Tel Aviv University
The Lester & Sally Entin Faculty of Humanities
The Shirley & Leslie Porter School of Cultural Studies

SIMILARITY, VARIATION, AND CHANGE: INSTABILITY IN HEBREW WEAK VERBS

THESIS SUBMITTED FOR THE DEGREE
"DOCTOR OF PHILOSOPHY"

by

Gila Zadok

Submitted to the Senate of Tel Aviv University

December 2012

This work was carried out under the supervision of

Prof. Outi Bat-El

TABLE OF CONTENTS

ABSTRACT

The Hebrew verb system is changing. This dissertation sets out to investigate the nature of the change, what triggered it, what its driving forces are, and where it is going.

We know that the verb system is changing because of the large amount of surface variation attested particularly in the group of verbs known as the 'weak verbs'. This group comprises many distinct paradigm types (sub-classes) that group verbs together according to their weak segment and its position in the stem. The Hebrew verb system comprises five classes of verbs (named *binyanim*) that determine the overall shape of the verb (including prefixes, if any, and vocalic pattern). With each class (*binyan*) comprising several distinct sub-classes, the overall number of distinct paradigm types is extremely large, but systematic. So why change?

The prime suspects in this investigation are the pharyngeals and the glottals, collectively known as the 'gutturals', which are on the verge of extinction. Their loss leaves no surface cues for sub-class classification, and so speakers are forced to memorise or regularise. Thus, the loss of the gutturals wreaked havoc in the verb system, causing increased similarity between once distinct paradigms. The increase in similarity triggered the migration among the sub-classes within the binyan.

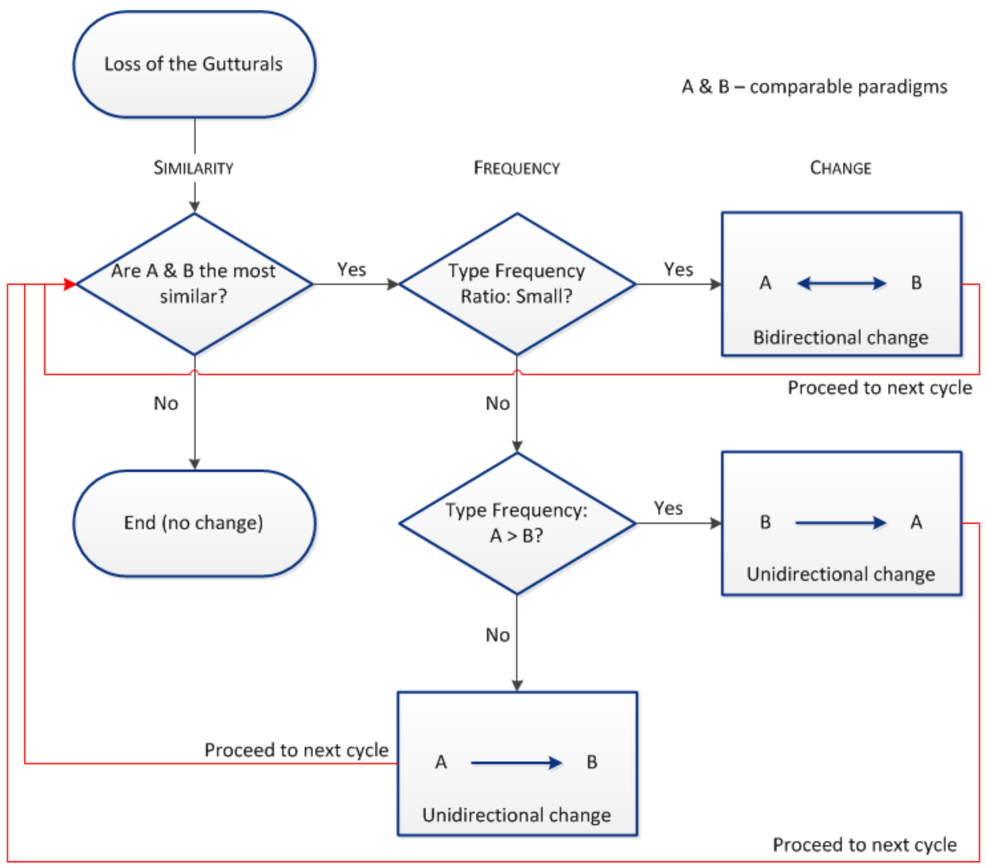
The migration among the sub-classes, I claim, is systematic. I propose a model for quantifying similarity between full-fledged paradigms that shows that the migration from any specific sub-class is ***only*** to the most similar sub-class within the confines of the binyan.

Once similarity identifies the merging sub-classes, the question of directionality is raised. I show two types of migration paths: unidirectional and bidirectional. In the unidirectional path, the members of one paradigm type, A, migrate to the most similar paradigm type, B, until paradigm A becomes extinct. In the bidirectional path, the members of paradigm A migrate to the most similar paradigm B as members of paradigm B migrate to paradigm A. Thus, the bidirectional path is seemingly

superfluous, the ultimate goal of the change is to reduce the number of paradigm types, as both paradigm types survive.

What determines the direction of the change is the frequency of the two comparable paradigm types. Verbs from the smaller group will typically migrate to the larger group, the paradigm type with more verb types. Thus, the goal of reducing the number of paradigm types is achieved while minimally disturbing the stability of the system. While bidirectionality is not ideal for reducing the number of paradigm types, it occurs when speakers cannot tell which is the larger group. Overall, the migration among the sub-classes manifested in the observed surface variation in B1 only, will reduce the number of paradigm types by 30%.

The flow of analogical change in the Hebrew verb system is summarised in the following chart.



Keywords: morphology, similarity, change, variation, Hebrew, verbs, analogy, frequency, Stochastic OT, Gradual Learning Algorithm, evolOT

Acknowledgements

Deciding on the verb system as the topic of my dissertation was possibly the easiest decision I have ever made. Coming up with a thesis, was a different story altogether. Only two things were clear from the very beginning: that the surface variation indicated a system in transition and that the change had to be systematic because… I wouldn't have it any other way.

It is impossible to describe the amount of work that goes into writing a dissertation to anyone who hasn't done it. If it were, far less dissertations would be written. I would not have managed to see this endeavour through if it weren't for the amazing guidance and support of my advisor, Outi Bat-El. Outi is the perfect advisor: on numerous occasions she adapted herself to my crazy schedule, spent hours upon hours discussing my ideas, my progress, and sometimes lack thereof, reading countless drafts, critiquing my work, and offering invaluable advice.

One of the most memorable critiques that I received from Outi, was that when I got it right, a pattern would emerge. For the longest time, there was no pattern. I shuffled verbs from one pile to another attempting to organise and reorganise them in a way that made sense, until finally that longed for pattern emerged. I can only hope this means that I got it right.

This journey was anything but a smooth ride. I dealt with several challenges that sometimes felt too much. Outi managed to pull me through some of these moments and convince me to persevere. When she was unsuccessful, she had the wisdom to ask Galit Adam to do it. Luckily, as it turns out, I cannot resist them both.

Galit was incredibly supportive in reading several drafts since the early days of the proposal, and offered her unique perspective. She also pre-reviewed my final draft before submission, to which I am immensely grateful. I know it was not easy.

I also owe a debt of gratitude to Evan Cohen, who has also kindly pre-reviewed my final draft. Evan has always shown interest in my research and I often managed to get some insight from him in every one of our mini chats on my way from one place to another.

I would also like to thank my readers, Prof. Stephan Anderson and Prof. Shmuel Bolozky, for their comments and insights both on the proposal and on the final draft. This dissertation has undoubtedly benefited from their comments. I gave great attention to everyone’s comments and tried to implement them to the best of my abilities. Needless to say, I am solely responsible for any errors that still remain.

Special thanks to my collegues at Alvarion who allowed me to record them. As promised, they shall remain anonymous. I would also like to thank my managers at Intucell for allowing me time off to write the final draft and for showing such incredible and, admittedly, unexpected support.

I also owe special thanks to my brother, Uri, and his wife, Tal, for helping me with some of the calculations. And to all my dear friends whom I bored with my endless rambling about verb migration; I know that they probably could not care less, but I figured out many things when trying to explain it to them.

Finally, I would like to thank my mother, Jasna, for holding back on her requests just so I could finish already.

CHAPTER 1. INTRODUCTION

The Modern Hebrew (MH) verb system is restrictive in terms of the number of patterns it exhibits and it is also quite systematic relative to the nominal system. Within the verb system, the irregular verbs, known as the weak verbs, allow a much greater number of patterns than the regular verbs, and they also exhibit a great degree of phonological variation. The seemingly chaotic nature of the weak verbs, manifested by the variation, stems primarily from the historical change in the language's inventory, namely the loss of the segments *ʕ*, *ʔ*, *ħ*, and *h*, collectively known as the 'gutturals'. The effect of this change in the segmental inventory is twofold:

a. The motivation for a distinction between paradigms has been lost: the gutturals are notorious for influencing their surrounding vowels and their loss thus caused a great deal of opacity with respect to the surface vowels. Compare for example, the near-minimal pair *saxár* 'he rented' and *saxár* 'he traded'. A near identity of two verb stems would normally imply identical inflectional paradigms. This is, however, not the case with these two verbs, whose plural forms differ in the number of surface syllables: *saxrú* 'they rented' vs. *saxarú* 'they traded'. The reason for this discrepancy lies in the origin of the consonant *x*. The *x* in *saxár* 'he traded' historically originates from the pharyngeal fricative *ħ*, which has merged with the dorsal fricative *x*, whereas the *x* in *saxár* 'he rented' originates from the dorsal fricative *x*. That is, the distinction between the paradigms was once phonologically motivated (*saxár–saxrú* 'rented' vs. *saħár–saħarú* 'traded'). The *ħ*, a member of the gutturals, required a following low vowel, while the non guttural *x* did not. Following the merger of *ħ* and *x*, the appearance of the low vowel after *x* in *saxarú* 'they traded' has become opaque.

b. The surface cues for a distinction between paradigms has been lost. The loss of the gutturals makes the distinction among some paradigms very difficult because the most significant cues that made the distinction possible have disappeared. Consider the three normative verb forms, *milʔú* 'they filled', *nisú* 'they attempted’, and *bit͡sʕú* ‘they executed’. These forms differ in the final stem consonant: *milʔú* ‘they filled’ has a glottal stop in this position, *nisú* ‘they attempted’ has no consonant in this position (i.e. it is vowel final), and *bit͡sʕú* ‘they executed’ has a pharyngeal consonant in final position. The (normative) first person singular forms in the past tense indicate that the three forms pertain to distinct paradigms, manifested by the stressed vowel: *mil**é**ti* ‘I filled’, *nis**í**ti* ‘I attempted’, and *bit͡s**á**ti* ‘I executed’. The selection of the vowel depends solely on the final segment of the stem: final-*ʔ* entails an *e*, a vowel-final stem entails an *i* in that position, and a final pharyngeal entails an *a* in the same position. The subsequent loss of the gutturals caused these once distinct forms to become indistinct in some forms of the paradigm: *milʔú* → *milú* ‘they filled’, *nisú* ‘they attempted’ (no change) and *bit͡sʕú* → *bit͡sú* ‘they executed’. The cues necessary for medial vowel selection in the other forms of these paradigms, as described above, have been lost.[1]

The change in the segmental inventory, which is external to the verb system, has greatly affected the verb system, with ***the resulting opacity leading to variation, and variation in turn, leading to change***. The degree of variation in the verb system makes the system appear chaotic. The main goal of this study is to show that ***while variation is widespread in the MH verb system, it is, actually, quite systematic***. I argue that the variations exhibited can be grouped into two types of change: substitution and merger.

a. **Substitution** is typically unidirectional - a form is replaced by another form. For example, normative *yeesóf* ‘he will collect’ is being replaced by *yaasóf*, which assumes the structure of *yaavód* ‘he will work’. This substitution results in the loss of the *CeeCóC* pattern.

[1] The weakening of the glottals and pharyngeals were already observed in Biblical Hebrew (Bolozky 2003b).

b. **Merger** is typically bidirectional - verbs belonging to comparable groups migrate in both directions. For example, normative *miléti* ‘I filled’ has a surface variant *milíti*, which assumes the structure of *nisíti* ‘I attempted’. However, normative *nisíti* ‘I attempted’ also has a variant *niséti*, which assumes the structure of *miléti* ‘I filled’. In this case, *e*-verbs are migrating to the *i*-verbs group while *i*-verbs are migrating to the *e*-verbs group. Thus, migration occurs in both directions and neither pattern is lost.

The distinction between merger and substitution is thus very fine and lies solely on directionality. But although the migration in merger is bidirectional, the migration for a specific verb is, nevertheless, unidirectional. A specific verb migrates from one group to the other. So in essence, the process of change manifested in merger and substitution is one and the directionality is regulated by other factors.

I claim that ***the process of change in the verb system is triggered by similarity and its direction is regulated by the type frequency of the similar paradigms, such that typically the less frequent form or pattern is replaced by a more frequent form or pattern***. The ratio between the type frequencies of two comparable paradigms affects the direction of the change.

The similarity among verbs triggers both types of change in the verb system (substitution *and* merger). I claim that it is not merely similar paradigms that interact in the change process, but it is in fact only the ***most*** similar paradigms that do so. To demonstrate this, I propose a model for quantifying similarity among full-fledged paradigms (Chapter 4) and show that the model is able to predict which two comparable paradigms will interact in the change process (either merge or be substituted).

The type frequencies of the comparable paradigm pair determines the direction of the change. If the motivation for the change is to reduce the number of class paradigm allowed in the language (as claimed in §3.3), then the actual direction is not important, as long as it is unidirectional, because bidirectionality does not result in fewer patterns. Nevertheless, in §3.3.1, I show that the paradigm with the higher type frequency (i.e. the paradigm type hosting more verb members) is the one more likely to survive. Verbs from the smaller group will typically migrate to the larger group. This achieves the goal for less patterns while minimally disturbing the stability of the system.

Group size regulates the direction of change, but it is not the absolute number of group members that determines which of the paradigm types is larger. In §4.3.3, I suggest that the ratio between the two groups is the determining factor of group size. Thus, if the ratio between the two comparable paradigm types is large, this is taken to mean that the groups are sufficiently different in size to enable speakers to select the larger group. If, however, the ratio is small, this is taken to mean that the difference in size between the two groups is insufficient for speakers to determine which is the larger group, even though one group may indeed be larger as it hosts more verb members than the comparable group. Therefore, when the ratio is small, the migration is bidirectional.

1.1. Overview

The dissertation is organised as follows:

The relevant ***background on Modern Hebrew*** (§1.2), begins with a working definition for the terms ‘normative’ and ‘colloquial’ that will serve the purposes of this study (§1.2.1). The segmental inventory of the language is then presented in (§1.2.2), followed by a discussion on the Gutturals (§1.2.3), which receive a great deal of attention throughout the dissertation.

The ***theoretical background*** (§1.3) provides a high-level description of the theoretical frameworks used for the analysis. This includes a short introduction to traditional Optimality Theory (OT; Prince and Smolensky, 1993/2004) and its stochastic variant (Boersma, 1998, Boersma and Hayes, 2001). Two applications based on Stochastic OT (which is further discussed in §3.4) are presented: the Gradual

Learning Algorithm (Boersma and Hayes, 2001), which can test whether or not a proposed grammar is learnable, and evolOT (Jäger, 2002b), which simulates language evolution based on the Gradual Learning Algorithm. I also present in this section additional topics that are used for the analysis of the data. These include ‘frequency’ (token frequency and type frequency), analogy and paradigm levelling, and similarity (analogy, paradigm levelling and similarity are further discussed in Chapter 4).

In §1.4, I describe the ***sources of the data*** I used.

In Chapter 2, I present the ***Hebrew verb system*** and its five morphological classes (*binyanim*), with emphasis on the relevant distinction between the regular and weak (irregular) verbs. The paradigms in this chapter and the morpho-phonological alternations are drawn from the normative register.

In Chapter 3, I discuss ***variation and change*** in the Hebrew verb system. I claim that the variation exhibited in the verb system is, by and large, an indication of two kinds of change: merger and substitution. These types of change differ in their directionality. While substitution (§3.3.1) is typically unidirectional, merger (§3.3.2) is typically bidirectional. Two rare types of change are discussed as well: the multipath (§4.3.7) which is a special case of unidirectional change and is claimed to be a result of diphthong simplification in the guttural-final class, and also split paradigms (§3.3.3). I then present the grammar of change (§) where I explain how the probability of occurrence of each variant is calculated, based on the Gradual Learning Algorithm. I apply the principles of GLA on the *i* ~ *e* alternation exhibited in one of the classes (*binyanim*).

Chapter 4 is devoted to the ***similarity*** in the Hebrew verb system and its result in the levelling of (some of) the paradigms. I draw a distinction between intra-paradigm levelling, which is the traditional focus of paradigm levelling, and inter-paradigm levelling (analogical levelling). For this purpose, I provide examples of attempts to formalise analogy in the early days of proportional analogy (§4.1.1) and more recently in OT (§4.1.2).

I argue that similarity is the basis of the analogy exhibited in the verb system (§4.2). I present the hierarchical organisation of the sub-paradigms of the Hebrew verb system and claim that similarity diminishes higher up in this hierarchy. Within this hierarchical organisation, the majority of the observed variation occurs at the level of the sub-class because at this level there is no relation among the various sub-classes, neither derivational nor inflectional. Therefore, changes occurring at this level will not affect either derivation or inflection.

Similarity, though, only has meaning if it has a function. Therefore, I limit the discussion of similarity to the properties of the binyan and propose a model of similarity that is gradient based on the relative distance between comparable paradigms (§4.3). Within this model, paradigms are compared such that each member of the paradigm is compared to its parallel member in the comparable paradigm. Each dissimilarity between the comparable members is counted and the sum of all the dissimilarity units from all the comparable forms result in the final score that reflects the degree of similarity of the comparable paradigm pair. By quantifying similarity, the model is able to predict which sub-class interactions are possible and which are not. A paradigm type will merge only with the paradigm type most similar to it.

Once the most similar paradigms have been identified, the question of directionality arises. The directionality is shown to be greatly affected by type frequency, such that members of the paradigm type with the lower type frequency typically migrate to the paradigm type with the higher type frequency. I discuss three types of migration paths: unidirectional, bidirectional, and a special case of multipath migration, where the migrating forms can choose between two paradigm types. In §4.3.7, the multipath type of migration is shown to be just another type of unidirectional migration path, where the second path is a special case of diphthong simplification, rather than merger with another paradigm type that is not the most similar.

Type frequency can only predict unidirectional change, from the less frequent to the more frequent paradigm type. To explain bidirectionality, I propose that the ratio between the type frequencies plays a role, such that a small ratio means that there is a greater chance for bidirectionality. In §4.3.9, I suggest how the model can be refined to accommodate additional distinctions that may be required in other languages.

Chapter 5 attempts to ***predict how the variation will resolve itself***. The prediction is that in both the unidirectional and the bidirectional variation, the newer variants will survive and the older variants will gradually become extinct. For this purpose, I apply evolOT (Jäger, 2002b), a software implementation of the Gradual Learning Algorithm (Boersma and Hayes 2001) to simulate language development in the course of time (§5.2). The simulation is consistent with the findings of this study regarding the surviving forms and the role of frequency in the progression of change.

Chapter 6 summarises the findings of this research.

1.2. Language Background

Modern Hebrew, also known as 'Israeli', or 'Contemporary' Hebrew (Rosén, 1973, Schwartzwald, 1985), is the primary language spoken in Israel. It is a member of the Canaanite languages of the Northwestern Semitic family. The history of the language dates back to the Hebrew Bible, reflecting the Hebrew language (referred to as Biblical Hebrew) of circa 1000-500 BC. By 200 AD, Aramaic had replaced it as the spoken language and Hebrew was used primarily for liturgical purposes. Hebrew was not spoken as a native language until its revival at the turn of the 20th century (Rabin, 1972, Schwarzwald, 2001). Nevertheless, Biblical Hebrew is the major source of the Modern Hebrew vocabulary (Ravid, 1995). Throughout its history, the language had many influences on all linguistic aspects, primarily from neighbouring languages, and most recently from English, Arabic, Yiddish, and Slavic languages (Wexler, 1990, Zuckermann, 2008).

1.2.1. Normative vs. Colloquial

There is much disparity between what is officially considered ‘correct’ and what is actually used by native speakers, calling for a distinction between ‘normative’ and ‘colloquial’.

English dictionaries make a clear-cut distinction between the terms ‘normative’ and ‘colloquial’. The Oxford English Dictionary defines ‘colloquial’ as ‘conversational, in or of talk, oral... not used in formal or elevated language’. Conversely, ‘normative’ is defined as something ‘of, deriving from, or implying a standard or norm; prescriptive’. So the dictionary distinction is between strictly spoken and strictly standard. However, when analysing natural spoken language, these terms are not as easily identifiable, as some forms are found both in the standard grammar books and in natural speech while others are commonly accepted as the ‘norm’ but depart from the standard grammar book forms.

