, Stefan. A., Pierrehumbert, Janet. B., and Michael B Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, *22*(1), 179-228.

Frisch, Stefan A. and Bushra Zawaydeh. 2001. The Psychological Reality of OCP-Place in Arabic, *Language* 77, 91–106.

Fudge, Erik C. 1969. Syllables. *Journal of Linguistics 5*: 253-287.

Goldsmith, John. A. 1976. *Autosegmental phonology* (Vol. 159). Bloomington: Indiana University Linguistics Club.Yip 1998

Greenberg, Joseph. 1950. The patterning of morphemes in Semitic. *Word*, 4, 196–208.

Johnson, Keith. 2012. *Acoustic and Auditory Phonetics 3rd Edition*. Wiley-Blackwell, West Sussex. Pages 115-123

Kaisse, Ellen. 1988. Modern Greek continuant dissimilation and the OCP. *Ms, University of Washington*

Kawahara, Shigeto. 2007. Half rhymes in Japanese rap lyrics and knowledge of similarity. Journal of East Asian Linguistics 16:113-144.

Kawahara, Shigeto. 2011. Modes of phonological judgment. *Ms, Rutgers University. Available as ROA-1140 from the Rutgers Optimality Archive.*

Kawahara, Shigeto, Hajime Ono, and Kiyoshi Sudo. 2006. Consonant co-occurrence restrictions in Yamato Japanese. *Japanese/Korean Linguistics* 14:27-38.

Laks, Lior. 2011. *Morpho-phonological and Morpho-thematic Relations in Hebrew and Arabic Verb Formation*. PhD Dissertation, Tel-Aviv University.

Leben, William Ronald. 1973. *Suprasegmental phonology* (Doctoral dissertation, Massachusetts Institute of Technology).

McCarthy, John. 1979. Formal problems in Semitic morphology and phonology. *Dissertation MIT, Cambridge, MA.*

McCarthy, John. 1981. A prosodic theory of nonconcatenative morphology.*Linguistic inquiry*, *12*, 373-418.

McCarthy, John. 1986. OCP effects: Gemination and antigemination. *Linguistic inquiry*, *17*(2), 207-263.

McCarthy, John. 1988. Feature Geometry and Dependency: A Review, *Phonetica* 43, 84–108.

Mielke, Jeff. 2012. A phonetically-based metric of sound similarity. *Lingua* 122:145-163.

Odden, David. 1988. Anti antigemination and the OCP. *Linguistic inquiry*, *19*(3), 451-476.

Padgett, Jaye. 1995. *Stricture in feature geometry.* SCLI Publications.

Pierrehumbert, Janet B. 1993. Dissimilarity in the Arabic Verbal Roots, in *Proceedings of the North East Linguistics Society* 23, pp. 367–381.

Prince, Alan, and Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Technical report RuCCSTR-2. Rutgers Center for Cognitive Science.

Rice, Keren. D. 1993. A reexamination of the feature [sonorant]: the status of sonorant obstruents. *Language*

Rose, Sharon. 2000. Rethinking geminates, long-distance geminates, and the OCP. *Linguistic Inquiry*, *31*(1), 85-122.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language*, *80*(3), 475-531

Sagey, Elizabeth C. 1986. *The Representation of Features and Relations in Non-linear Phonology*. Ph.D. dissertation, MIT.

Schwarzwald, Ora. 2002. *Studies in Hebrew Morphology*. The Open University of Israel [in Hebrew]

Yeverechyahu, Hadas. 2012. *OCP in Hebrew Verbs - an Empirical View*. A poster presented at *The Second* International Graduate Student Conference on Diverse Approaches to Linguistics (IGDAL 2), Tel-Aviv University.

Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic inquiry*, *19*(1), 65-100.

Zadok, Gila. 2012. *On the Instability of Hebrew Weak Verbs: Similarity, Variation, and Change*. Ph.D. dissertation, Tel-Aviv University

תקציר

בשפות שמיות, עיצורים הומורגניים (עיצורים החולקים מקום חיתוך) נוטים שלא להופיע באותו גזע (Greenberg 1950). מחקרים קודמים (כדוגמת McCarthy 1981, 1986; Frisch et al. 2004) מציעים כי הגבלות אלו נובעות מדמיון בין עיצורים, כך שככל ששני עיצורים (הומורגניים) דומים יותר זה לזה, כך הסבירות שיופיעו באותו גזע נמוכה יותר. המחקר הנוכחי בוחן את ההגבלות על מופעי עיצורים במערכת הפועל בעברית. אני שואלת כיצד דמיון בין עיצורים תורם להגבלות, והאם ההגבלות נובעות מאילוצים אוניברסליים או שהן מושפעות מגורמים תלויי שפה.

למחקר שלושה חלקים עיקריים. ראשית, התאמתי את מודל הדמיון של Frisch et al. (2004) למצאי העיצורים בעברית. שנית, ניתחתי את לקסיקון מערכת הפועל בעברית, תוך התמקדות בעיצורים הראשון והשני של הגזע (C_1-C_2) בבניינים קל ופיעל. הניתוח הראה מתאם ברמת מובהקות גבוהה בין מודל הדמיון והלקסיקון, וכן מציע כי למקום החיתוך של העיצורים יש השפעה רבה, בהשוואה לתכוניות אחרות, בקביעת ההגבלות. על מנת לחזק ולהרחיב את הניתוח הלקסיקלי, ערכתי שני ניסויים פסיכובלשניים: מטלת זיהוי לקסיקלי ומטלת שיפוטי סבירות עבור לא-מילים, שניהם בוחנים את ההגבלות במערכת הפונולוגית של הדוברים. ניתוח שיפוטי הסבירות הראה מתאם ברמת מובהקות גבוהה בין שיפוטי הדוברים ומודל הדמיון מחד, ובין שיפוטי הדוברים ומחקר הלקסיקון מאידך. כמו כן, הניסויים גם כן מדגישים את תפקיד מקום החיתוך של העיצורים בהגבלות על מופעים.

ממצאים אלו מציעים כי ישנן הגבלות מבוססות דמיון על שני העיצורים הראשונים בגזע הפעלים בעברית, גם בלקסיקון וגם במערכת הפונולוגית של הדוברים. כמו כן, הם מציעים כי למקום החיתוך של העיצורים השפעה רבה בקביעת ההגבלות, כך ששני עיצורים החולקים מקום חיתוך ראשי נוטים שלא להופיע באותו גזע. אף על פי כן, הניסויים אינם יכולים להציע האם השפעת הדמיון על המערכת הפונולוגית היא ישירה, או עקיפה דרך השפעות לקסיקליות.

TEL AVIV UNIVERSITY אוניברסיטת תל-אביב

הפקולטה למדעי הרוח ע״ש לסטר וסאלי אנטין

החוג לבלשנות

תפקיד הדמיון בהגבלות על רצפי עיצורים:

עדות ממערכת הפועל בעברית

חיבור זה הוגש כעבודת גמר לקראת התואר

״מוסמך אוניברסיטה״ באוניברסיטת ת״א

על ידי

הדס יברכיהו

העבודה הוכנה בהדרכת:

פרופ׳ אותי בת-אל

ד״ר אוון-גרי כהן

דצמבר 2014