Grammar books, often employ the terms ‘normative’ and ‘colloquial’ synonymously with ‘correct’ and ‘incorrect’. Thus, ‘colloquial’ refers to incorrect language found strictly in the spoken language, and ‘normative’ refers to what is officially considered correct language, typically associated with formal high register, as employed by the national broadcasting network, teachers, newspapers, and literature (Ravid, 1995). However, the terms ‘normative’ and ‘colloquial’ say nothing about the correctness or acceptability of a word. Consider, for example, the following pairs of verbs:[2]

(1) Normative - Colloquial frequencies example

	Normative	Frequency	Colloquial	Frequency	
a.	miléti	5%	milíti	95%	‘I filled’
b.	nisíti	97%	niséti	3%	‘I attempted’
c.	makír	36%	mekír	64%	‘I/you/he recognise(s)’

2 The frequencies are calculated from the data collected for this study. See §1.4 for details.

When considering the variation among speakers, as well as within individual speakers, there may be variants that are more frequently used than others and as such, may also be perceived as the norm, although they are not necessarily the normative form prescribed by the standard grammar books. For example, colloquial *milíti* 'I filled' (1a) and *mekír* 'he recognises' (1b) are used more frequently than their normative forms *miléti* and *makír*, though they are not the prescribed normative forms. There are also forms that are formally accepted as 'correct', but are less frequently used by native speakers. For example, the vowel-final verbs *niséti* 'I attempted' and *kivéti* 'I hoped' have been accepted by the Academy of the Hebrew Language as correct in 1996, but their frequency of use remains low).

So strictly spoken vs. strictly literary, correct vs. incorrect, and also frequent vs. infrequent, none of these can make an accurate distinction between normative and colloquial. For the purposes of this study, which seeks to analyse diachronic change from a synchronic perspective, normative and colloquial are regarded simply as synchronic variants where one (the colloquial form) is ***newer*** than the other (the normative form). This definition simultaneously captures the diachronic nature of the change (new vs. old) and also all the above mentioned effects: change occurs in spoken language and the frequency is indicative of the stage of the progression of the change and not an inherent feature of the word. I take the degree of acceptability of the variants to be reflected by their frequency of use (see the discussion in §1.3.6) and so the degree of acceptability also depends on the progression of the change and is not an inherent feature of the word.

When referring to variants in the course of diachronic change, I refer to form only and do not consider change in meaning. Thus, I do not consider cases in which a new meaning is allotted to an old form (e.g. *ganúv* exhibits variation in meaning: the old form meaning 'stolen' and the new form has an additional meaning 'cool'). I only consider cases of segmental variation, such as those in (1) above.

In the following sections, I provide a brief sketch of the language's segmental inventory.

1.2.2. The Segmental Inventory

The changes observed in the MH verb system stems from the change in the language's segmental inventory. The segmental inventory of MH is provided in the following tables:

(2) MH vocalic inventory

	Front	Back
High	i	u
Mid	e	o
Low		a

MH has five phonemic vowels. Phonetically they are all [-ATR], except the back mid vowel /o/, which is [+ATR]. Based on acoustic evidence, the low vowel, /a/, is grouped with the back vowels (Cohen, 2009, Most et al., 2000), though Laufer (1990) classifies it as central.

(3) MH consonant inventory (adapted from Laufer, 1990)

	Bilabial	Labio-dental	Alveolar	Palato-alveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal
Stop	p b		t d			k g			ʔ
Fricative		f v	s z	ʃ ʒ		x	ʁ	ħ ʕ	h
Affricate			t͡s	t͡ʃ d͡ʒ					
Nasal	m		n						
Liquid			l						
Glide	w				y				

Much attention is given in the present study to the pharyngeals and the glottals (see discussion in §1.2.3), since their historical change is the major reason for what seems to be the chaotic nature of the verbal system of today's Hebrew. The pharyngeals *ħ* and *ʕ* appear in the speech of some speakers of oriental descent, not always systematically (see a recent analysis in Pariente 2006, Pariente, 2010). For most speakers, they are not part of the consonantal inventory; the *ħ* is replaced by *x*, while *ʕ* is replaced by *ʔ* or is simply not pronounced. They do, however, survive in the orthography. Given the high degree of literacy of Hebrew speakers, this could contribute to intra-speaker variation.

The glottals *ʔ* and *h* are mostly omitted in regular speech. They appear sporadically in careful speech, but they are rapidly disappearing from the language's inventory (Berman 1981a, b).

In Biblical Hebrew, the pharyngeals and glottals formed the class traditionally named 'gutturals'.[3] For most speakers, the members of this group have not survived (see §1.2.3). As will be seen further in this study, the changes in this class of segments have wreaked havoc on the verb system.

Historically, the velar fricative *x* has three sources: as an allophone of *k* (e.g. *katáv* 'he wrote' – *yixtóv* 'he will write'); as an independent phoneme (e.g. *rixél* 'he gossiped'); and from the historical guttural *ħ* (e.g. *maħáq → maxák* 'he erased'), as discussed above. Only the first source exhibits alternation within the paradigm. In Biblical Hebrew, *x* was in complementary distribution with the velar stop *k*, an allophone that only surfaced post vocalically (e.g. *k̲aváʃ* 'conquered' – *yix̲bóʃ* 'will conquer'; *max̲ár* 'sold' – *yimk̲ór* 'will sell'). In Modern Hebrew, most speakers exhibit variation by sometimes alternating and sometimes not while some speakers do not alternate at all; their grammar represents the end state of the change (Adam, 2002). And so *x* is both an allophone of *k* and an independent phoneme in the language today.

t͡ʃ, *d͡ʒ*, and *ʒ* are not natively part of the language's phonemic inventory; they appear either as allophones due to assimilation and truncation, as in */ʃgia/ → [ʒgi.á]*

3 *r* is sometimes grouped with the gutturals (McCarthy 1994). In this study, I refer to the gutturals excluding *r*, unless otherwise mentioned.

‘an error’, */tiʃava/* → *[t͡ʃa.vá]* ‘swear!’, or as phonemes in loanwords, as in *[t͡ʃips]* ‘chips’, *[d͡ʒip]* ‘jeep’, *[gaʁáʒ]* ‘garage’ (Bat-El, 2002a, Bolozky, 1979, Cohen, 2009).

In Biblical Hebrew, *w* existed as a phoneme, however it did not survive, and was subsequently replaced by *v* (e.g. *walad* → *valad* ‘child’). Today’s *w* is a relatively new phoneme and appears only in loanwords (Cohen, 2009).

ʁ is realised in a variety of ways, as a uvular fricative or approximant, palato-alveolar, or a flapped variant (Bolozky, 1997, Bolozky and Kreitman, 2007, Cohen, 2009, Schwarzwald, 2001). The phonetic realisation of *ʁ* is not pertinent to this study. I use the symbol *r* instead.

1.2.3. The Gutturals

The gutturals are a class of consonants comprising the pharyngeals *ħ* and *ʕ* and the glottals *ʔ* and *h* (McCarthy, 1994, Faust, 2005, Pariente, 2006). In Biblical Hebrew, the gutturals differed in several respects from the other consonants. They did not geminate as other consonants, and instead, lowering of the preceding vowel occurred.[4]

(4) Biblical Hebrew Gemination

	Verb with Gemination		**Verb with Guttural**	
a.	dibber	‘spoke’	t**e**ʔér	‘described’
b.	məd**u**bbár	‘spoken’	məz**o**hám	‘contaminated’
c.	hitnaggéd	‘objected’	hitpaʔér	‘glorified’

In Biblical Hebrew verbs, gemination occurs in two conjugation patterns, traditionally named *piél* and *hitpaél* (and also in *puál*, the passive form of *piél*). When a guttural falls in the geminated slot, gemination is blocked and lowering occurs (4a and b), unless the preceding vowel is already low (4c).

4 Compensatory lengthening also occurred where gemination of gutturals was blocked (Lowenstamm and Kaye 1986). I do not deal with this here, as there is no length distinction in MH. Also, lowering was blocked in some environments in BH (e.g. before a guttural in *Piél*: *nihél* ‘managed’, *niʕér* ‘shook’, *niħeʃ* ‘guessed’).

Gemination did not survive in MH, but the lowering of the vowel preceding a historical guttural (where gemination was blocked) has survived even where the guttural is replaced by a non-guttural. Thus the lowering remains opaque.

(5) Normative Modern Hebrew vowel lowering (compared to the regular verbs)

Regular Verb		**Verb with Lowering**		
n**i**xtáv	'was written'	n**e**xʃáv	*n**i**xʃáv	'was considered'
h**i**xtív	'was written'	h**e**xlít	*h**i**xlít	'decided'
d**i**bér	'spoke'	t**e(**ʔ)ér	*t**i(**ʔ)ér	'described'
medubár	'spoken'	mezohám	*mezuhám	'contaminated'

Yet another case of opacity is the insertion of the low vowel *a* in words ending in a guttural other than *ʔ*, if the final vowel is not already *a*. The glottal stop is not allowed to close a syllable.

(6) Normative Modern Hebrew *a* insertion (compared to verbs with no guttural)

Non guttural-final Verbs		**Guttural-final Verbs**	
hiʃmíd	'destroyed'	hiʃmía(ʕ)	'sounded'
ʃamén	'fat'	taméa(h)	'wondered'
simén	'marked'	siméax	'made happy'
gadál	'grew'	samáx	'was happy'

1.3. Theoretical Background

This dissertation deals with free variation in the course of language change. Gradient acceptability within the context of variation is also discussed as a by-product (§1.3.6). These issues, which go hand in hand with diachronic change, have always been problematic for deterministic theoretical models aimed at providing a single grammar of synchronic linguistic knowledge. The model I propose for free variation is couched within Optimality Theory (§1.3.1), more specifically, within its stochastic variant, Stochastic OT (§1.3.2).

Stochastic OT (StOT), enables to capture not only the grammar of free variation, but also the probability of its occurrence by postulating a range of application for every constraint. If the ranges of two constraints overlap, then the probability of variation increases. If their ranges do not overlap, then the ordinal ranking of the constraints is maintained as defined in traditional OT and the probability of the opposite ranking occurring is extremely low.[5] The probability of occurrence of each variant, which I take to mean the probability of its usage, provides an insight to the natural degree of acceptability of each variant within the perspective of the language. StOT is thus able to capture both free variation and gradient acceptability that are associated with change (see §1.3.6 for a discussion on the relationship between 'frequency', 'probability', and 'acceptability').

In the remainder of this section, I provide an overview of traditional Optimality Theory and the approaches to variation within its framework (§1.3.1) and an overview of StOT (§1.3.2) followed by its application in a learning model, the Gradual Learning Algorithm (§1.3.3). A description of an application of the Gradual Learning Algorithm (GLA) to simulate language evolution, evolOT, is provided in §1.3.4. I then discuss additional theoretical topics that are pertinent to this study. In §1.3.5, I discuss type frequency and token frequency. In §1.3.6, I explain the connection between frequency, probability, and acceptability, as employed in this study. I claim that both probability and acceptability are reflected in the frequency of use. In §1.3.7, I present an overview of analogy and paradigm levelling that account for similarity within the paradigm. Finally, in §1.3.8, I discuss similarity between paradigms and suggest that the levelling among paradigms, which comes in the form of merging of paradigms, is between the most similar paradigms.

1.3.1. Optimality Theory

Optimality Theory (Prince and Smolensky, 1993/2004, henceforth OT) is a constraint-based model of grammar that simultaneously evaluates a set of output candidates for a

---

5 Note that because the theoretical framework of StOT assumes a normal distribution of application for each constraint, the distributions inevitably overlap bordering infinity, even if the ranges of the constraints do not overlap. However, the probability of this overlap ever having a surface manifestation is extremely low.

given input using a set of universal constraints with a language-specific ranking. The constraints often compete with one another and they are inherently violable. The candidate selected as the actual output is the one to least violate the constraint hierarchy. First, the candidates violating the higher ranked constraints are eliminated, if there is a candidate that satisfies them. The surviving candidates are then evaluated against the lower ranked constraints, until only one candidate survives. The winning candidate is the surface output form.

The OT grammar consists of a generator, GEN, and an evaluator, EVAL. GEN generates all the output candidates for a given input. EVAL compares the output candidates against the constraint hierarchy.

Traditional OT selects a single output for each input, and so the notion of variation, which means multiple outputs for a single input, poses a challenge to the theory. Some of the proposals for dealing with variation within OT include Multiple Grammars (Anttila, 2002b, Kiparsky, 1993, Kroch, 1989) and partial ranking of constraints (Prince and Smolensky, 1993/2004, Anttila, 1997a, Anttila, 2002a, Anttila and Cho, 1998).

First, let us consider the partial ranking of constraints. In traditional OT, unless there is evidence to the contrary, all non-rankings of constraints are considered non-crucial, such that all combinations of ranking of the non-ranked constraints will yield the same result. However, there is evidence that for competing constraints that should be ranked with respect to one another, both rankings can apply, each yielding a different attested output, as in the case of variation. Using partial ranking, or crucial non-ranking, OT can derive variation from a single grammar.

Consider the following example of MH spirantisation. MH exhibits alternation between stops and fricatives, whereby fricatives normatively appear post-vocalically, and stops elsewhere. Therefore, the stop-fricative alternation surfaces where the past and future forms alternate prosodically (Adam, 2002).

(7) Normative Stop-Fricative Alternation[6]

Past	**Future**	
ka**f**át͡s	yik**p**ót͡s	‘jump’
ta**f**ás	yit**p**ós	‘catch’
patáx	yi**f**táx	‘open’

Following Adam’s (2002) analysis, the constraints responsible for this state of affairs are as follows:

(8) Spirantisation constraints:

*$_{\sigma}$[CONT: A fricative is not allowed in an onset position

*V-STOP: A stop is not allowed in a post-vocalic position

IDENT-F$_{[\text{CONT}]}$: Corresponding segments S1 and S2 have identical values for the feature [CONT] (i.e. a fricative in the input is realised as a fricative in the output and a stop in the input is realised as a stop in the output).

The ranking *V-STOP » *$_{\sigma}$[CONT » IDENT-F$_{[\text{CONT}]}$ accounts for the alternations in (7) above.[7]

(9) *tafás – yitpós* ‘catch’

Input: ta**f**as	*V-STOP	*$_{\sigma}$[CONT	IDENT-F$_{[\text{CONT}]}$
a. + ta**f**as		*	
b. ta**p**as	*!		*
Input: yit**f**os			
a. yit**f**os		*!	
b. + yit**p**os			*

[6] *t* does not alternate. *k* originating from the historical pharyngeal *q*, also does not alternate. I simplify the discussion here to show how OT deals with variation. See Adam (2002) for a full discussion of the Stop-Fricative alternation in MH.

[7] Because the markedness constraints are ranked here above the faithfulness constraint, it makes no difference if we assume a fricative (*ta**f**as*) or a stop (*ta**p**as*) in the input. Both inputs yield the same resulting output.

The OT grammar is able to capture the stop-fricative alternation. However, due to changes in the language, MH now also allows non-alternating paradigms, as shown in (10). The existence of both alternating and non-alternating paradigms results in surface variation. Recall from §1.2.1, that the terms 'normative' and 'colloquial' refer here to 'old' and 'new' respectively.

(10) Variation in Colloquial Modern Hebrew Spirantisation (Adam, 2002)

Past	**Future**		
	Normative	**Colloquial**	
ka**f**a͡ts	yik**p**ó͡ts	yik**f**ó͡ts	'jump'
ta**f**as	yit**p**ós	yit**f**ós	'catch'
katav	yi**x**tóv	yi**k**tóv	'write

The ranking of $*_{\sigma}$[CONT above the faithfulness constraint IDENT-F$_{[\text{CONT}]}$ is responsible for selecting the candidate with the stop (normative *yit**p**ós*). The opposite ranking selects the candidate with the fricative (colloquial *yit**f**ós*). That both candidates surface in the language, means that both rankings are possible. By not ranking these two constraints with respect to one another, we are able to capture optionality. The constraints are thus 'crucially unranked' with respect to one another. (Crucial non-ranking is marked with a broken line.)

(11) Crucial non-ranking - Spirantisation

Input: ta**f**as	*V-STOP	$*_{\sigma}$[CONT	IDENT-F$_{[\text{CONT}]}$
a. + ta**f**as		*	
b. ta**p**as	*!		*
Input: yit**f**os			
a. + yit**f**os		*	
b. + yit**p**os			*

Now that these two constraints are not ranked with respect to one another, i.e. both $*_{\sigma}$[CONT » IDENT-F$_{[\text{CONT}]}$ and IDENT-F$_{[\text{CONT}]}$ » $*_{\sigma}$[CONT are possible, both candidates are equally optimal. The same grammar accounts for the alternating paradigm *tafás*–

yitpós as well as for the non-alternating paradigm *tafás–yitfós*. The crucial non-ranking of constraints enables the selection of more than one candidate, accounting for free variation. If both are equally optimal, speakers have the option to choose which variant to produce.

The problem with the non-ranking account of variation, is that the two variants are equally optimal. If they are equally optimal, they have an equal probability of surfacing. This means that we would expect their distribution to be 50% each. Traditional OT cannot account for any other distribution. These two variants, however, are not equally distributed in the language. So while the mechanism of variation can be described in traditional OT, the degree of their usage (reflecting the degree of their acceptance) cannot.

Another approach to variation within OT is the Multiple Grammars theory (Anttila, 2002b, Kiparsky, 1993, Kroch, 1989) proposing that variation results from competing invariant grammars. The idea that individual speakers have multiple grammars is independently necessary to account for multilingualism. We need to assume two grammars in order to account for the competence of multilinguals in (at least) two different languages. This idea can be extended to all phenomena that require a ranking of the constraints that departs from the standard ranking assumed, including dialects, registers, and so why not free variation as well? After all, inter-speaker variation is a form of dialect (or idiolect) and intra-speaker variation is a form of register, where one variant can be applied in more careful speech and the other in casual speech. In this account, free variation would mean that the speaker's grammar includes several different grammars; a grammar for each variant type.

The number of possible grammars depends on the number of constraints. If for example, a language has five constraints, then the number of possible grammars is 5! (i.e. 120); this is the number of possible ranking permutations (see Prince and Smolensky, 1993/2004 for the notion of factorial typology).

The number of ranking permutations is reduced if two different rankings yield the same result. Suppose that a speaker has three of the grammars predicted by factorial typology. As each grammar predicts a slightly different output, this accounts for variation. Each variant results from at least one of the grammars. The multiple grammars theory not only accounts for variation, it also provides a mechanism for predicting the variants' frequency, thus improving upon the partial ranking mechanism. If a candidate wins by *n* grammars and *t* is the total number of grammars provided by factorial typology, then the candidate's probability of occurrence is *n/t* (Anttila, 1997a, van Oostendorp, 2004).

The problem is that the theory makes falsifiable predictions. Returning to the spirantisation example above, the three constraints in (8) yield 6 possible grammars, as follows:

(12) Logically Possible MH Spirantisation Grammars

a. *V-STOP » *$_{\sigma}$[CONT » IDENT-F$_{[\text{CONT}]}$
b. *V-STOP » IDENT-F$_{[\text{CONT}]}$ » *$_{\sigma}$[CONT
c. *$_{\sigma}$[CONT » *V-STOP » IDENT-F$_{[\text{CONT}]}$
d. *$_{\sigma}$[CONT » IDENT-F$_{[\text{CONT}]}$ » *V-STOP
e. IDENT-F$_{[\text{CONT}]}$ » *V-STOP » *$_{\sigma}$[CONT
f. IDENT-F$_{[\text{CONT}]}$ » *$_{\sigma}$[CONT » *V-STOP

Assuming a UR with a fricative, grammars (12b), (12e), and (12f) all yield the non-alternating paradigm *tafás–yitfós*. Grammars (12a), (12c), and (12d), yield the alternating paradigms *tafás–yitpós*. So from the six possible ranking permutations, only two grammars emerge. Note, that these two grammars correctly predict that **tapas* will never emerge. The two paradigms, alternating and non-alternating, are predicted in three of the six possible grammars, thus they each have a 50% probability of occurrence. Even if the current state of the language supports a 50% probability for each paradigm (which it does not), the theory cannot account for any change in their probability of occurrence, which is translated to frequency of use. In other words, the theory cannot account for gradual change, difference in registers, style variation, etc.

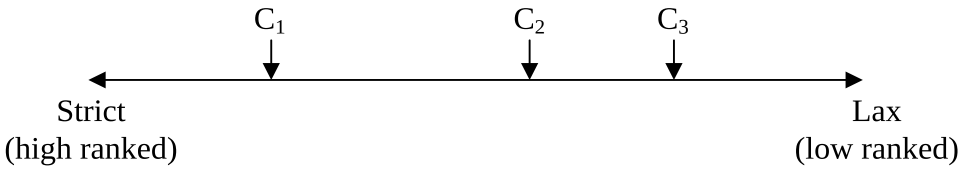
Anttila (2007) compares these two approaches to variation, showing that each approach makes different predictions as to the possible variations. He concludes that the partial ranking approach is empirically superior to the Multiple Grammars approach as it is more restrictive and it excludes patterns that are predicted to be possible under the Multiple Grammars Theory. In the following section, I describe a third approach to variation within OT, Stochastic OT, which is able to make more accurate quantitative predictions about variation than the partial ranking approach. By considering frequency of use, Stochastic OT is able to account for gradual change, as the frequency of use is dependent upon the progression of change. To account for other parameters of interest, such as difference in registers and style variation, all that is required is that the data be so encoded. Thus, the frequency of use of a specific variant may be found to have different frequencies depending on register, style, etc.

1.3.2. Stochastic Optimality Theory

Stochastic OT assumes the same basic mechanism of traditional OT: a generator responsible for generating the output candidates, an evaluator that evaluates the set of candidates according to ranked violable constraints to select the optimal output for a given input. Stochastic OT differs, however, from traditional OT in that it presupposes a continuous scale of constraint strictness (Boersma, 1998, Boersma and Hayes, 2001). Each constraint receives a value that reflects its position on the ranking scale and at every evaluation, a noise component is added to the ranking value, slightly changing the distance between any two constraints. The distance between two constraints determines their interaction. The farther they are from one another, the stricter the ranking. The closer they are to one another, the more lax the ranking becomes, until variable outputs can be produced. The following diagrams, taken from

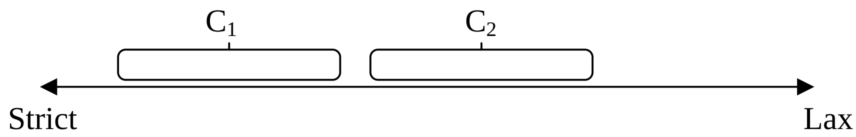
Boersma and Hayes (2001), illustrate this for the constraint ranking C_1 » C_2 » C_3. The diagram in (13) shows a categorical ranking of three constraints, as in traditional OT.

(13) Categorical ranking along a continuous scale



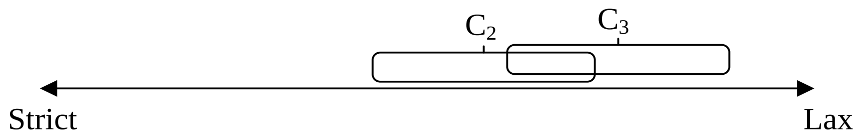
During evaluation, a random value is temporarily added to each constraint to create a value range for the constraints. If the ranges do not overlap, as in (14), then the ordinary categorical ranking of traditional OT is maintained in each evaluation.

(14) Categorical ranking with ranges



If, however, the ranges overlap, as in (15), there will be free ranking and both C_2 » C_3 and C_3 » C_2 may be possible.

(15) Free ranking



Following Boersma and Hayes (2001), the constraint ranges are interpreted as probability distributions determining the probability that the selection point during evaluation time will be at any given distance from the centre of the range. Each constraint is thus represented by a normal (Gaussian) distribution. When distributions overlap, the probability of variation can be calculated allowing to make predictions about the candidates' relative frequencies (see §3.4.1 on calculating probabilities and 3.4.2, including subsections, for an example).

Some may object to applying the stochastic model in this way, claiming that grammar licenses forms and is not a predictor of usage and that competence and performance should not be confused. I would agree that "grammar is grammar and usage is usage" (Frederick Newmeyer's title of his 2003 article in *Language*), but I think that rather than confusing competence and performance, StOT makes an interesting correlation between them. It adds to the grammar a layer of information to which speakers are exposed and it is the driving force of change. I contend that inter-speaker variation affects one's grammar in much the same way as it does the language as a whole.

The grammar that we as linguists describe should be able to capture the dynamic nature of change. If the grammar merely licenses forms and is completely independent of use, then it cannot address the question of bias towards one output alternate over another, nor can it address the process of change. It can only relate to two states: before the change and after the change. The stochastic nature of the model is applied in this dissertation such that not only production is stochastic (which would be reflected by intra-speaker variation), but also inter-speaker variation is stochastic. At the point when a language learner selects one of the licensed alternates, the selection will be made according to the bias already existing in the variable input. In selecting a specific output alternate, the learner thus contributes to the bias, slightly changing the grammar composition, even if the individual speaker's grammar is non-variable.

1.3.3. Learning Algorithms and the Gradual Learning Algorithm

Every theoretical framework must be learnable. Learning algorithms are computational implementations of theoretical models of grammar that tell us whether or not the grammar assumed by the theoretical framework is in fact learnable. If the grammar converges, that is, if it yields a result on every training set, it is assumed to be learnable. The Gradual Learning Algorithm (GLA; Boersma and Hayes, 2001) is the learning algorithm for Stochastic OT.

The input for GLA consists of an underlying representation, a set of constraints, a set of candidates, the frequency of each candidate in the language and markings of each candidate's violations of the constraints. The only thing that is not fed into the algorithm is the assumed ranking of the constraints. The algorithm assumes an initial non-ranking where all constraints are equal. The actual ranking is derived based on the output and its frequency in the language (more specifically, in the corpus being analysed). More detailed information on GLA is provided in Chapter 5.

1.3.4. evolOT

In addition to proving the learnability of a grammar within a theoretical framework, learning algorithms such as the GLA can also be used to simulate language development. If a running of the algorithm simulates a child's learning process, and we assume that the grammar that the algorithm is fed is the language's only grammar (that is, disregarding diversity), then the running of the algorithm simulates the learning process of a generation of speakers. If so, then running the algorithm numerous times can theoretically simulate the learning of numerous generations, assuming that the output of one generation is the input for the next generation.

For this purpose, Jäger (2002b) developed evolOT, a software that implements the GLA and Stochastic OT for simulating language evolution. Given a corpus with frequencies and a set of unranked constraints, the GLA 'learns' the grammar and provides a stochastic constraint ranking. evolOT includes a random generator that produces a sample corpus from this stochastic grammar. The size of the sample corpus is assumed to be the same size as the initial corpus and the frequencies of the inputs are assumed to be the same as in the initial corpus. What may change with each cycle of learning and production (i.e. with each 'generation'), are the relative frequencies of the outputs, reflecting the constraints' ranking and their degree of overlap.

In Chapter 5, I employ evolOT to test the predictions that each type of change makes with respect to the end state of the change, assuming, of course, that nothing intervenes, causing the change process to take a different path.

1.3.5. Frequency

Stochastic OT, the GLA and their implementation in evolOT, all take into account the frequency of use of the variants in the language. Greenberg (1966) investigates the role of frequency in what is today referred to as the theory of markedness, demonstrating that unmarked items throughout the grammar are typically more frequent than marked items.

More recently, frequency has been argued to play a role in all aspects of language: syntax (Givón, 1979, Haiman, 1994 and others), acquisition (Ferguson and Farwell, 1975, Lindblom, 1992), phonology (Bybee, 2001), morphology (Bybee, 1984, Bybee et al., 1994), loanwords (Cohen, 2009), etc. Usage-based models (Bybee, 1985) investigating the role of experience in the formation of linguistic categories and representations began to emerge alongside probabilistic models (Albright, 2008a, Boersma and Hayes, 2001, Daugherty and Seidenberg, 1994, Rumelhart and McClelland, 1986 among others). The frequency of the input was found to be an important factor to the modelling of linguistic systems. High-frequency expressions tend to undergo sound change before low-frequency expressions, as is seen in the extreme reduction of high-frequency phrases, such as *going to → gonna, how are you → hi* (Bybee, 2001). Regularisation has been shown to affect low-frequency paradigms before high-frequency ones (Lieberman et al., 2007). High frequency words and phrases grow strong with repetition, while low-frequency words and expressions are less prominent but gain stability by conforming to patterns used by other items (Bybee, 2007).

Two types of frequency counts are commonly employed: token frequency and type frequency. Token frequency counts the number of times a unit appears in a corpus. Type frequency counts the number of distinct items represented by a specific pattern. For example, English past tense ablaut (e.g. *know–knew*) has a lower type frequency than the regular pattern of *[-ed]* suffixation. This means that there are more English verbs that take the *[-ed]* suffix to form the past tense than verbs whose past tense is formed through ablaut. However, the paradigm *know–knew* may have a higher token frequency than *knit–knitted*, meaning that in a given corpus, *know* and *knew* appear more times than *knit* and *knitted*.

Bybee (2001, 2007) argues that token frequency and type frequency have different effects: *The conserving effect*: Repetition strengthens memory representations for linguistic forms and makes them more accessible. In experiments where subjects are asked to say whether or not a string is a word in their language, they respond much more quickly to high-frequency words than to low-frequency words. This suggests that each token strengthens the memory representation for a word or phrase. Their strength explains why they resist change on the basis of comparison with other forms. Also, within a paradigm, it is usually the higher-frequency form that serves as the base for change. *The reducing effect*: It is a common observation that oft-repeated phrases, such as greetings and titles, tend to reduce phonetically (e.g. *god be with you* → *goodbye*; *how are you* → *hi*). Reductive sound change applies probabilistically across all frequency levels, affecting high-frequency items more quickly and radically than low-frequency items. *Autonomy*: The term refers to the extent to which a word is likely to be represented in the lexicon as a whole and separate unit. Autonomy is probabilistic and is influenced, among other things, by frequency. Highly-frequent words can be accessed independently of related items weakening their connections to other forms, leading in extreme cases to suppletion (e.g. *went* split from *wend* and became the past of *go*). This occurs only in the highest frequency paradigms of a language, in inflectional morphology. In derivational morphology, high-frequency derived forms tend to split off semantically from their bases if they are more frequent than their base.

Type frequency is a major factor determining the degree of productivity of a construction. Constructions that apply to a high number of distinct items also tend to be highly applicable to new items.

In the context of the Hebrew verb system and the changes it is undergoing, I show that low frequency patterns (templates) are replaced by high type-frequency patterns. In other words, ***the direction of change is determined by type frequency. The token frequency of each variant is used for predicting the pace of progression of the change***. Once the direction is set (by type frequency), the forms with the lower token frequency are first to migrate.

Newmeyer (Newmeyer, 2003) argues against stochastic grammars claiming that while language users and hence their grammars are sensitive to frequency, it does not follow that frequency is part of their grammar. Therefore, he concludes that “grammar is grammar and usage is usage” (p.702). I agree. Frequency of use is not assumed to be part of grammar, but is information that is accessible to speakers and is part of their knowledge. This information resides on a tier separate from grammar and therefore studies on grammar are found independently of studies on usage. In this dissertation, I combine the study of grammar and usage and show that they work in tandem with each other and affect one another.

1.3.6. Frequency, Probability, and Acceptability

In this study, I take ‘frequency’ as evidence for both ‘probability’ and ‘acceptability’. This calls for an explanation, as these terms are not naturally synonymous. As discussed in §1.3.5, frequency refers to a count of actual occurrences of an event in the language.[8] Probability refers to the predicted count of an event in the language. Thus, frequency refers to events that have already occurred and probability to events that have not yet occurred. If the calculated probability is then proven true, then the probability of the event reflects its actual frequency in the language.

---

8 An ‘event’ refers to any property under investigation, e.g. a specific verb, tense, syllable type, etc.

Because language is by nature infinite, both frequency and probability can only be calculated based on a finite corpus. The larger the corpus, the more accurately the calculation reflects the event's actual frequency and therefore also its probability of occurrence in the language.

Suppose we have a bag with 50 blue balls and 50 white balls. The fact that we know the number of blue balls and white balls means that we know their actual frequency (which is 50%-50%). If we were to do a sampling experiment, we would expect to extract a blue ball from the bag in 50% of the samplings. This is the blue balls' probability of occurrence, which equals their actual frequency. Now suppose we actually carried out the sampling experiment. We blindly extract a ball from the bag, and it turns out to be blue. If we were to stop the experiment at this point, the measured frequency of the blue balls based on the experiment would be 100% (1 blue ball out of 1 sampling). This is nowhere near the blue balls' frequency or their calculated probability. If we were to continue the experiment and next sample another blue ball, and then a white ball, after the third sampling, the measured frequency of the blue balls will have dropped to 66.67% (2 out of 3 samplings). If we were then to sample another white ball, their frequency would drop further to 50% (2 out of 4 samplings), and so on. The more samplings we do, the closer the result to the balls' actual frequency.

In the above example, the bag with the balls is our finite corpus. Based on this finite corpus, we calculate the frequency of the events (the balls) in the language. The larger the corpus is, the better the measured frequencies in the corpus reflect their actual frequencies in the language. Based on this frequency, we can predict their probability of occurrence in another similar finite corpus. So the calculated frequency of an event in a corpus ideally equals its probability of occurrence in a similar corpus. I further claim that the frequency of the variable event in a corpus also reflects the degree of its acceptability.

‘Acceptability’ refers to the degree in which native speakers regard an event as ‘correct’, and is typically based on tests where subjects are required to score the acceptability of a datum. The tests typically include a list of data under investigation, and subjects are provided with a scale according to which they are to rate the degree of acceptability of each datum. In order to show gradient acceptability and gradient grammaticality, the measurement scale must also be sufficiently gradient. A nominal scale allowing to rate the datum as ‘acceptable’ or ‘unacceptable’, says nothing about the relative acceptability of comparable data. To show relative acceptability, an ordinal scale must be used. But the ordinal scale must also be able to measure the difference between every two selection points. Consider, for example, the following symbol-based scale: 0, ?, *, ** (where 0 denotes an acceptable datum, ? a less acceptable datum, etc.). This is a typical ordinal scale where each symbol indicates less acceptability than the previous symbol on the scale. On such a scale, it is impossible to say whether the difference between ‘*’ and ‘?’ is the same as the difference between ‘**’ and ‘*’ and therefore no meaningful calculations can be done. To enable such calculations, the intervals between successive pairs of measurement points must be controlled. Interval scales, for example 1-5 or 1-10, not only enable better control of the intervals, but they also enable to perform more accurate mathematical operations on the results. But in order for the calculations to be accurate (i.e. to have meaning in the physical world), a reference point, according to which all judgments are compared must be stipulated. Without a shared reference point among the subjects, it is impossible to say whether one subject’s ‘4’ is better than another subject’s ‘3’, even if both subjects use the same interval scale. This is questionable even with a shared reference point (Bard et al., 1996).

The acceptability tests are flawed not only in the measurement scale. It is also impossible to say whether the acceptability judgments reflect the subjects’ own grammar or their impression about what may be plausible, what may be considered a higher register, etc. They may judge forms as perfectly acceptable even if they would never use them themselves. So whose grammar do the acceptability judgments reflect?

For the purposes of this study, I take a more natural approach to acceptability, assuming that speakers use words that are acceptable to them. This is not to say that speakers never use unacceptable forms, however, the more unacceptable they find them, the lower their frequency is expected to be. Within this approach, degree of acceptability is thus also reflected by frequency of use.

1.3.7. Analogy and Paradigm Levelling

From the mid nineteenth century, the Neogrammarians used analogy to account for exceptions to regular sound laws. This view was not unopposed, and analogy was dismissed for not being restrictive enough. Nevertheless, Paul (1891, cited and discussed in Downing et al., 2005) states that words form groups in our mind. He distinguishes between ‘material’ groups and ‘formal’ groups. Material groups include words that have a common element of meaning (e.g. a common stem) whereas formal groups share morphological properties (e.g. all 3rd person singular forms). He further claims that inflectionally related words are more tightly connected than derivationally related words, and that within the inflectional groups, verbs sharing a tense feature are more tightly connected than verbs sharing number or person features.

Seeing words as groups, or paradigms, enabled the discussion of analogical change in terms of the paradigm. The tighter the connection among members of the paradigm, the stronger the preference for uniformity within the paradigm. Inflectionally related words are, therefore, more likely to be subject to levelling than derivationally related words. Thus, analogy is constrained by the confines of the paradigm.

Additional restrictions discussed involved the direction of levelling. Within the paradigm, which members are more likely to change in order to resemble other members of the paradigm? One approach predicts the direction of change based on typological tendencies; levelling is to the isolation form, the most frequent form, the unmarked form, etc. (Bybee, 1985, Kuryłowicz, 1949, Mańczak, 1958, Paul, 1891). Another approach suggests that levelling is to a single surface base from within the paradigm that preserves the most contrasts, predicting that contrasts preserved in the base will be maintained while those neutralised in the base will be levelled (Albright, 2002a, Albright, 2002b, Albright, 2006b).

Within OT, Output-Output correspondence has been suggested to account for the similarity among morphologically related words (Benua, 1997, Kenstowicz, 1996, Raffelsiefen, 1995, among others). Output-Output constraints requiring identity of some feature among morphologically related words are ranked higher than the regular Input-Output constraints.

Two approaches to paradigm uniformity have been proposed within OT. One approach assumes a base (or multiple bases) against which all members of the paradigm are independently evaluated through Output-Output correspondence constraints (Bat-El, 2005, Benua, 1995, Benua, 1997, Burzio, 1998, Steriade, 1999, Albright, 2008b). Under this approach, there is a base that influences all members of the paradigm. Another approach assumes that all members of a paradigm can influence all other members of the paradigm (Kenstowicz, 1996, McCarthy, 2005). This approach also assumes Output-Output correspondence constraints that are responsible for identity within the paradigm, but it assumes no base. Instead, all members of the paradigm are simultaneously evaluated by the constraints.

1.3.8. Similarity

In the previous section, analogy and paradigm levelling were discussed as accounts for similarity among forms, which would otherwise display regular alternations. However, ***the merging of paradigms pertaining to different sub-classes within the Hebrew verb system requires reference to a different kind of similarity: that between paradigms*** and not among the members of a specific paradigm. While token frequency may affect the selection of the base to which all members of the paradigm align, levelling between class paradigms is regulated by type frequency. Consider, for example, the following class paradigms:

(16) Type frequency regulating directionality

	Past	Future	Future Variant		Type Frequency[9]
a.	nixnás	yi-kanés	-	'enter'	113
	nidbák	yi-dabék	-	'be glued'	
b.	nirdám	ye-radém	yiradém	'fall asleep'	13
	nirtáv	ye-ratév	yiratév	'get wet'	

The two class paradigms (16a) and (16b) differ in the prefix vowel of the future tense. In (16a) the future prefix takes an *i* while in (16b) it takes an *e*.[10] However, the paradigm in (16b) has a variant with *i*, which indicates paradigm levelling in progress. The direction of levelling between these paradigms is determined by the number of verbs that follow each paradigm, i.e. type frequency. The paradigm in (16a) has a much higher type frequency (113 verbs) than that in (16b) (13 verbs), and so levelling is to the paradigm with the higher frequency, as evidenced by the variation. The *ye-* prefix (16b) has a *yi-* variant, but the *yi-* prefix (16a) does not have a *ye-* variant. The change is thus to *yi-* rather than to *ye-*.

As described in Chapter 2, the Hebrew verbs are divided into configurations (templates), traditionally termed *binyanim* (here denoted as B1-B5 – see Chapter 2), which are further divided into sub-classes. Each *binyan* has formal characteristics, such as affixes and vowel patterns that are specific to that binyan only. Within the binyan, verbs group into sub-classes, as described below, on the basis of the position of specific consonants (the weak consonants) in the stem.

9 Tarmon, Asher, and Uval, Ezri. 1998. Hebrew Verb Tables. Jerusalem: Tamir Publishers.

10 The inter-vocalic *r* is grouped here with the gutturals, exhibiting lowering as in *yeʕadér* 'will be absent' (see also fn.2).

(17) Hebrew Normative Classes (B1 3rd person **past**)[11]

Stem C		**Position**					
		Initial		**Medial**		**Final**	
Glottal	ʔ	**ʔ**aháv	'love'	ʃa**ʔ**ál	'ask'	saná[12]	'hate'
	h	**h**aláx	'walk	ma**h**ál	'dilute'	tamá**h**	'wonder'
Pharyngeal	ʕ	**ʕ**amád	'stand'	ta**ʕ**án	'claim'	ʃamá**ʕ**[13]	'hear'
	ħ	**ħ**aʃáv	'think'	da**ħ**áf	'push'	bará**ħ**	'escape'
Nasal	n	**n**afál	'fall'	-		tamá**n**	'hide'
(special)	l	**l**akáħ	'take'	-		-	
Glide/V	y	**y**arád	'descend'	k**á**m	'rise'	kan**á**	'buy'
All others		gadál	'grow'				

Each sub-class may have its own peculiarities creating dissimilarities to varying degrees among the sub-classes, as shown in the following table, where the future forms of the verbs presented in (18) are provided.

(18) Hebrew Normative Classes (B1 3rd person **future**)[14]

Stem C		**Position**					
		Initial		**Medial**		**Final**	
Glottal	ʔ	y**o**háv	'love'	**yi**ʃ**ʔ**ál	'ask'	**y**i**s**ná	'hate'
	h	y**elé**x	'walk	yim**h**ál	'dilute'	yitmá**h**	'wonder'
Pharyngeal	ʕ	y**aʕa**m**ó**d	'stand'	yitʕán	'claim'	yiʃmá**ʕ**	'hear'
	ħ	**yaħ**ʃ**ó**v	'think'	yid**ħó**f	'push'	yivrá**ħ**	'escape'
Nasal	n	yi<u>p</u>**ó**l	'fall'	-		yitm**ó**n	'hide'
(special)	l	yi<u>k</u>áħ	'take'	-		-	
Glide/V	y	**y**e**r**é**d**	'descend'	yak**ú**m	'rise'	yikn**é**	'buy'
All others		yigdál	'grow'				

[11] A minus sign (-) means that all verbs with this stem C in the specified position follow the pattern of the regular verbs, and not that there are no verbs with this stem C in the specified position. They do not form an independent sub-class.

[12] From /sanáʔ/

[13] In B3 and B5, the class of pharyngeal finals exhibit a diphthong in the second syllable, e.g. B3 *yid**éa**ʕ* 'he notified', *bit**éa**ħ* 'he insured' and B5 *hitpakéaʕ* 'he burst' *hitbadéaħ* 'he joked'. See Chapter 2 for a list of the *binyanim*.

[14] A minus sign (-) means that all verbs with this stem C in the specified position follow the pattern of the regular verbs, and not that there are no verbs with this stem C in the specified position.

I claim that ***the merging of sub-classes within a binyan occurs between similar sub-classes***. I show that the observed variations that stem from the merging of sub-classes, is systematically between two similar paradigms, the two most similar sub-classes within the binyan. I propose a mechanism for quantifying similarity, such that only the two most similar paradigms can merge. Once the merging is complete, there is nothing to stop the next most similar paradigms from merging. I show a special case of a daisy chain effect that suggests that merging can continue and that patterns are not limited to a single merging cycle. However, they cannot skip a step and they cannot merge with two patterns simultaneously. The merging process is thus restricted.

1.4. Data Sources

This study relies on data obtained from several sources used in different manners for different purposes. The primary source was obtained from recordings of spontaneous speech from radio and TV shows and also of conversations with friends and co-workers. Recordings took place in closed rooms to eliminate background noise that may affect the quality of the transcription. The participants were sometimes told that they were being recorded and other times they were not. Participants who were told that the conversation was being recorded, were not told the purpose of the recording. No significant differences were found between the sets of data where the participants knew they were being recorded and when they did not.

Radio and TV shows are easily accessible and downloadable from the internet and are therefore an excellent source of data. Only shows that include spontaneous casual speech were used. The productions of the hosts of these shows were not transcribed as radio and TV hosts undergo training in ‘correct’ Hebrew. However, studies employing acceptability tests may benefit from comparing the results with generalisations on the characteristics of the language of such trained professionals.

Out of entire recorded conversations, only verbs were transcribed, in all conjugation forms, including past, present, future, infinitive, and imperative. Wherever the author participated in the conversation, my own productions were not transcribed. The transcribed forms were then analysed morphologically according to tense, person, binyan, and whether or not they deviated from the normative form. Deviations from the normative forms are considered as errors, even if they are widely accepted. Because my analysis relies on an error-driven algorithm, the Gradual Learning Algorithm (Boersma and Hayes, 2001), I apply the term ‘error’, although it does not reflect any judgement; it is simply used in order to demonstrate the process of change. The idea is that at the onset of change, deviations are regarded as errors. As the error becomes more frequent, it is regarded as a variant and is more accepted, until it finally becomes the norm itself.

Two hours of recordings were transcribed and analysed with a total of 2964 tokens (see Appendix B). The transcription reveals that in the conversations recorded for this study, speakers utter an average of 23 verb tokens per minute. However, while the amount of data from natural speech is immense, the diversity (types) is limited.

One of my goals in using natural speech was to attempt to obtain complete paradigms from individual speakers. In many cases, especially in the verbs with the highest frequency, this goal was for the most part achieved. Due to the nature of the verb system, it is possible to deduce some of the missing forms based on those that did surface. However, this did not prove to be necessary.

Recording of natural speech is by far the best source of data for this type of study; an experimental setting often leads the participants to provide data that do not necessarily reflect their own language use. However, the process of transcription has proven to be extremely time consuming and requires hours and hours of diligent work in order to achieve a sizeable corpus. As time was not found in abundance, other sources of data were used to enhance the research.

Another source that I used is Bolozky's list of 500 most frequent verbs (Bolozky, 1996). Bolozky compiled the list from a database of 5.3 million words and the verbs were listed by their basic form with the number of appearances in the entire corpus. The corpus was gathered from written sources. From this list it is only possible to comment on the relative frequency of the verb in general, without any information on the frequency of its various forms in the paradigm.

The errors in the recorded data were compared to the normative forms. The normative forms were taken primarily from Tarmon and Uval (1998), listing 3,600 fully conjugated verbs. The Even Shoshan dictionary (1982) was often used to verify normative data. Any deviation from the normative form was regarded as a potential colloquial form.

Some verbs were recorded sporadically out of context. These verbs were analysed in the same way as the recorded conversations, but were not used in running the algorithms; they were used only to provide additional examples.

CHAPTER 2. THE HEBREW VERB SYSTEM

2.1. The Verb Configurations

Hebrew verbs are conjugated according to specific configurations (structures), traditionally named *binyanim* (sg. *binyan*). Structurally, the binyan is essentially the combination of the prosodic structures (also referred to as templates in McCarthy, 1981, McCarthy, 1985) with the vocalic patterns, and affixes (if any). These templates have been shown to be derived from general constraints, such as the minimal word constraints requiring words to be disyllabic and alignment constraints, targeting language specific elements, such as the vocalic pattern and affixes (Bat-El, 2003, Bat-El, 2011).

Modern Hebrew has five binyanim, typically referred to by their 3rd person sg. past stem configuration:

(19) Hebrew Verb Binyanim

B1	CaCáC	katáv	'he wrote'
B2	niCCáC	nixnás	'he entered'
B3	hiCCíC	hizkír	'he reminded'
B4	CiC(C)éC	tipés	'he climbed'
B5	hitCaC(C)éC	hitlabéʃ	'he got dressed'

There are three additional patterns, *huCCáC*, *CuC(C)áC*, and the more recent *hitCuC(C)áC*, all sharing the vocalic pattern {ua} and serve as the passive forms of B3, B4, and B5 respectively.[15] These patterns can be argued to be derived through passivisation processes that change the quality of the vowel via melodic overwriting (Bat-El, 2002b), rather than being independent binyanim (cf. Aronoff, 1994). All verbs that take one of these forms have active counterparts. Also, they differ from the five binyanim in (19) in that they do not have infinitive and imperative forms.

15 Laks (2006) argues that *hitCuC(C)áC* is formed via the blending of *hitCaC(C)éC* and *CuC(C)áC* rather than a passive form generated through the regular passivisation process (Bat-El 2002b). Whatever the status of these three configurations in the language, they are not considered in this study.

A verb stem can be conjugated in any of the binyanim, subject to the limitations of each banyan, and its meaning differs from one binyan to another, as in (20). However, not many verbs are used in all the binyanim.

(20) Verb conjugation example

Binyan	**Past**	**Future**	
B1	katáv	yixtóv	'write'
B2	nixtáv	yikatév	'be written'
B3	hixtív	yaxtív	'dictate'
B4	kitév	yexatév	'inscribe'
B5	hitkatév	yitkatév	'correspond'

The vocalic pattern of verbs is morphologically conditioned. According to Bat-El (2003), the shape of the binyan is regulated by the interaction of the constraints in (21) and a set of constraints on the vocalic patterns.

(21) Constraints on the form of the Binyan

a. ALL FEET RIGHT/LEFT (ALLFTR/L) (McCarthy and Prince, 1993)
The right (or left) edge of every foot is aligned with the right (or left) edge of the prosodic word

b. FOOT BINARITY (FTBIN) (Prince and Smolensky, 1993/2004)
Feet are binary on the moraic or syllabic level

c. *COMPLEX: Complex margins are prohibited

d. MAXV: Deletion of vowels is prohibited

e. FINALC: Words end in a consonant

The vocalic patterns (VP) associated with the binyanim are also regulated by a set of morphological constraints that are unranked with respect to one another: VP1{aa} (e.g. *sagár* 'he closed'), VP2{ia} (e.g. *nisgár* 'he was closed'), VP3{ii} (e.g. *hisgír* 'he extradited'), etc., where the number denotes the binyan and the vowels in the curly brackets denote the vowel pattern of the past form required for the specific binyan. The input is thus specified for the binyan required in the output, and the vocalic pattern constraint must match the binyan specification.

Related verbs share the same stem consonants, and also some level of meaning (e.g. *gadál* 'grew in size', *higdíl* 'enlarged, caused to grow'), though not always. This is the base for the traditional root-based theory of Semitic morphology, which has been advocated in more recent studies in generative phonology (McCarthy, 1981, McCarthy, 1985). That is, in the above example, *gdl* would be the root meaning 'grow in size'. However, this view did not go unchallenged (see Bat-El, 1994, Bat-El, 2003, Bat-El, 2011, Ussishkin, 1999 and references therein).

Because this study focuses on change in form only and does not consider meaning at all, I abstract away from this discussion. Both views yield the same results with respect to the change process and its influencing factors. Wherever I do mention the stem consonants, I make no claim as to their status as a distinct morpheme that carries meaning and simply refer to them as consonants in the verb stem.

B3, B4, and B5 do not alternate prosodically. The prosodic structure of their stems is preserved throughout the tense paradigms (22). B1 and B2 stems do alternate prosodically (Adam, 2002, Bat-El, 1994, Bolozky, 1978b).

(22) Verb stem alternation

	Past	**Present**	**Stem**	**Future**	**Stem**	
B1	ʃamár	ʃomér	Cv.CvC	yiʃmór	vC.CvC	'guard'
B2	nixnás	nixnás	vC.CvC	yikanés	Cv.CvC	'enter'
B3	hixnís	maxnís	vC.CvC	yaxnís	vC.CvC	'insert'
B4	gidél	megadél	Cv.CvC	yegadél	Cv.CvC	'raise'
B5	hitlabéʃ	mitlabéʃ	Cv.CvC	yitlabéʃ	Cv.CvC	'dress'

2.2. Regular Verbs

The configurations presented above refer to regular verbs. The stems of regular Hebrew verbs are disyllabic and they invariably end in a consonant. The second stem vowel deletes before a vowel-initial suffix when the preceding syllable is open (e.g. *gadál-u* → *gadlú* 'they grew', but *nixnás-u* → *nixn**e**sú* 'they entered'), thus reducing the number of syllables (Bat-El, 2008). Most verbs comprise three stem consonants throughout the paradigm, however there are verbs with four and five stem consonants, especially denominative verbs (e.g. *xintréʃ* 'talked nonsense'; *hiʃprít͡s* 'squirted'; *hitbalgén* 'became messy'). Denominative verbs are mostly restricted to the prosodically non-alternating binyanim, B3, B4, and B5 (Bat-El, 1994, Bolozky, 1978b).

Excluding cases of vowel deletion before a vowel-initial suffix, the stems of regular verbs exhibit only two surface prosodic structures, CvCvC and vCCvC; 7 vocalic patterns, out of the 25 possible combinations given the five vowels in the language.

(23) Regular verbs prosodic patterns

Prosodic Structure	**Example**	
CvCvC	gadál	'he grew'
	gadél	'he is growing'
	hitganév	'snuck in/out'
vCCvC	tagdíl	'she will enlarge'
	higdíl	'he enlarged'
	magdíl	'I/he is enlarging'

(24) Regular verbs vocalic patterns

Vocalic Pattern	**Example**	
<a a>	gadál	'he grew'
<i a>	nixnás	'he entered'
<i e>	dibér	'he spoke'
<i i>	higdíl	'he enlarged'
<a e>	mitlabéʃ	'he is getting dressed'
<i o>	yiʃmór	'he will guard'
<a i>	yagdíl	'he will enlarge'

2.3. Weak Verbs

Any verb that does not adhere to the definition of the regular verbs is considered weak. Accordingly, weak verbs are defined as verbs that exhibit at most two stem consonants in at least one form in the paradigm. This definition may be a continuum spanning from the regular verbs (verbs with at least three stem consonants throughout the paradigm) at one extreme to the 'weakest' weak verbs (verbs with no stem consonants anywhere in the paradigm, if there are any: e.g. *hayí-ti* [aíti] 'I was') at the other extreme.[16]

Hebrew weak verbs exhibit a significantly larger number of prosodic patterns and vocalic patterns compared to the regular verbs: seven prosodic templates and 18 vocalic patterns. A sample list is provided below (see Appendix B for type frequencies).

(25) Weak verbs prosodic patterns

Prosodic Structure	**Example**	
CvC	sám	'he put'
CvCv	rat͡sá	'he wanted'
vCCv	hilvá	'he lent (money)'
vCCvC	yamt͡síʔ̤u	'they will invent'
vCvC	horíd	'he lowered'
CvCvC	mihér	'he hurried'
Cv	bá	'he came'

(26) Weak verbs vocalic patterns

Vocalic Pattern	**Example**	
<a a>	mat͡sá	'he found'
<a>	sám	'he put'
<a o>	nasóg	'he retreated
<a e>	yitmalé	'he will be filled'
<a i>	yasím	'he will put'
<a u>	yarút͡s	'he will run'
<e u o>	nesugó-ti	'I retreated'
<e e>	yeléx	'he will go'
<e e a>	neherág	'he was killed'
<e a>	hekám-ti	'I established'
<e i>	hekím	'he established'

16 This is an extreme and rare case in Hebrew, where some forms in the paradigm may be pronounced without the stem glide.

Vocalic Pattern	**Example**	
<i a>	hiʃmá-ti	'I sounded'
<i e>	hilvé-ti	'I lent (money)'
<i i>	himt͡sí	'he invented'
<i i a>	hiʃmía	'he sounded'
<o a>	nolád	'he was born'
<o e>	odéd	'he encouraged'
<o i>	horíd	'he lowered'

The weak verbs are traditionally classified according to the type of the weak consonant and its position in the stem: initial, medial, or final. These sub-classes are named *gzarot* (single *gizra*). A weak consonant is often null in the surface representation and there could be more than one weak consonant in a stem.

The significance of classifying the weak verbs becomes apparent when dealing with levelling. Verbs with weak stem consonants in different positions (or with different weak consonants in a specific position) have different conjugation patterns and can therefore be argued to belong to different paradigms. Paradigm levelling, as laid out in McCarthy (2005), has been argued to operate within the paradigm. Thus, if different types of weak verbs belong to different paradigms, levelling among the sub-classes should not be possible.

In the discussion on variation in the verb system (Chapter 3), I show that variation is more widespread and seemingly chaotic in the weak verbs than in the regular verbs. If this is a continuum, then we could predict that variation increases as verbs are closer to the weak extreme (where fewer stem consonants have a surface realisation). Nevertheless, as variation is strongly affected by similarity, the weakest verbs are actually immune to change as they are typically not similar to any other class of verbs.

In the following sections, I lay out some of the normative alternations exhibited in the weak classes that are relevant to this study.

2.3.1. *Sub-class-based Alternations*

Modern Hebrew exhibits alternations that are restricted to a sub-class of verbs within the binyan.

2.3.1.1 *Verbs with a stem initial glottal stop* in B1 may take one of four future patterns. The glottal stop may originate in a glottal stop in a previous phase of the language (27b-d), or from the historical guttural ʕ (27a). The glottal stop is often not realised phonetically and is therefore marked in brackets.

(27) Stems with an initial glottal stop (B1)

	Past	Future	
a.	(ʔ̲)avád	ya(ʔ̲)avód	'work'
	(ʔ̲)amád	ya(ʔ̲)amód	'stand'
b.	(ʔ̲)asáf	ye(ʔ̲)esóf	'gather'
	(ʔ̲)asár	ye(ʔ̲)esór	'forbid'
c.	(ʔ̲)aráx	ye(ʔ̲)eráx	'last'
	(ʔ̲)azál	ye(ʔ̲)ezál	'deplete'
d.	(ʔ̲)axál	yoxál	'eat'
	(ʔ̲)amár	yomár	'say'

Synchronically, there are no cues in the surface form, and so there is no way of predicting which verb will take which future form. See further discussion in §3.3.1.

2.3.1.2 *Verbs beginning with the glide* /y/ in B1, like the glottal initial verbs, also have two future forms:

(28) Stems with an initial /y/ (B1)

	Past	Future	
a.	yaráʃ	yiráʃ	'inherit'
	yanák	yinák	'suckle'
b.	yarád	yeréd	'descend'
	yaʃáv	yeʃév	'sit'

Here too, verbs with an initial glide must be marked for the future form that applies to them as there is no way of otherwise predicting the shape of their future form.

2.3.1.3 The stem-initial *n* in B1 is deleted when in coda, but only in some verbs.

(29) Stems with an initial *n* (B1)

	Past	Future		
a.	Deleted in coda position			
	nafál	yipól	(*yinpól)	'fall'
	nasá	yisá	(*yinsá)	'travel'
b.	Preserved in coda position			
	naʃám	yinʃóm	(*yiʃóm)	'breathe'
	naváx	yinbáx	(*yibáx)	'bark'

In Biblical Hebrew, the rule applied regularly to all *n*-initial verbs whenever the *n* falls in coda position. This happens in B1 in the future and infinitive and also in B3 throughout the paradigm. In Modern Hebrew, B3 verbs like *hipil* 'dropped' and *hisia* 'transported' are assumed to have been reanalysed as /*hipil*/ and /*hisia*/ respectively, without underlying *n* (Barkai, 1975, Schwarzwald, 1973). The reanalysis is made possible because there is no prosodic alternation in the B3 paradigm. Consequently, the underlying *n* never surfaces, and so speakers have no reason to assume an underlying *n*. This view is debatable, because while there is no alternation in B3, the stem sometimes alternates between the B1 and B3 binyanim: B1 *nafál* 'fell' ~ B3 *hipíl* 'dropped, caused to fall'; B1 *nasá* 'travelled' ~ B3 *hisía* 'transported'. Given the transparent semantic relation, speakers are able to retrieve the missing consonant.

The deletion rule of the stem-initial *n*, however, does not operate on all *n*-initial stems in Modern Hebrew. In fact, most *n*-initial verbs do not undergo deletion (as in (29b)) and so the few verbs that do (29a) must be marked somehow for the deletion to apply.

2.3.2. *Multiple Sub-classes-based Alternations*

Verb sub-classes are characterized by the position of the weak consonant, i.e. the position of the consonant that does not surface in one (or more) form in the paradigm. Thus, if a verb has more than one such position, it belongs to two sub-classes simultaneously. In such cases, the paradigm takes the characteristics of both sub-class paradigms (30a), but not always (30b). In the following table, the past and future forms of B1 verbs with two weak positions are presented in the left columns and are compared with the two single-weak position classes in the two right columns (labelled A and B). The first vowel of each future form is compared with the first vowel in the form in A, and the second vowel with the second vowel in the form in B. The comparable vowel in A and B is bolded and underlined.

(30) Multiple sub-classes-based alternations (B1)

	Past	Future			A	B
a.	(ʔ̲)afá	yofé	'bake'	cf.	y**<u>o</u>**már	yiʃt**<u>é</u>**
	natá	yité	'be inclined'	cf.	y**<u>i</u>**pól	yiʃt**<u>é</u>**
	nasá	yisá	'marry'	cf.	y**<u>i</u>**pól	yisn**<u>á</u>**
b.	yará	yiré	'shoot'	cf.	y**<u>e</u>**ʃév	yiʃt**<u>é</u>**
	her(ʔ̲)á	yar(ʔ̲)é	'show'	cf.	h**<u>i</u>**ʃ(ʔ̲)íl	yarʃ**<u>é</u>**

The verbs in (30a) have more than one weak segment and they exhibit characteristics of more than one sub-class of verbs. *(ʔ̲)afá* has an initial weak glottal stop in the past tense, which disappears in the future tense and the first stem vowel surfaces as [o], as in other glottal initial B1 verbs, (e.g. *(ʔ̲)amár–yomár* 'say'). But as *(ʔ̲)afá* is also vowel final, it has a final *e* in the future tense rather than the regular *a*, as in other V-final B1 verbs (e.g. *ʃatá–yiʃté* 'drink' cf. *gadál–yigdál* 'grow'). *natá* 'was inclined' and *nasá* 'carried' both have an initial *n* that is deleted in the future (as in *yipól* 'he will fall'), and both are V-final. The two paradigms *nasá* 'carried' and *natá* 'was inclined' differ in their final vowel as the former paradigm exhibits *ʔ-Ø* alternation (*nasá* 3rd sg. ms. – *nasʔ̲á* 3rd sg.fem.) whereas the latter does not (*natá* 3rd sg.ms. – *natetá* 3rd sg.fem. **natʔ̲á*). Therefore, *nasá* ‘carried’ is comparable to other *ʔ*-final paradigms (e.g. *saná–yisná* ‘hate’) whereas *natá* ‘was inclined’ is comparable to other V-final paradigms (e.g. *ʃatá–yiʃté* ‘drink’).

As in (30a), the verbs in (30b) also have more than one weak segment, but they are exceptional. The verb *yará* ‘he shot’ has an initial *y* and a final V. As expected, it follows the normal V-final pattern, exhibiting a final *e* in the future form (as in *yiʃté* ‘will drink’). It has, however, an exceptional initial vowel. In other cases in B1, an initial *y* is followed by an *e* rather than by the regular *i*, as in the regular verb *yigdál* ‘will grow’. Speakers need to memorise the exceptions as in (30b). *her(ʔ̲)á* ‘showed’ has a medial glottal stop and a final V. Here too, the verb follows the normal V-final pattern. However, it does not have the regular initial vowel, *i*, as appears in other B3 medial glottal verbs (e.g. *hiʃʔ̲íl–yaʃʔ̲íl* ‘lend’). Speakers need to either memorise or regularise these exceptions.

2.3.3. Morpheme-based Alternations

Some morphophonemic alternations in the verb system are limited to specific verbs or a small number of verbs and no new verb entering the language is expected to exhibit these alternations. The following are examples of such alternations. In the examples, the relevant consonant is underlined.

(31) Verbs with an initial *l* (B1)

	Past	**Future**	
l ~ *Ø* alternation	l̲akáx-ti	ʔekáx	‘take (1st sg.)’
	l̲akáx	yikáx	‘take (3rd ms.sg.)’
	l̲akáx-tem	tikx-ú	‘take (2nd pl.)’
No alternation	l̲aváʃ-ti	ʔel̲báʃ	‘wear (1st sg.)’
	l̲axáʃ	yil̲xáʃ	‘whisper (3rd ms.sg.)’
	l̲amád-ta	til̲mád	‘study (2nd ms.sg.)’

(32) *n*-final verbs (B1)

	V-initial suffix/ No suffix	**C-initial suffix**	
n~ *Ø* alternation	natán	natá-ti	'gave 3rd ms.sg.-1st'
	natn-á	natát-a	'gave 3rd fem.sg.-2nd ms.sg.'
	natn-ú	natá-tem	'gave 3rd pl.-2nd pl.'
No alternation	ratán	ratan-ti	'grumbled 3rd ms.sg.-1st'
	tamán	tamán-ta	'concealed 3rd ms.sg.-2nd ms.sg.'
	karán	karán-tem	'radiated 3rd ms.sg.-2nd pl.'

The stem-initial /l/ appears in the past form *lakáx* 'he took' (31), but not in the future paradigm. This alternation does not occur in any other verb with a stem-initial /l/. Similarly, the stem-final *n* appears in the forms with a vowel initial suffix or with no suffix, as in e.g. *natán* 'he gave' (32), but not in forms with a consonant initial suffix. Again, this alternation does not occur in any other verb. In both these cases, these alternations must include a reference to these particular verbs (Bolozky, 1978a).

CHAPTER 3. VARIATION AND LANGUAGE CHANGE

This dissertation relies heavily on the synchronic variation observed in the Hebrew verb system, which indicates that the system is in the course of change. In phonology, the term 'variation' refers to a state in which one input yields multiple outputs (Anttila, 2006, Anttila, 2007). The term 'change' refers to the process that a language undergoes where one linguistic element is replaced by another linguistic element in the course of time (Shin-ichiro, 2009). As change is gradual, there is an interim phase in which variation occurs and both linguistic elements, the old and the new, coexist. Accordingly, diachronic change always involves variation, although the converse is not always true (Weinreich et al., 1968). In the course of change, during the interim stage, variation can be among speakers (inter-speaker variation), where some speakers have adopted the new linguistic element while others stick to the old (e.g. age difference among speakers). It can also be within an individual speaker (intra-speaker variation), where the same speaker uses both linguistic elements (e.g. different registers).

In this study, I explore ***synchronic variation in the context of diachronic change***, showing that the variation observed in the Hebrew verb system is indicative of a system in change. All of the observed variation is triggered by changes in the language's segmental inventory. Some of the observed changes are the direct result of segmental loss, causing distinct paradigms to merge due to increasing similarity. Other changes, as the stop-fricative variation described in (7)-(11) above, do not result in increasing similarity. In the stop-fricative alternation example, the variation is between alternating and non-alternating paradigms (e.g. *ta**f**ás–yit**p**ós* ≈ *ta**f**ás–yit**f**ós* 'he caught–he will catch'). According to Adam (2002), changes in the language's segmental inventory has caused the process to become opaque and triggered the change. However, the segmental loss did not result in increased similarity, at least not in the sense in which similarity is used in the proposed model, where stem consonants are overlooked (see §4.3.1).

Variation and change have been topics of interest since the nineteenth century, with the Neogrammarians' discussion of sound change (Bloomfield, 1933, Saussure, 1916, 1959). The Neogrammarians viewed sound change as a purely phonetic process that is automatic and exceptionless. Thus, sound change and the variation that comes with it, characterises performance (speech), and not linguistic competence (Anderson, 1985). However, grammar must somehow be involved as it does impose structural constraints on variation and change, preventing rules from applying blindly and without exceptions. Grammar contains variation and change by requiring them to apply selectively and perhaps also gradiently.

Other studies (Kiparsky, 1968, Kiparsky, 1988, Kiparsky, 1995, Anttila, 1997a, Anttila, 1997b, Anttila and Cho, 1998, Reynolds, 1994, Wang, 1969) have attempted to integrate variation within a formal model of linguistic knowledge. In this study, I treat variation in the course of change as an inherent part of speakers knowledge.

3.1. Intra-speaker and Inter-speaker Variation

Language change is gradual and generally follows an S-shaped curve, where the change is slow at first, then proceeds very rapidly before slowing down again (Bailey 1973, Kroch 1989). In the course of change, as a new form spreads through a speech community, speakers do not suddenly jump from always using the old form to always using the new form. Change is gradual and there is always a period in which both forms are available to individual speakers as well as to communities of speakers (Weinreich et al. 1968).

Of course, not all speakers necessarily go through a period of intra-speaker variation. Some language learners in the course of language change may infer the new form from the variable input as others may infer the old form, thus contributing to inter-speaker variation without experiencing intra-speaker variation. However, some speakers may acquire one form and switch to the other form during the course of their life. For them, intra-speaker variation is inevitable.

Following this, I assume in this study that both intra- and inter-speaker variation exist in the course of change in general, and in the change that the Hebrew verb system is currently undergoing in particular. I also assume that inter-speaker variation feeds intra-speaker variation, and vice versa.

It was the intention of this dissertation to describe the change in the verb system of the language rather than the grammar of individual speakers and this is why the particular methodology described in §1.4 was chosen. Only natural data was used because an experimental setting often leads the participants to provide data that do not necessarily reflect their own language use. I recorded spontaneous speech where the speakers gave their consent to being recorded several weeks before the actual recording began and they were not told when they were being recorded. This method of data collection does not cater for intra-speaker variation as there is not enough data from any one speaker because only a few verb forms appear more than once within the same conversation.

However, the corpus does show some evidence of intra-speaker variation, as follows:

- The exact same form repeated differently by the same speaker: e.g. ***e***_xláteti_ vs. ***i***_xláteti_ 'I decided' (where the former is the normative form)
- The same stem produced differently in different forms within the same paradigm: e.g. ***i***_txálti_ 'I started' vs. ***e***_txálnu_ 'we started' (for normative _itxálnu_)
- The same sub-class pattern produced differently for different verbs: e.g. _m**e**vín_ 'he understands' vs. _m**a**xín_ 'he prepares' (for normative _mexín_)

These data do not bear statistical significance, however, none of the attested examples counter the merging patterns discussed in this dissertation. The assumption made in this dissertation is that intra-speaker variation exists to a yet unknown degree. The degree of variation (inter- and intra-speaker) is taken as indication of the progression of the change and therefore does not impact the conclusions on the actual process of change.

3.2. Variation in the Hebrew Verb System

Regular verbs show little variation between normative and colloquial forms (see §1.2.1 for definitions of the terms ‘normative’ and ‘colloquial’). mostly in the stop-fricative alternations (e.g. *ʃa**f**áx* ≈ *ʃa**p**áx* ‘he spilt’, ***k**ibés* ≈ ***x**ibés* ‘he laundered’; Adam, 2002), in the vowels of the past tense of B3 (e.g. *hirgíʃ* ≈ *hergíʃ* ‘he felt’, *higdíl* ≈ *hegdíl* ‘he enlarged’; Bolozky, 1980a), and in the replacement of the 1st person sg. prefix by the 3rd person ms.sg. prefix (e.g. *ani **ʔe**sróf* ≈ *ani **yi**sróf* ‘I will burn’).[17] These variations are observed across age groups. Some variation is found in the surroundings of the historical *ħ*, which in a previous state of the language mandated a following low vowel (Berman, 1978, Schwarzwald, 2001). In the current state of the language, at least for most speakers, *ħ* has merged with *x*. Consequently, the following low vowel has almost entirely disappeared (e.g. *yaxazór* ≈ *yaxzór* ‘he will return’, *t͡soxakím* ≈ *t͡soxkím* ‘we/they are laughing’).

A rare type of variation found in the verb system (both in regular and irregular verbs) involves split paradigms (§3.3.3), where some forms of the paradigm follow one binyan, and some forms of the paradigm follow another binyan (e.g. *paxád–yifxád* ≈ *paxád–yefaxéd* ‘be afraid’, where the past tense *paxád* follows the B1 pattern and the future tense follows either the B1 pattern *yifxád* or the B4 pattern *yefaxéd*). Split paradigms are discussed further in §3.3.3 The variation involved in regular verbs is summarized in (33).

(33) Variation in regular verbs

Type	**Normative**	**Colloquial**	
Stop ≈ Fricative	**k**ibés–yexabés	**x**ibés–yexabés	‘launder past-future’
i → e	h**i**gdíl	(h)**e**gdíl	‘he enlarged’
1st→3rd pr. prefix	**(ʔ)e**sróf	**(ʔ/y)i**sróf	‘I will burn’
a ≈ ∅ / x__	yax**a**zór	yaxzór	‘he will return’
Split paradigm	paxád–**yifxád** (B1)	paxád–**yefaxéd** (B4)	‘be afraid past-future’

[17] I use ≈ to denote variation and ~ to denote alternation.

Weak verbs exhibit the same types of variation as the regular verbs (34). However, the variation in the quality of the vowel is more complex than that in regular verbs, as shown in (35).

(34) Variation in weak verbs

Type	**Normative**	**Colloquial**	
Stop ≈ Fricative	**k**isá–yexasé	**x**isá–yexasé~ye**k**asé	'cover'
i → e	h**i**mt͡sí	(h)**e**mt͡sí	'invent'
1st→3rd pr. prefix	**(ʔ)a**ví	**(ʔ/y)a**ví	'bring'
a ≈ ∅ / x__	te(ʔ)ax**a**rú	te(ʔ)axrú	'you pl. will be late'
Split paradigm	(ʔ)amár–**yomár** (B1)	(ʔ)amár–**yagíd** (B3)	'say'

(35) More variation in weak verbs

a. e ≈ i (bidirectional)

Normative	**Colloquial**		**Normative**	**Colloquial**	
miléti	milíti	'I filled'	nisíti	niséti	'I attempted'
hevín	hivín	'he understood'	hikír	hekír	'he recognised'

b. e ≈ i (unidirectional)

Normative	**Colloquial**		**Normative**	**Colloquial**
exér	ixér	'was late'	-	
yeradém	yiradém	'fall asleep'		

c. e ≈ a (bidirectional)

Normative	**Colloquial**		**Normative**	**Colloquial**	
milé	milá	'he filled'	gilá	gilé	'he disclosed'
hitmalé	hitmalá	'was filled up'	hitnasá	hitnasé	'he experienced'

d. Diphthong simplification (multipath) (discussed in §4.3.7)

Normative	**Colloquial**		**Normative**	**Colloquial**	
bit͡séa	bit͡sá	'executed'	bit͡sáti	bit͡séti	'I executed'

The data above show that in the weak verbs, the variation in the vowel pattern is sometimes bidirectional (35a), both from normative [i] to colloquial [e] and also from normative [e] to colloquial [i]. Similarly, the bidirectional variation is observed from normative [e] to colloquial [a] as well as from normative [a] to colloquial [e] (35c).

It is important to note that the data do not reflect two dialects of Hebrew. Rather, the two forms, normative and colloquial, coexist in the language, sometimes among speakers and often within individual speakers. Furthermore, some of the variation may be specific to register (e.g. *bit͡séa* ≈ *bit͡sá* 'he executed' is limited to formal, high register speech), or socio-economic background (e.g. *niséti*, variant of *nisíti* 'I attempted' is limited to lower socio-economic background). However, I do not address these distinctions; rather, I look at the language as a single system. It is reasonable to assume that variants that are limited to a specific register or to a specific group of speakers will have a lower token frequency in the language than other variants. The effect of these variants on the progression of change will therefore be limited.

3.3. Change

As a diachronic process, a change is from a previous non-variable state to a different non-variable state; in between, variation resides. The process of change always involves variation, but within the context of change, variation is not the desirable (optimal) state and so any variation that results from change is expected to resolve itself until a non-variable state is reached.

Change is inevitable in any living language. Kiparsky (1995) describes change as lexical diffusion, i.e. as an optimisation process that eliminates complexity from the system. The elimination of complexity, or the simplification of paradigms, reflects a general tendency of languages towards regularisation. A well documented example is the change in the English verb inflection (Hare and Elman, 1995). Old English had at least ten different past tense markings: at least six 'strong' classes of verbs that were inflected through a stem vowel change (ablaut), as in *give–gave*, and four subclasses of 'weak' verbs that took variants of the suffixes *-t* or *-d*. According to Hare and

Elman (1995), the system has since simplified dramatically as the weak classes coalesced into one, and the change gradually spread to the ablaut classes, resulting in the modern system in which the regular suffix *-d* applies to most verbs with some exceptions, remnants of the ablaut strong classes.

So change is typically to a simpler system, which in the English past tense example means fewer classes. But before changing, the more complex system was stable for a while. If indeed there is pressure towards simplification, why was the more complex system able to exist? And what happened to eventually undermine its stability and cause it to change?

In Hebrew, a number of factors have contributed to the apparent stability of the verb system prior to change. Perhaps the most prominent factor is the degree of regularity in the system. Although the Hebrew verb system has a number of distinct patterns (sub-classes) within the system of the binyanim, as long as there were cues that enabled speakers to identify each verb as belonging to one of the sub-classes, the system was able to remain relatively stable. However, Hebrew underwent phonological changes that affected the morphological system of the verbs, causing it to become unstable.

Three of the segments that underwent change are: *ʔ*, h, and *ʕ*. In previous phases of the language, the glottal stop appeared in onset position, but was banned from appearing in coda position. The glottal fricative appeared in coda position, but was rare. The historical guttural *ʕ* was allowed in both onset and coda positions (Berman, 1978, Sumner, 2002, Sumner, 2003, Bolozky and Kreitman, 2007). Two major changes that are relevant to the discussion at hand occurred in the language's segmental inventory: the guttural *ʕ* has merged with the glottal stop *ʔ*, and the glottals, both *ʔ*, *h*, and the *ʔ* derived from the historical *ʕ*, have become optional. For example, *ʔaháv* ≈ *ʔaáv* ≈ *aháv* ≈ *aáv* 'he loved' are all synchronically accepted variants, as are

ʔasá ≈ *asá* ‘he did’ from historical *ʕasá*.[18] As a result, the frequency of use of the glottals is rapidly diminishing (Berman, 1981a, Berman, 1981b) and they are on the verge of extinction. These changes in the segmental inventory are external to the verb system, but affect it considerably. Indeed, they are the primary cause of the variation found in weak verbs, as described further in this chapter.

In what follows, I demonstrate the change in the Hebrew verbs and show that much like in the case of English past inflection, in Hebrew too, a simpler system means fewer patterns, rather than simpler structures. I claim that the ***Hebrew verb system exhibits variation as a result of two kinds of process: merger and substitution***. I demonstrate in §4.3 that merger is based on similarity between comparable forms and that similarity is based on structural identity. Substitution is shown to be regulated by frequency, such that the more frequent forms/patterns replace the rarer forms/patterns.

A number of strategies can be applied to achieve fewer patterns. Patterns may be eliminated through extinction or substitution (see §3.3.1). Some patterns can cease to exist altogether, resulting in a gap in the paradigm, or they can be replaced by another pattern, thus becoming extinct without leaving gaps in the paradigm. The frequency of the pattern plays a key role in predicting which pattern will survive. The more frequent the pattern is, the more likely it is to survive.

Patterns can also merge to form a single paradigm (§3.3.2). In this case, the change is bidirectional and is therefore likely to operate at a slower pace. The end result may be a single non-variable paradigm comprising forms from both paradigms, although this outcome is not considered here.[19] For merging paradigms, similarity

---

[18] From the perspective of the language, nowadays the glottals are completely optional. It is possible, however, that their surface distribution varies among registers. The collection of data for this study does not take register into account.

[19] In order to explore whether or not such an outcome is possible, much more data must be collected from each subject to ascertain that they have mixed paradigms rather than intra-speaker variation. The data collected for the purposes of this study do not cater for this.

plays a key role in the selection of the patterns that can merge. Similar patterns are more susceptible to merging than dissimilar ones.

3.3.1. Extinction/Substitution

In a previous non-variable state of B1 (Phase I in (36) below), the glottals and gutturals surfaced regularly in stem-initial position, the past tense of stems with an initial guttural followed the regular *CaCáC* pattern, and the future tense followed the pattern according to the stem-initial 'weak' segment, *yaCaCoC* for *ʕ* and *h*, and *yeCeCoC* or *yeCeCaC* for *ʔ*.[20]

(36) Extinction and substitution in B1 (the weak segment is underlined)

	Phase I		Phase II		
	Past	**Future**	**Past**	**Future**	
Survives	ʕavád	yaʕavód	avád	yaavód	'work 3rd past/future'
	halám	yahalóm	alám	yaalóm	'hit 3rd past/future'
Substituted	ʔasáf	yeʔesóf	asáf	yaasóf	'collect 3rd past/future'
Extinct	ʔaráx	yeʔeráx	aráx	Ø	'last 3rd past/future'

In Phase I, the past tense has a single pattern, *CaCaC*, but three future patterns. Verbs with a stem initial *ʕ* or *h*, have only one available future pattern, *yaCaCoC*.[21] However, for verbs with an initial glottal stop, there are two future patterns to choose from, *yeʔeCaC* and *yeʔeCoC*. In order to choose the correct pattern, speakers need to list in the lexicon at least some *ʔ*-initial verbs with their future form. Note that the regular B1 verbs, where the stem-initial consonant is 'strong', take a completely different future pattern (*yiCCaC*).

[20] This non-variable state is a theoretical phase that may or may not have actually existed.

[21] The fact that there is only one future pattern for *ʕ*- and *h*-initial B1 stems could be accidental, or perhaps it is to avoid a sequence of three syllables with a low vowel, *yaCaCaC*, a pattern which is not found anywhere in the verb system (cf. *xazaka* 'strong fm.' adjective). The first low vowel is mandated by the gutturals, and the second low vowel is achieved through harmony with the first vowel, as is also exhibited in *ye(ʔ)esof* 'will gather' (Bolozky 1980a). The glottal stop differs from the other gutturals in that it mandated a preceding front mid vowel *e* rather than *a* as the other gutturals, or *i* as all other consonants.

The merger of ʕ and ʔ and the subsequent evanescence of the glottals altogether has resulted in the loss of contrast between the three future patterns. With no weak segment on the surface (Phase II Past in (36)), speakers are not left with sufficient indication when to use which pattern. To choose the correct future pattern, speakers now need to list all verbs in the lexicon along with their conjugation pattern.

Instead of loading the lexicon with listed paradigms, speakers reduce the number of patterns by resorting to either substitution or extinction. *yeesóf* has been replaced by *yaasóf*, and *yeeráx* has become extinct (**hayeʃiva* ***taarox*** *ʃaa* 'the meeting will last one hour') and a different verb is more likely to be used instead (*hayeʃiva* ***timaʃex*** *ʃaa* 'the meeting will last one hour').[22] Thus, of the original three future patterns only one survives. The surviving pattern (*yaCaCoC*) is the one with the highest type frequency (36). The more infrequent patterns, *yeCeCaC* with a type frequency of only two verbs, and *yeCeCoC* with a type frequency of 14 verbs, are the ones to become extinct. Note that the two verbs with the rarest future pattern (*yeCeCaC*) have completely lost the future paradigm, whereas the *yeCeCoC* pattern is the one replaced. In Chapter 4, I propose a model for quantifying similarity. Within this model, *yeCeCoC* is more similar to *yaCaCoC* than *yeCeCaC*.[23] As one of the identical vowels is achieved through harmony, the *yeCeCoC* pattern differs from *yaCaCoC* in only one vowel (<e o> vs. <a o>), whereas *yeCeCaC* in two (<e a> vs. <a o>). See more on similarity effects in Chapter 4.

3.3.2. Merger

As mentioned above, patterns can also merge. The data in (37) show three distinct paradigms in B4. The distinctive elements are not only the weak segments, but also

---

22 See Raffelsiefen (2004), Albright (2006a), Rice (2006) and references therein on the conditions of paradigm gaps.

23 It could be argued that it is more similar because the stressed vowel is identical, however, other such cases suggest that stress does not play any role in the change in the verb system. Stress is very regular and appears on the final syllable, except when a consonant initial suffix is attached, in which case it is penultimate.

the vowel in the past tense in forms with no suffix or with a following consonant initial suffix.

(37) Merger in B4 – Phase I (contrastive segments are underlined)

Paradigm type	**Past**			**Future**	
	3rd sg.	**1st sg.**	**3rd pl.**	**3rd pl.**	
ʔ-Final	milé	miléti	milʔú	yemalʔú	'fill'
V-Final	nisá	nisíti	nisú	yenasú	'try'
ʕ-Final	bit͡séʕ^{a}	bit͡sáʕti	bit͡sʕú	yevat͡sʕú	'execute'

The data in (37) show that in Phase I, the three types of paradigms remain distinct throughout the entire paradigm. Following the loss of the weak segments, the contrast between the three paradigms is almost completely lost (38). The future tense is identical for all three paradigms as are some forms in the past tense. The final stem vowel in the past is the only remnant of the old system, leaving speakers without sufficient cues when to use which contrastive vowel in the past tense.

(38) Merger in B4 – Phase II (contrastive segments are underlined)

Paradigm type	**Past**		**Future**	
	3rd sg.	**3rd pl.**	**3rd pl.**	
ʔ-Final	milé	milú	yemalú	'fill'
V-Final	nisá	nisú	yenasú	'try'
ʕ-Final	bit͡séa	bit͡sú	yevat͡sú	'execute'

To resolve this, many speakers have merged the three paradigms, resulting in a single past pattern: *milá*, *nisá*, and *bit͡sá*, respectively.[24] However, as mentioned in (35) above, both *miléti* and *milíti* and also *nisíti* and *niséti* coexist. This indicates that speakers have not merely replaced one pattern with another, as described in §3.3.1, but they have juxtaposed the two forms and merged them. When compared, the two competing forms are equal, and so some speakers may choose one pattern while other

[24] The *bit͡sá* 'he executed' type of paradigm is found primarily in high register and has therefore a lower frequency in the language than *bit͡séa*.

speakers choose the other, resulting in variation among speakers. Merging patterns coexist alongside one another and it is impossible to predict whether they will continue to exist as distinct paradigms but with different members, or whether they will fuse into a single paradigm that will be a combination of the original paradigms. Because change is bidirectional, the process takes longer as the frequency of the two paradigms changes at a slower pace. This is because as members defect to one group, new members join from the other group. For a more detailed discussion, see Chapter 4.

3.3.3. Split Paradigms

Split paradigms are a phenomenon in which some forms in the paradigm follow one pattern and other forms in the same paradigm follow a different pattern. In MH, this phenomenon has several types of manifestations, which, as observed in Bolozky, 1980b, mostly tend to split along tense lines.

(39) Hebrew Split Paradigms (Bolozky, 1980b)

Split Paradigm				**Expected Complete Paradigm**[25]			
Past		**Future**			**Past**	**Future**	
amár	(B1)	yagíd	(B3)	(B1)	amár	yomár	'say'
				(B3)	*higid	yagíd	
yaxól/yaxál	(B1)	yuxál	(B3 passive)	(B1)	yaxól/yaxál	*yexel[26]	'be able to'
				(B3)	*huxal	yuxál	
nigáʃ	(B2)	yigáʃ	(B1)	(B1)	*nagaʃ	yigáʃ	'approach'
				(B2)	nigáʃ	*yinageʃ	
erá	(B4)	yeerá	(B1)	(B1)	*ará	yeerá	'occur'
				(B4)	erá	*yeara	
paxád	(B1)	yefaxéd	(B4)	(B1)	paxád	yifxád	'fear'
				(B4)	*pixed	yefaxéd	
ʃamán	(B1)	yaʃmín	(B3)	(B1)	ʃamán	*yiʃman	'become fat'
				(B3)	hiʃmín	yaʃmín	
bagár	(B1)	yitbagér	(B5)	(B1)	bagár	*yivgar	'mature'

[25] This column provides the complete paradigm pair for the split paradigm in the left column. Forms with an asterisk have either never existed or are obsolete.
[26] Cf. *yaʃáv-yeʃév* 'sit'

Split Paradigm				Expected Complete Paradigm[25]			
Past		**Future**			**Past**	**Future**	
				(B5)	hitbagér	yitbagér	
neeʃám	(B2)	yuaʃám	(B3 passive)[27]	(B3 passive)	huaʃám	yuaʃám	'be accused'
				(B2)	neeʃám	yeaʃém	
neerát͡s	(B2)	yuarát͡s	(B3 passive)	(B3 passive)	huarát͡s	yuarát͡s	'be admired'
				(B2)	neerát͡s	yearét͡s	
yasád	(B1)	yeyaséd	(B4)	(B1)	yasád	*yisad	'establish'
				(B4)	yiséd	yeyaséd	
nit͡sav	(B2)	--		(B2)	nit͡sav	*yiyat͡sev	'stand motionless'

In (39) above, the paradigms that are synchronically split are presented in the left column. For each split paradigm, the complete paradigm pair is provided in the right column. These are the two paradigms involved in the formation of the split paradigm. The forms with the asterisk have either never existed or are obsolete. Paradigms where both past and future forms exist (i.e. where the paradigm is complete synchronically) have a lower token frequency in the language as they are typically found in high register only.

Bolozky (1980b) argues that split paradigms typically arise "where intra-binyan relationships are obscured by considerable morphophonological change from one subparadigm to another" (p.122). That is, paradigms split because of morphophonological opacity of the tense relationship within the binyan. As no split paradigms involving recently formed verbs exist in MH, he concludes that the process is probably not very productive.

Productivity can only be measured synchronically based on the grammar at a specific point in time. Paradigms split as a result of diachronic change, and so they cannot be productive synchronically, only diachronically. However, while it is not possible to predict how or which paradigms will split, some indication of which

27 Bolozky states that the future forms used for *neeʃám* and *neerát͡s* are *hooʃám* and *hoorát͡s* respectively, which are the normative *hufal* forms for the guttural initial verbs. However, the *hufal* paradigm has completely levelled to match the rest of the paradigm with the {ua} vowel pattern characteristic of the passives.

paradigms ***could possibly*** split is available synchronically, as argued below. Before going into the conditions of split paradigm formation, I briefly describe the types of paradigm splits.

There are several types of split paradigms. One type is when a verb has one or more of its forms in the tense paradigm realised with a morphologically unrelated word (*suppletion*), as in English *go–went*, where the past tense *went* (from obsolete present **wend*) is not morphologically related to *go*. In MH, this type of split is rare, found only in one form: the B1 past form *amár* 'he said' takes a morphologically unrelated form *yagíd* 'he will say' as the future form. English and Hebrew are in this respect at different stages of change. In English, neither the past tense of *go*, nor the present tense of *went* exist synchronically, and the language is left with a 'pure' split paradigm. In Hebrew, the future tense of *amár* (*yomár*) 'say' still exists, but is nowadays used in more formal language. The past tense of *yagíd* (**higid*) is easily formed and understood, but is rarely used in any context. So for the *yagíd* form, the paradigm is truly split as is the English case, but for the *amár* form, variation exists for the future form *yomár* ≈ *yagíd* where the former is formal and the latter is more casual.

A more common type of paradigm split in MH is known as *binyan split*, where the forms of the two tenses of a verb are morphologically related (they have the same stem consonants) but they take different binyanim. In some cases, binyan split happens in order to fill a gap, where a form is missing, for example the B1 form *yaxólta* 'you were able' (or colloquial *yaxálta*) takes the future *hufál* form *tuxál* (see Gesenius, 1910 §53u and §69r). Neither the B1 future nor the *hufál* past exist. According to Gesenius, *yuxál* could have derived from *yoxál* in order to distinguish it from the future tense of *ʔaxál* 'he ate', *yoxál* 'he will eat'. But why would the future tense of *yaxál* be anything like that of *ʔaxál* and not like that of *yaʕáv* (with an initial *y*)? Gesenius brings an alternative explanation whereby the future of *hufál* is used instead. Although elsewhere (§53u) a claim is made whereby several supposed future tense of *hufál* verbs are actually future forms of passive B1 verbs, and although *yuxál* is not listed among these verbs, nevertheless, the connection between the future tense of *hufál* and that of B1 (*paál*) is made. And in both cases, the verb's structure and especially its vowel pattern is indicative of passive even though the meaning of *yuxál* is active.

Split paradigms need to be memorised as there is no morpho-phonological process deriving them. They are the result of frequency of use. When two paradigms with a similar meaning coexist, in time, some forms of one paradigm may become more formal (and therefore less frequent). As the frequency of these forms decreases, the frequency of the parallel forms in the other paradigm increases. Given enough time, the low-frequency forms disappear and a single paradigm arises with the high-frequency forms of the two paradigms. When a new verb enters the language, it receives the productive regular morphology at the time. This makes it more transparent and easier to learn.

The two paradigms *amár* 'he said' and *higíd* 'he told' existed simultaneously in Biblical Hebrew.

(40) *amár – yagíd*

Past	**Present**	**Future**	**Infinitive**	
amár	omér	yomár	lomár	'say'
higíd	magíd	yagíd	lehagíd	'tell'

For whatever reason, *yomár* and *lomár* became less frequent. As this happened, the frequency of *yagíd* and *lehagíd* increased, slowly replacing the infrequent forms. In parallel, as the future and infinitive of *amár* were being replaced, the past and present of *higíd* became obsolete. Some remnants of the older forms exist (e.g. *magédet atidot* 'fortune teller'; *klomár* 'that is to say'), but as verbs, these forms are either extinct, or highly literary.

Going back to the idea of productivity, as mentioned above, the productivity of split paradigms cannot be measured synchronically. However, the existence of morphological variation may be indicative of diachronic productivity. Morphological variation in the binyan system is defined as two or more related verbs that occur in different binyanim, but share the same thematic grid and meaning (Laks, 2011).

(41) Morphological variation (Laks, 2011)

a. nirtávti ≈ hitratávti ba-géʃem 'I got wet (B2≈B5) in the rain'
b. dan niftár ≈ hitpatér me-ha-orxím 'Dan got rid (B2≈B5) of the guests'

In the examples in (41), there is at least one context in which the two related forms are interchangeable. Such examples set the right conditions for paradigm splitting. This type of variation can resolve itself in two ways: either one paradigm will become dominant at the expense of the other, until it will replace it completely, or the forms in the two binyanim will merge to create a single, albeit split, paradigm. In addition, recall the example in (36) above, where out of the three comparable B1 paradigms, one is the attractor (*yaavód* 'he will work') and one is substituted (*yeesóf* 'he will collect'). That is, its members slowly migrate to the *yaavód* class paradigm. The third paradigm (*yeeráx* 'he will last') is the one with the lowest type frequency. It loses its future tense altogether. The loss of part of the paradigm is not accidental. It is the offending part that is lost; the part that for whatever reason could not be levelled. By eliminating the future tense of the *aráx* paradigm, the levelling of the future tense of the entire group is not compromised. The gap that is created in the *aráx* paradigm is another prime setting for a paradigm to split. Whether or not this gap will be filled by another verb is unpredictable. Notice, however, that as Bolozky claims, here too, the split is along tense lines.

It is possible, perhaps even probable, that some paradigms will be replaced and others will be split. Which paradigm will take which option will most likely depend on the frequency of use of all of its members.

It is not a coincidence that many of the split paradigms involve weak verbs. Weak verbs, especially the low-frequency ones, already involve some degree of memorisation, and so the lower-frequency forms in these paradigms (typically the future forms) are more likely to split (e.g. *neerác–*yearéc*, *nicáv–*yiyacév*, *yasád–*yisád*).

3.4. The Grammar of Change: A Stochastic Analysis

So far in this chapter, I have described the types of changes that the patterns of weak verbs undergo, namely extinction/substitution and merger. I have also suggested that both frequency and similarity are important factors in determining which patterns are likely to survive (the more frequent ones) and also which patterns are more susceptible to merger (the more similar ones). In the following sections, I propose a way of testing these claims using the principles of Stochastic OT and Boersma and Hayes' (2001) Gradual Learning Algorithm described in §1.3.2 and in Chapter 5.

Recall that the Gradual Learning Algorithm adds a small noise factor whenever a constraint is called into action. The algorithm is also error driven. This means that whenever a deviation from the expected output is encountered in a specific corpus, the constraints involved move a little bit either farther apart or closer together, depending on the nature of the error. Taking into account that each constraint has a range of application and also a distribution of application (such that the constraint is more likely to apply closer to the mean than at the edges), the probability of variation can be calculated if the two constraints overlap. The following is an example in which the future prefix for 1st person sg. is gradually being replaced by the 3rd person ms.sg. prefix (Bolozky, 1999, Bolozky, 2003a, Ravid, 1995).

It is possible that the proximity of the pronoun's high vowel to the vowel initial verb (following the loss of the glottal stop of the 1st pr. future prefix) has triggered the change. Thus, *ani ʔekné* → *ani (y)ekné* → *ani yikné* 'I will buy'. Whatever the trigger may be, it no longer applies as the 3rd pr. ms.sg. prefix is also exhibited when the pronoun and the vowel initial verb are not adjacent (e.g. *ani lo yikné* ‘I will not buy’).

The following table demonstrates the progression of change in the corpus recorded for the present study (see §1.4), assuming that no external interferences occur.

(42) 1st person sg. prefix → 3rd person sg. prefix

	Future Prefix	Total 1st pr.	1st pr. yi-	Total yi-
a.	**1st sampling**			
	Tokens	27	8	54+8
	Error rate	8/27 = 29.63%		
b.	**2nd sampling**			
	Tokens	19	14	62+6
	Error rate	6/19 = 30%		
c.	**3rd sampling**			
	Tokens	9	17	68+3
	Error rate	3/9 = 30%		

In the example in (42a), the corpus includes 54 instances of the 3rd pr. future (e.g. *yigdál* ‘he will grow’) and 27 instances of verbs in the 1st pr. future. Out of these 27 instances, 8 instances appear with the 3rd pr. future prefix *yi*- (i.e. *yigdál* instead of *egdál* ‘I will grow’). The error rate is approximately 30%. To clarify, the corpus includes 27 tokens of verbs in the 1st pr. future form (regardless of the prefix used) and 54 tokens of verbs in the 3rd pr. future forms. Out of the 27 1st pr. tokens, 8 had a *yi*- prefix instead of the expected *(ʔ)e*- prefix. This means that the language learner is actually exposed to the 3rd pr. future prefix 8 instances more (i.e. 62) and to the 1st pr. future prefix 8 instances less (i.e. 19). These 19 instances of the 1st pr. prefix are the input for the next sampling, as in (42b). Assuming a constant error rate of 30%, every running of the algorithm on the same corpus results in fewer 1st pr. future prefixes and more 3rd pr. future prefixes (as shown in the 2nd and 3rd sampling in 42b and 42c).[28]

[28] Of course the error rate in an actual diachronic change process varies depending on the progression of the change. The error rate is probably low at first, then it increases quite rapidly as the new form becomes increasingly frequent, until it finally nears a plateau as the number of instances of the

This tendency will continue until the 1st pr. future prefix becomes extinct. The process slows down as the number of tokens of the prefix to which the speakers are exposed diminishes. Of course, the rise in the use of the 3rd pr. prefix for 1st pr. marking raises the frequency of the 3rd pr. prefix (*yi*-) in the language. This is important because no distinction is made between *yi*- for 1st pr. and *yi*- for 3rd pr. This distinction must be made by the preceding pronoun (*ani yigdál* 'I will grow' vs. *hu yigdál* 'he will grow'). Moreover, it should be noted that it is the 1st pr. prefix that is being replaced as the 3rd pr. prefix remains intact. This is determined by frequency; however, since type frequency is not relevant here, token frequency selects the surviving prefix. It is the prefix with the higher token frequency that survives.

In the following section, I employ the laws of probability and the principles of Stochastic OT to calculate the rate at which a given error is likely to occur in a given corpus.

3.4.1. Calculating Probabilities

The probability of an occurrence is interpreted as the frequency of that occurrence. The difference between probability and frequency is that the former calculates the likelihood that an event that has not yet occurred will occur and the latter calculates the number of events that have already occurred in a given data set (corpus). Because change is slow, we can predict the probability that an event will occur based on the frequency in which it has already occurred. Corpora differ in size, and so the number of events differ accordingly. However, the relative frequency of an event converges to its true probability as the number of experiments increases. This is known as the Law of Large Numbers (Bod et al., 2003). So the more times we run the algorithm on the corpus (i.e. the more sampling we do), the closer the result will be to the actual frequency of the event in the language, until the calculated probability of the

---

original form is reduced considerably, thus forming an S-shaped graph. I use here a constant error rate to simplify the explanation. This does not affect the end result, that the number of instances of the original form, in this example the 1st pr. prefix, diminishes over time.

occurrence of the event equals its current frequency in the language (see §1.3.6 above).

To calculate probability, the following will be employed:

(43) Definitions (Bod et al., 2003)

Ω = the sample space (corpus). This is the total number of events.
A = an event that is a subset of Ω
|A| = the number of items in A
P = probability

(44) The probability formula of sampling

$$\mathrm{P(A)} = \frac{|\mathrm{A}|}{|\Omega|}$$

The probability of an event equals the number of counted events divided by the total number of all events (i.e. the sample space). To illustrate this, I provide a simple example in (45). In this example, we have an imaginary corpus of 50 verbs. 20 verbs are in the past tense, 25 in the present tense, and five verbs are in the future tense. Let us calculate the probability of sampling a verb in the past tense.

(45) Probability calculation example

Ω = 50
A = Past tense
|A| = 20

$$\mathrm{P(\{Past\})} = \frac{|\{\mathrm{Past}\}|}{|\Omega|} = \frac{|20|}{50} = 0.4$$

The probability of sampling a verb in the past tense in this corpus is 0.4 (or 40%), assuming that every sampling yields a result.[29]

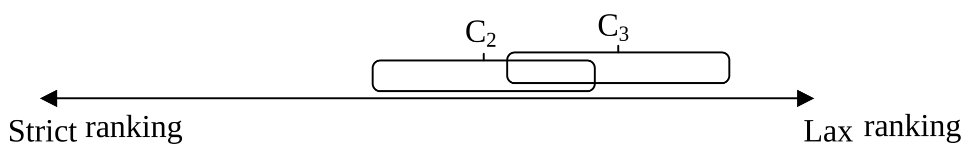
When dealing with constraints and their ranking, especially if constraints are assumed to have ranges, this formula is not enough. Recall that according to the Gradual Learning Algorithm, each constraint is assigned a value that reflects its relative ranking. Then, during evaluation a random value is added to create a range of

29 This is not always the case; sometimes sampling does not yield a result (e.g., some types of slot machines). For this type of probability, the formula is more complex. This is not the case here.

application. If the ranges of two constraints overlap, then there is a chance that the opposite ranking will occur. The greater the overlap, the greater the chance that the opposite ranking will occur. In such cases, we can calculate the probability of occurrence of the opposite ranking. This will determine the error frequency of a specific event in the language. Recall also that at every evaluation, a noise value is added to slightly alter the position of the constraint. This means that the calculated probability constantly changes as events occur. This is the error rate, or more precisely, the rate of change, assuming that the errors are here to stay and nothing else will influence them.

As shown in (46), two constraints that are relatively close slightly overlap when added a range.

(46) Free ranking (repeated from (15))



The probability of occurrence of each of the constraints is interpreted as a normal (Gaussian) distribution, in which the constraint is more likely to apply around the centre of the distribution and less likely to apply towards the edges.

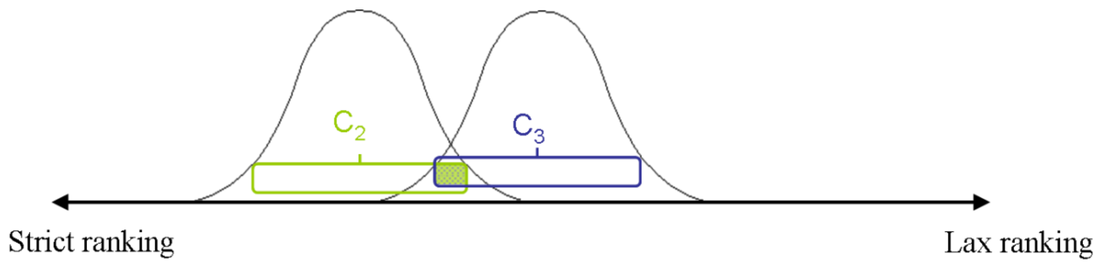


Figure 1: Constraints' Distribution of Application

The overlapping area between the two distributions is the probability of occurrence of both rankings ($C_2 \gg C_3$ and $C_3 \gg C_2$).

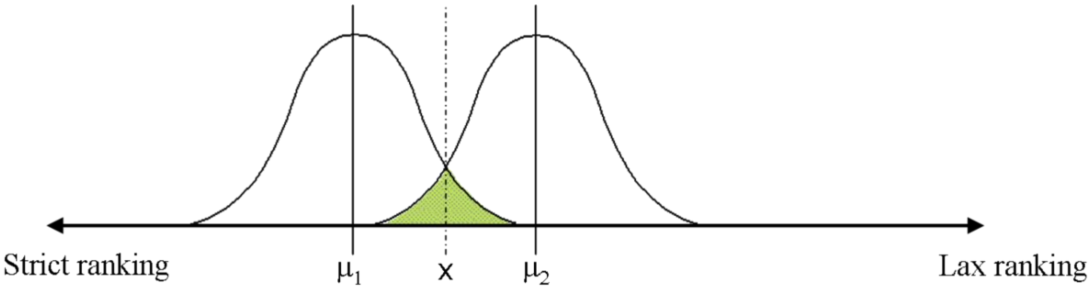


Figure 2: Overlapping Constraints' Distributions

Half of this area is the probability of occurrence of the ranking C_3 » C_2. To calculate this probability, we calculate the area above x (the point where the two distributions cross). The following formula normalises the distribution in order to calculate the distance (in standard deviations) between μ_1 and x and also between μ_2 and x.

(47) Normalisation formula for calculating the distance from the mean (in standard deviations)

$Z = \frac{x-\mu}{\sigma}$

Where

x is the meeting point of the two distributions

μ is the mean of each distribution

σ is the standard deviation which is interpreted as the evaluation noise (and is an arbitrarily selected value)

A Z-table (Appendix C) is then used to calculate the two areas. The result of the subtraction is the area between the two points. In this study, this area is interpreted as the frequency of error, or the probability of occurrence of the opposite ranking.

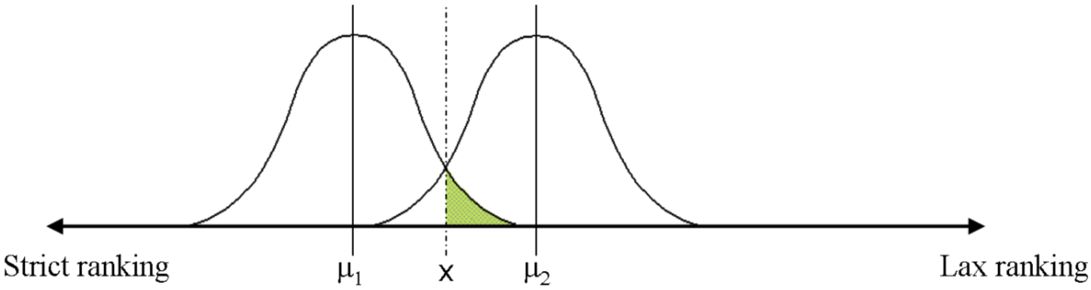


Figure 3: Frequency of Error

3.4.2. *i ~ e Alternation in B3*

In this section, I apply the principles of the GLA described above to the *i ~ e* alternation in B3. First, I present the alternation in the prescriptive (normative) language, and take it as a starting point for a hypothetical phase in the language where presumably there was no variation (§3.4.2.1). There may not have been such a phase in the language's history. However, given the principles of change (that it is from a non-variable grammar to a different non variable grammar), it stands to reason that there could have been a stage in the language in which there was alternation but no variation. Then, I describe the current state, in which the alternation is preserved, but with variation in the output forms (§3.4.2.2). Finally, I apply the GLA to simulate the change in the *i* and *e* distribution in B3 in an attempt to predict the future non variable state (§3.4.3).

3.4.2.1 In a non-variable phase of the language, in B3 (*hifíl*) past tense, some of the verbs have an *i* in the initial syllable, and some have *e*. The distribution of *i* and *e* at this historical stage, represented here by the normative prescriptive language, is complementary. *i* surfaces in an initial closed syllable (48a and 48b) and *e* in an initial open syllable (48c).[30]

(48) *i* and *e* distribution (B3)

a. Before a cluster
higdíl 'enlarged'
hixtív 'dictated'

b. Before a geminate (historical)
hikkír 'recognised'
hibbít 'looked at'

c. Before a CV sequence
hekím 'established'
hevín 'understood'

30 The analysis provided in this section refers to a hypothetical historical stage in which gutturals and geminates surfaced regularly and accounted for the distribution of the vowels. Synchronically, as gutturals and germination no longer exist, a different account is proposed further into this study.

In regular verbs, where all stem consonants surface, *i* is observed in the first syllable (48a). When the initial stem consonant is a geminate, *i* also surfaces in the initial vocalic position (48b).[31] In the weak verbs where there are only two stem consonants underlyingly, no compensatory gemination occurs, and the initial vowel is in an open syllable, and is lowered to *e*.

The constraints responsible for this alternation are as follows:

(49) *i* ~ *e* Alternation Constraints

a.	$*V_{[+high]}]_{\sigma}$:	High vowels are prohibited in open syllables
b.	IDENTV:	Every vowel in the output must be identical to the corresponding vowel in the input

A high vowel is prohibited in an open syllable (49a). When in this position, lowering occurs to satisfy this constraint. The satisfaction of this constraint is at the expense of IDENTV (49b), which requires identical vowels in the output as in the input. For lowering to occur, the ban on high vowels in open syllables must outrank IDENTV.

$*V_{[+high]}]_{\sigma}$ is a member of a family of constraints banning high vowels in general, and in every position in particular. This ban is reiterated for every vowel in the language's vocalic inventory. The ban against high vowels in general, though, must be ranked quite low, at least lower than IDENTV (as shown in 50), and does not affect the discussed verbs.

(50) $*V_{[+high]}]_{\sigma}$ » IDENTV » $*V_{[+high]}$

When the initial stem consonant is a guttural, a vowel is inserted to rescue the guttural from a coda position. Now that the guttural is in the onset, lowering occurs

[31] In this group of verbs, known as the defective verbs, the initial underlying stem consonant is deleted and gemination occured historically to compensate for the lost segment (Lowenstamm and Kaye 1986). Synchronically, germination no longer exists and so speakers have no reason to assume an underlying '*n*'.

because the initial stem vowel is in an open syllable.[32] The process is summarised in (51).

(51) hiħlíf (cf. higdíl 'he enlarged') → hiħ**e**líf → h**e**ħelíf 'he replaced'

(52) B3 verbs with an initial guttural

a.	heħelíf	'replaced'
	heħemí	'complimented'
b.	heʔemín	'believed'
	heʔeʃím	'accused'
c.	heʕesik	'employed'
	heʕerix	'appreciated'

Epenthesis in this environment does not always occur. There are 14 verbs where the guttural does occur in the coda, as in the examples in (53).

(53) heħtím 'sign (someone)'
heʕtík 'copied'

The verbs in (53) have a guttural in the initial stem position, and yet vowel epenthesis does not occur and the guttural surfaces in the coda. And although the first syllable is closed, *i* does not surface and vowel lowering does occur. An additional constraint (54) must be postulated to account for this.

(54) $*V_{[+high]}C_{[+low]}$: A sequence comprising a high vowel followed by a low consonant is not allowed

The constraint $*V_{[+high]}C_{[+low]}$ must outrank IDENTV to enable lowering. The low consonant can be deleted to satisfy this constraint, but deletion is banned by MAX, requiring all input segments to surface in the output. Indeed, additional constraints are necessary to ban lowering to *o* or to *a* instead of to *e*. But I disregard them as they are not pertinent to the discussion at hand.

32 The lowering of the vowel could be due to vowel harmony. This is immaterial to the discussion.

The two markedness constraints, $*V_{[+high]}]_\sigma$ and $*V_{[+high]}C_{[+low]}$ are responsible for the distribution of *i* and *e* in the initial vocalic position of the B3 verbs. The two constraints do not compete with one another, but they both outrank IDENTV at this stage of the language. The ranking is provided in (55).

(55) *i* ~ *e* alternation constraints ranking

$*V_{[+high]}]_\sigma$, $*V_{[+high]}C_{[+low]}$ » IDENTV » $*V_{[+high]}]$

The following tableaux illustrate this for all input types:

(56) *i* ~ *e* alternation (B3)

i.

/higdil/	$*V_{[+high]}]_\sigma$	$*V_{[+high]}C_{[+low]}$	IDENTV	$*V_{[+high]}]$
a. + hig.dil				**
b. heg.dil			*	*

ii.

/hiħtim/	$*V_{[+high]}]_\sigma$	$*V_{[+high]}C_{[+low]}$	IDENTV	$*V_{[+high]}]$
a. hiħ.tim		*!		**
b. + heħ.tim			*	*

iii.

/hikim/	$*V_{[+high]}]_\sigma$	$*V_{[+high]}C_{[+low]}$	IDENTV	$*V_{[+high]}]$
a. hi.kim	*!			**
b. + he.kim			*	*

iv.

/hinkir/	$*V_{[+high]}]_\sigma$	$*V_{[+high]}C_{[+low]}$	IDENTV	$*V_{[+high]}]$
a. + hikkir				**
b. hekkir			*!	*

The tableaux in (56) demonstrate the distribution of *i* and *e* in the deterministic grammar that is assumed to exist, or could have existed, in a previous phase of the language. In the regular verbs in (56i), where all stem consonants surface and the initial stem consonant is not a guttural, the *i* of the binyan's vocalic pattern surfaces unchanged in the first syllable, as is mandated by IDENTV (see §2.1). When the initial stem consonant is, however, a guttural (56ii), lowering occurs in order to satisfy the higher ranking constraint $*V_{[+high]}C_{[+low]}$. The faithful candidate (56iia) is eliminated for violating this constraint. When only two stem consonants exist underlyingly, the output inevitably consists of an open syllable, and so *e* surfaces in order to satisfy the high ranked markedness constraint banning a high vowel in an open syllable, $*V_{[+high]}]_\sigma$ (56iiib). In the last group of verbs, known as the defective verbs (56iv), the underlying consonant is deleted and although a stem consonant is missing, the underlying vowel *i* is rescued by the compensatory gemination (Lowenstamm and Kaye, 1986).

There is a relatively small group of weak verbs, where *o* surfaces in the initial vocalic position, replacing an initial *y*, which is found in other morphologically related verbs in a different binyan (e.g. *horíd* 'lower'; cf. *yarád* 'descended').

The following table illustrates the distribution of B3 verbs at this phase of the language with their type frequency.

(57) *i* ~ *e* alternation in B3 past tense (Normative)

	Paradigm	initial *i*	% (types)	initial *e*	%	initial *o*	%
a.	Regular	higdíl 'enlarged'	95.67% (287)	heħʃíx 'darkened'	4.33% (13)	--	
b.	Weak	hikkír 'recognised'	34.21% (78)	hevín 'understood'	55.26% (126)	horíd 'lowered'	10.53% (24)

At first glance, it appears that the regular (57a) and the weak paradigms (57b) are alike, as both include forms with initial *i* as well as forms with initial *e*. However, observe the percentages of each form. This is the percentage out of the total number of B3 verbs per paradigm type (i.e. out of the total number of regular/weak B3 verbs). The regular paradigms predominantly have an *i* in the initial vocalic position, whereas the weak paradigms predominantly, though less prominently, have an *e* in this position. It would appear, then, that speakers are more likely to select *i* in the initial position of a regular verb than an *e*, and when a consonant is missing in the paradigm (i.e. in weak verbs), speakers are more likely to select an *e* in the initial position, though to a lower probability than in the regular paradigms.

The following tables show the type distribution of *i*, *e*, and *o* in B3 verbs relative to the total number of B3 verbs with *i/e/o* in initial position (**%R**), and relative to the total number of B3 verbs (**%T**). The table in (58) refers to all B3 verbs in Hebrew (including rare and obsolete verbs). The table in (59) shows the same information but from a database of actual verbs used in MH (taken from Bolozky's list of 500 most frequent verbs).

(58) B3 verbs type frequency (counted from Tarmon and Uval (1998))

	i	%R	%T	*e*	%R	%T	*o*	%R	%T	Total	%
Regular	287	78.63	54.36	13	9.35	2.46	0	0	0	300	56.82
Weak	78	21.37	14.78	126	90.65	23.86	24	100	4.54	228	43.18
Total	365		69.13	139		26.33	24		4.54	528	

(59) B3 verbs in use type frequency (counted from Bolozky 1996)

	i	%R	%T	*e*	%R	%T	*o*	%R	%T	Total	%
Regular	44	70.97	41.51	12	33.33	11.32	0	0	0	56	52.83
Weak	18	29.03	16.98	24	66.67	22.64	8	100	7.55	50	47.17
Total	62		58.49	36		33.96	8		7.55	106	

The tendencies in (58) and (59) are similar. The majority of verbs have an *i* in the first syllable (69.13% in (58) and 58.49 in (59)), compared to *e* (26.33% in (58) and 33.96% in (59)). In the regular verbs, the larger group consists of verbs with initial *i* (95.67%, 287 out of the 300 regular verbs), whereas in the weak verbs, the larger group consists of verbs with an initial *e* (55.26%, 126 out of the 228 weak verbs). Note that out of the 528 B3 verbs in the dictionary, only 106 are in actual use. The rest have either become obsolete, or are extremely rare. But also in the verbs in use (59), the larger group of verbs have an *i* in the first syllable of regular verbs (78.57%, 44 out of the 56 regular verbs) and an *e* in the first syllable of the weak verbs (48%, 24 out of the 50 weak verbs).

The following table shows the token distribution of the B3 verbs in Bolozky's list.

(60) B3 verbs in use (Bolozky) token frequency[33]

	i	%R	%T		e	%R	%T		Total	%
Regular	18,378	57.02	31.31		3,192	17.02	5.4		21,570	36.75
Weak	13,854	42.98	23.61		15,564	82.98	26.52		37,118	63.25
Total	32,232		54.92		18,756		31.96		58,688	

The token frequency of the verbs in use in Bolozky's corpus reveals the same tendencies as the type frequency, namely that there is a preference of verbs with *i* in initial position over verbs with *e* in the same position. Within these two groups, though, the weak verbs receive a much higher token frequency than type frequency. This is not surprising, as the weak verbs are the most commonly used verbs. In fact the ten most frequent verbs on Bolozky's list are all weak. The data from the recorded corpus are presented in §3.4.2.2.

In the following section, I present the variation in the B3 verbs as is manifested in the recorded database and I apply the principles of the Gradual Learning Algorithm on these data to determine the stochastic relations among the constraints that is responsible for the variation.

3.4.2.2 Current State

In the current stage of the language, variation is observed and the initial *i* and *e* in B3 occur in free variation. Both the normative and the colloquial forms can be found within the same register. Consonant gemination has long been lost and the *i* in the resulting open syllable has become opaque.

(61) *i* and *e* variation in MH (B3)

Type		**Normative**	**Colloquial**	
Regular	a.	higdíl	hegdíl	'enlarged'
	b.	hextím	hixtím	'signed someone'
Weak	c.	hekím	hikím	'established'
	d.	hikír	hekír	'recognised'

33 The 7,700 B3 verb tokens with initial *o* have been removed due to space considerations. They are reflected in the calculations, though.

The merging of the guttural *ħ* with the velar *x*, caused the verbs with the historical guttural (61b) to resemble the regular verbs with no guttural (61a). The loss of gemination in the language has similarly caused the defective verbs with historical gemination (61d) to resemble the weak verbs with no gemination (61c). This explains why /*hixtím*/ surfaces as [*hextím*], as in the previous stage, alongside [*hixtím*], and [*hekír*] alongside [*hikír*].

The historical guttural *ʕ* has merged with *ʔ*, and the *ʔ* has been steadily disappearing in natural speech. So now that the gutturals have disappeared from the language's segmental inventory, the constraint militating against a sequence of a high vowel followed by a guttural no longer plays a role in the language. To account for the co-occurrence of *hegdíl* and *higdíl*, the general constraint against high vowels discussed in (50) above, must now outrank IDENTV. And yet for verbs like *hikím* to surface, IDENTV must outrank $*V_{[+high]}$. Both rankings are simultaneously possible.

That both rankings are simultaneously possible suggests that the ranking is not fixed. Traditional OT accounts for variation through the crucial unranking of constraints (Anttila, 1997a, Kiparsky, 1993, Prince and Smolensky, 1993/2004, Reynolds, 1994). Thus, unlike regular unranking, where two constraints are not in competition, and so whether one is ranked above or below the other yields the same result, crucial unranking means that both rankings are possible, and they each yield a different output. This provides a theoretical model that caters for variation, but it says nothing about how often each ranking is predicted to apply (see §1.3.1).

While the variant forms in (61) are considered to occur in free variation, they do not all occur in the same frequency. The ranking applies stochastically, mirroring the variants' frequency in the language. This means that for the variation in (61) to occur, the ranges of the constraints must overlap to a varying degree. In traditional OT, the location of the constraints on the scale is not important. However, in Stochastic OT, the distance between the constraints determines the frequency in which each variant is likely to occur.

The frequency of occurrence of *i* and *e* in B3 is as follows (DB=database):

(62) *i* and *e* token frequency

Vowel	**Token Frequency**	
	Recorded DB	**Bolozky's DB (see 60)**
i	64.49% (189)	54.92% (32,232)
e	33.88% (104)	31.96% (18,756)

Information about the variation is included in the recorded database (but not in Bolozky's database). The frequency of occurrence of the new variants (i.e. the 'error' rate) is important to determine the rate of change. The calculated error rates in the recorded database are as follows:

(63) Error rates in the distribution of *i* and *e* in B3

i $\rightarrow$ e: 12.62% 25/198
e $\rightarrow$ i: 15.79% 15/95
Total error rate: 13.65%

Out of the verb tokens with normative initial *i* in the database (189), 12.62% surfaced with an *e* instead, and out of the verbs with initial *e* in the database (95), 15.79% surfaced with an *i* instead, a total error rate of 13.65%.

The actual value of μ on the constraints' scale (which, recall, is the mean of the distribution of each constraint) is not important; only the distances between the constraints are. The initial value is picked randomly. Let us assume an initial value of 100 for the mean of IDENTV. Applying the calculations described in §3.4.1 with a standard deviation of 2.0 and a varying plasticity of 1 at the onset of learning and of 0.1 at the end of the learning cycle, the mean of the overlapping constraint $*V_{[+high]}]_\sigma$ is located at a distance of 2.2 deviation points. Each time *e* is encountered, the two constraints move closer together; whenever *i* is encountered, the two constraints move farther apart. If they move closer together, the rate of the variation will grow presumably until they grow apart again and the opposite ranking wins over. For the constraints to be far enough apart so that no overlapping occurs (thus one constraint is crucially ranked above the other and variation is presumed impossible), the distance between the means of the two constraints would have to be close to 5 deviation points.[34]

It could be argued that the difference between *hikír* and *hekír* lies not in the ranking of the constraints, but in the underlying form. That is, speakers who say *hikír* are aware of the missing consonant (i.e. their UR is something like /*hiCkir*/, and for them the initial vowel is in a closed syllable, even though the phonology of the language no longer enables long consonants. But if so, then intra-speaker variation would not be possible.[35]

It could also be argued that the change from *i* to *e* is simply a reduction of the vowel in an unstressed syllable. However, no such reduction occurs in B2 or B4, where the *i* also occurs in an open syllable.

(64) No *i* and *e* variation (B2 and B4)

	Binyan	Past		
a.	B2	nitán	*netán	'he was given'
b.	B4	dibér	*debér	'he talked'

That this happens only in B3 and that there is change in the opposite direction as well, namely from *e* to *i* (e.g. *hevíu* → *hivíu* 'they brought') suggests that this is not reduction at all, but rather a ranking issue.

In the normative language, there is a correlation between the past tense and the present tense, such that if there is an *i* in the past tense, there's an *a* in the present tense (e.g. *hikír–makír* 'recognise' past–present) and if there is an *e* instead, there is also an *e* in the present tense (e.g. *hekím–mekím* 'establish' past–present). The nature of the collection of the data for this study does not enable to check whether or not this

34 Note that even when constraint ranges do not overlap, the distributions of their application do overlap. But if the constraints are far enough apart, the overlap bordering infinity may not ever surface.

35 As mentioned in §1.4, the collection of the data for this study does not accommodate an analysis of intra-speaker variation, as there is not enough data from any one speaker. However, the database does show some instances of intra-speaker variation, particularly relating to the *hikír* paradigm.

correlation is retained in the midst of change, that is, if speakers who say *hivéti* ‘I brought’ also typically say *maví* in the present tense, as there is not enough data from any one speaker. To check this, much more data is required from each speaker.

3.4.3. Next Generations

The GLA is an algorithm for learning the ranking values of constraints in a stochastic OT grammar. It simulates the process of grammar acquisition for a language learner exposed to variable surface data. So the resulting ranking values reflect the current state, where the distribution of *higdíl* ≈ *hegdíl* and that of *hikím* ≈ *hekím* is approximately 60-40 percent. Running the algorithm, then, reflects the learning process of a single generation of learners. Each generation of learners is assumed to have the grammar of the previous generation at the onset of learning. This means that by applying the algorithm to additional cycles of learning, where the results of each cycle feeds the next cycle, we can make predictions about language change.

The size of the corpus in each cycle is assumed to be identical to that of the previous cycle. The frequency of the inputs is also assumed to be identical in each generation. The only parameter assumed to be varying from generation to generation is the frequency of the outputs. As these frequencies change, so does the distance between the constraints until finally they no longer overlap.

In Chapter 5, I predict the following (assuming nothing happens to obstruct the course of change):

a. For the unidirectional type of change, where the current ranges of the competing pair of constraints are close with some degree of overlap, the constraints will grow apart such that only one of the variants will survive (typically the new variant).

b. For the bidirectional type of change, the same change path is predicted as for the unidirectional change, but it will apply to both directions. This means that the classes of both variants will continue to exist. The variation is therefore predicted to continue incessantly, as long as the comparable paradigms fulfil the similarity requirement.

CHAPTER 4. SIMILARITY AND PARADIGM LEVELLING

Paradigm levelling refers to a diachronic linguistic phenomenon where a property that is found in some forms within a paradigm extends to other forms within that paradigm for the sole purpose of increasing the similarity among the members of the paradigm. The paradigm is levelled to a distinctive property found in one or more of its members. This phenomenon reflects a general preference of language for non-alternating paradigms.

There are several examples of paradigm levelling in recent literature, from German (Albright, 2008a), Yiddish (Albright, 2002a, Albright, 2006b), Latin (Kenstowicz, 1996), Hungarian (Rebrus and Törkenczy, 2005), Hebrew (Bat-El, 2005), among others. In Hebrew, the imperatives in (65) below are formed by truncating the 2nd person future form (the base). The normative imperative forms follow the old system, where complex onsets are banned from surfacing and spirantisation occurs post-vocalically. The result is an alternating paradigm, as in *ti**f**táx–**p**táx* ‘open! ms.sg.’ and *tift**e**xí–p**i**txí* ‘open! fem.sg.’. The colloquial imperatives, on the other hand, differ from their future base only in the truncated segments, but are otherwise identical to the base, resulting in a non-alternating future–imperative paradigm, as in ***ti**ftáx–ftáx* ‘open! ms.sg.’ and ***ti**ftexí–ftexí* ‘open! fem.sg.’ (see Bat-El (2005) for a complete analysis of Hebrew imperatives and Adam (2002) on MH spirantisation).

(65) Modern Hebrew truncated imperatives (Bat-El, 2005)

Base (Future)	**Output (Imperative)**		
	Normative	**Colloquial**	
tiftáx	ptáx	ftáx	‘open ms.sg.’
tiftexí	pitxí	ftexí	‘open fem.sg’
titlabéʃ	hitlabéʃ	tlabéʃ	‘dress ms.sg.’
titlabʃí	hitlabʃí	tlabʃí	‘dress fem.sg.’

Note that the stop–fricative alternation is not completely eliminated from the verb paradigm; it is maintained between past and future forms both normatively and colloquially (e.g. ***p**atáxta–ti**f**táx* ‘open 2nd pr. ms.sg. past–future).

The colloquial imperative forms are levelled to their corresponding future forms even though the verb system does not otherwise permit complex onsets.[36] That is, paradigm levelling eliminates paradigmatic contrasts that would have otherwise existed due to the high ranking of some constraint observed elsewhere in the same morpho-phonological environment, even if at the expense of violating otherwise unviolated constraints.

In pre-generative linguistics, such exceptions to ‘sound laws’ (‘constraints’ in today’s terms) were expressed in terms of analogy. The analogy to the surface future form has caused the levelling in the future–imperative paradigm. Within the framework of Optimality Theory, such exceptions to sound laws are expressed through correspondence between output forms (Benua, 1997, Kenstowicz, 1996, McCarthy and Prince, 1995).

The OT model is reminiscent of the analogy model of pre-generative linguistics. Both models essentially compare output forms. However, analogy, as observed in the Hebrew verb system, is between forms that do not share a common stem, whereas paradigm levelling is among the members of a paradigm that do share a common stem. In what follows (§4.1), I further discuss analogy in the pre-generative days (§4.1.1) as well as in OT (§4.1.2). ***I argue that analogy that is manifested in merging (see §3.3.1 and §3.3.2), relies heavily on similarity***. I provide a model of similarity (§4.3) that quantifies the similarity between words and between entire paradigms, enabling to predict which paradigms are likely to merge and which are not.

---

36 The ban on complex onsets is violated in denominative verbs (Bat-El 1994).

4.1. Analogy

The term 'analogy' has different meanings in different disciplines. Mathematics, natural history, philosophy, and literature are only some of the disciplines that employ this term. In language, 'analogy is a relation of similarity' (Anttila, 1972:88).

Ancient Western grammarians classified verbs and nouns according to the similarities and differences in their inflections. The regularity in the inflection, which was manifested by similarity among the forms, was interpreted as analogy; exceptions, which were manifested by differences among the forms, were explained by rules. Indeed, the very definition of 'analogy' (from Greek '*analogia*') is 'similarity'.

Contra their predecessors, in the late nineteenth century, the neogrammarians used analogy to address exceptions to regular sound laws. They viewed new analogical formations as 'incorrect' forms that replace existing forms and not as results of sound change (Lahiri, 2000 and references therein). For example, in Old Icelandic, the NOM sg. – GEN sg. paradigm **mann-r–mann-a* 'man' surfaced at a later stage as *maðr–manna*. The change **mannr* → *maðr* resulted from analogy to the NOM sg. – GEN sg. paradigm *guðr–gunna* 'battle', and cannot be attributed to any phonological change deriving -*ð*- from -*nn*- (Reiss, 1997).

4.1.1. Analogy in the Early Days

The Neogrammarians used the four-part proportional analogy synonymously with paradigm levelling (Anderson, 1985, Antilla, 1977, Downing et al., 2005). Saussure (1916, 1959) cites the change from *honōs* to *honor* 'dignity' in Latin as in (66) below (where ≅ denotes here analogy):

(66) Proportional analogy (Latin rhotacism)

ōrātōrem : ōrātor ≅ honōrem : x ; x = honor (replacing *honōs*)

'speaker-ACC sg.' : 'speaker-NOM sg.' ≅ 'dignity-ACC sg.' : 'dignity-NOM sg.'

If *ōrātōrem* and *honōrem* are analogous (as they are both accusative forms with the same suffix *-em*) then so are their nominative counterparts *ōrātor* and *honōs*. Consequently, *honōs* changes to *honor*.

However, proportional analogy is not sufficient to explain the change from *honōs* to *honor*. Without any guiding principles, there is nothing to prevent change in the opposite direction, i.e. taking *honōrem* : *honōs* as the model and changing *ōrātor* to **ōrātōs* instead:

(67) Unattested proportional analogy (Latin)
honōrem : honōs ≅ ōrātōrem : x
x= ōrātōs

This problem has not gone unnoticed. Sapir (1921) proposed that the forms most represented in the paradigm caused the others to change, such that change was to the property with the highest frequency in the paradigm. Saussure (1916, 1959) disagreed with this claim, saying that analogy is unpredictable and that the most frequent forms do not necessarily serve as the base of the analogical formations. Moreover, he claims that analogical forms are new creations that exist alongside the traditional forms whereas phonetic change annuls whatever preceded it and so analogical phenomena do not entail language change at all (p. 162-163). Others have suggested guidelines that regulate the direction of analogical change (Kuryłowicz, 1949, Mańczak, 1958). Kuryłowicz suggested that certain grammatical contrasts are more important than others and thus preserved in analogy (see further discussion in Anderson, 1992). This is shown in (68), where the preservation of number contrast (sg. vs. pl.) is argued to be more important than the preservation of case contrast (NOM vs. ACC):

(68) Change from Latin to Iberian Romance 'bread' (Kuryłowicz, 1949)[37]

	Latin		→	**Iberian Romance I**		→	**Iberian Romance II**	
	Sg.	**Pl.**		**Sg.**	**Pl.**		**Sg.**	**Pl.**
NOM	panis	panēs		panes	panes		pane	panes
ACC	panem	panēs		pane	panes		pane	panes

(69) Unattested change from Latin to Iberian Romance

	Latin		→	**Iberian Romance I**		→	**Iberian Romance II**	
	Sg.	**Pl.**		**Sg.**	**Pl.**		**Sg.**	**Pl.**
NOM	panis	panēs		panes	panes		pane	panes
ACC	panem	panēs		pane	panes		pane	pane

As a result of the sound change merging the front vowels *i* and *e* to *e* (i.e. from Latin NOM sg. *pan**i**s* to Iberian Romance NOM sg. *pan**e**s*) and the loss of final nasals (i.e. from Latin ACC sg. *panem* to Iberian Romance ACC sg. *pane*), the number distinction in the nominative was lost (in Latin, *panis* sg. and *panes* pl., whereas in Iberian Romance the same form *panes* is used for both sg. and pl.). The analogical change (*panes* 'ACC. pl.': *panes* 'NOM. pl.' ≅ *pane* 'ACC. sg.: *x* 'NOM. sg.', *x*=*pane*) restored the number contrast at the expense of the case contrast. That is, in both nominative and accusative, the distinction between singular and plural is maintained after the change, but the distinction between the two cases is completely lost. Kuryłowicz concludes that some distinctions, such as number, are more important than others (e.g. case). If case were more important than number, then the plural accusative *panes* would have possibly changed to *pane*, thus maintaining the case distinction (*panes* NOM, and *pane* ACC) while losing the number distinction.

4.1.2. Analogy in OT

The research on analogy continues along the lines of markedness. The notion of markedness was introduced and developed by Trubetzkoy (1939) within the Prague

[37] In Latin, case contrast is partial while number contrast is maintained throughout the paradigm. Therefore, it could be argued that the change from Latin to Iberian Romance maintains the widespread paradigm.

circle and was often related to the level of complexity (Anderson, 1985). The question as to what constitutes complexity is tricky in itself and has not yet been resolved, as it not only alludes to overt marking (e.g. affixation), but also to covert information (e.g. syllable structure, weight, etc.). Kiparsky (1968, Kiparsky, 1982, Kiparsky, 1988, Kiparsky, 2000) motivates analogy by grammar simplification, i.e. change that is constrained by the phonological system. He proposes that analogical changes are driven by well-formedness constraints, and that it is, therefore, grammar optimization. This view that 'language change' is essentially 'grammar change' completely eliminates the distinction between sound change and analogy. Analogy, in this respect, is another form of grammar change.

The properties of OT, that it is non-derivational, surface-based, and that it relies on faithfulness constraints as well as markedness constraints (Prince and Smolensky, 1993/2004), have made possible to incorporate analogy into mainstream theory. A purely analogical OT model would rely solely on surface faithfulness constraints. The only difference that requires bridging is that analogy tends to operate on words whereas constraints are more general.

Attempts to include paradigms into formal OT started with Kenstowicz (1996), Benua (1997), McCarthy (2005). According to Benua's (1997) Transderivational Correspondence Theory, a derived word is affected by its morphologically related simplex base. In the example of Hebrew imperatives ((65) above), the imperative form must be derived from the future base because the initial fricative in *ftáx* 'open! ms.sg.' cannot be drawn from the phonology of the language: it is not phonemic (this is exhibited in the stop–fricative alternation in *patáx–yiftáx* 'open 3rd ms.sg. past–future') and it is also not post-vocalic (the environment required for spirantisation to apply). The initial fricative must correspond to the surface future base (Bat-El, 2005, Bolozky, 1979). Accordingly, it is always the derived form that stands in correspondence with the base, and the influence is always of the base on the derived form. That is, within Benua's model, as in Kuryłowicz's model, levelling is asymmetrical, as it is always unidirectional; the derived form can never influence the base.

Kenstowicz (1996) proposes a symmetric account of levelling. His model of Uniform Exponence requires that morphologically related lexical items be as similar as possible. For example, recall the Latin analogical change in (66) above. The triggering factor for this change was the rhotacism rule replacing an intervocalic underlying */s/* with a surface [r]:

(70) Latin rhotacism (Kenstowicz, 1996)

NOM sg.	**GEN sg.**				
honōs	honōr-is	'honour'	(cf.	hones-tus	'honest')
arbōs	arbōr-is	'tree'	(cf.	arbus-tus	'wooden')

In the Latin examples above, stem final *s*, observed in the NOM sg., is replaced by an *r* when followed by the vowel-initial suffix of the genitive and thus between vowels. When a consonant initial suffix follows (e.g. *hones-tus*), the *s* is not in between vowels and so rhotacism is blocked.

At a later stage of the language, rhotacism has extended its application to the nominative forms as well, even though the *s* is not intervocalic (vowel length distinction is ignored as it is not pertinent to the discussion).

(71) Latin rhotacism (later stage)

NOM sg.	**GEN sg.**	
honor	honōr-is	'honour'
arbor	arbōr-is	'tree'

Kenstowicz explains the over-application of rhotacism by a constraint of Uniform Exponence (UE), requiring the stem to receive a consistent realisation throughout the paradigm. The UE constraint is ranked higher than the faithfulness constraint requiring the identify of *s* between input and output (FAITH-*s*). The constraint banning inter-vocalic *s* is similarly ranked above FAITH-*s*, resulting in the ranking: UE, *VsV

» FAITH-s. In this model, all related words are simultaneously evaluated by the grammar, selecting the optimal paradigm, as in (72).

(72) UE analysis of Latin rhotacism (Kenstowicz, 1996)

/hon:s, hono:s-is, hono:s-em,…/	UE	*VsV	FAITH-*s*
a. + honor honōris honōrem			***
b. honos honōris honōrem	*!		**
c. honos honōsis honōsem		*!*	

By virtue of UE, candidate (a) is the optimal paradigm, as all its members share the same stem final consonant. Candidate (c) is ruled out for the multiple violations of the constraint against inter-vocalic *[s]*, even though it not only conforms to UE, but it is also the most faithful candidate, as it maintains the underlying */s/* throughout the paradigm. Candidate (b) has a mixed paradigm; it conforms to the rhotacism rule by replacing the underlying /s/ only where it is intervocalic and allowing it to surface when it is not. However, as this does not fulfil the UE requirement, this candidate is also discarded, leaving candidate (a) as the winning paradigm, even though it is the most unfaithful.

Note that all members of the paradigm are evaluated together for any violation of the constraints. As there is no base, all members have equal status, and the model is therefore symmetric; any member could in principle affect all other members of the paradigm.

McCarthy (2005) explains that because of the asymmetrical nature of Benua's TCT model, which relies on the priority of a base, it holds for derivational morphology, but is incompatible with inflectional morphology. This is because inflectional paradigms have no base (cf. Albright, 2008b). Conversely, Kenstowicz's

Uniform Exponence fares well with inflectional paradigms, but runs into difficulty with derivations, as it over-predicts surface similarity. Therefore, McCarthy (2005) proposes the Optimal Paradigms model, which applies correspondence as used in TCT, and combines it with the symmetric evaluation of paradigms from Uniform Exponence.

In the Optimal Paradigms model, candidates consist of entire inflectional paradigms. The candidate paradigms are evaluated against markedness constraints, against faithfulness constraints requiring identity between each member of the paradigm and the underlying input (IO constraints), and between each member of the paradigm (OP constraints). The winning paradigm is the one that best satisfies the constraint hierarchy. McCarthy improves on Benua's model by enabling to simultaneously evaluate paradigms that do not share the same underlying stem.

4.2. Similarity

As mentioned in the previous section, the term 'analogy' means 'similarity'. So analogical change is connected in some way with similarity. Paradigm levelling is a synchronic requirement of the grammar that members of a paradigm be identical. So although 'analogy' and 'paradigm levelling' are seemingly synonymous, they are, nevertheless, different. Paradigm levelling is a type of analogy. While analogy is a diachronic process, paradigm levelling is the result of a synchronic requirement of the grammar, the effect of a constraint in OT terms. Another difference between analogy and paradigm levelling is that while paradigm levelling explains why members of a paradigm are similar, analogy has no explanatory power.

Similarity is the basis of analogy and the result of levelling. When speakers do not know how to conjugate a problematic or unfamiliar verb, they compare it to similar verbs that they do know and conjugate accordingly. Additionally, the grouping of words into paradigms is based primarily on similarity. The more similar two forms are, the more likely they are to be members of the same paradigm.

The term ‘paradigm’ alludes to the morphological relation among words. It has long been recognised that morphologically related words can influence each other’s surface structure. However, belonging to the same paradigm does not necessarily, in itself, say anything about the closeness of the relation. Nevertheless, the closer the relation, the greater the likelihood of mutual influence.

Paradigms are organised in hierarchies, such that every paradigm is part of a larger paradigm, as in Figure 4 below. That is, every paradigm is a sub-paradigm of a larger (sub-)paradigm (van Marle, 1985, Wurzel, 1989). Within this hierarchical organisation, there is a major distinction between inflectional and derivational paradigms. Forms in inflectional paradigms, as defined in Anderson (1992), are distinguished on the basis of syntactically-relevant features (e.g. tense, gender, number, etc.). The distinction among forms in derivational paradigms is less uniformed; it can be the different realisations of the same base in different derivational categories. Figure 4 shows a schematic hierarchy of the Hebrew verb system.

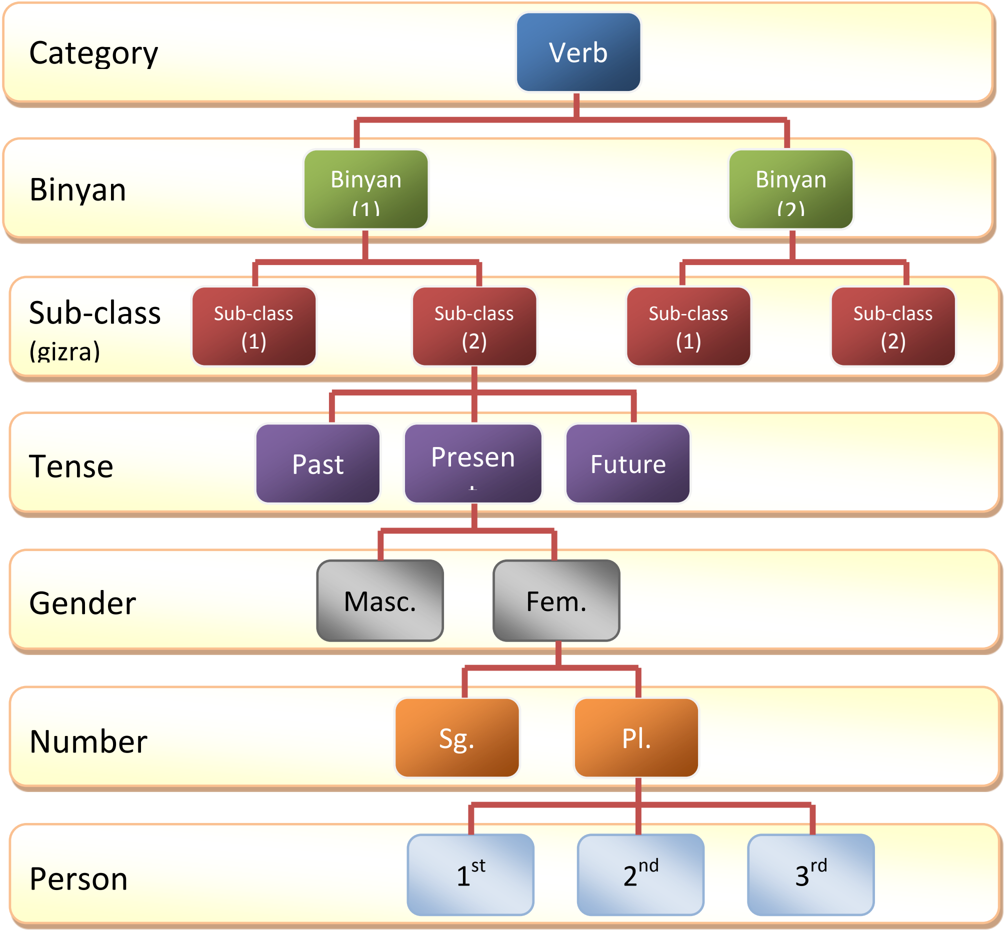


Figure 4: Hebrew Verb System Paradigm Hierarchy

The verb category comprises the various configurations, named *binyanim* (see §2.1). There are five verb configurations in Hebrew, which are derivationally related (e.g. B1 *gadál* 'to grow', B4 *gidél* 'to raise', an B3 *higdíl* 'to cause to grow').[38]

Within each *binyan* there are different sub-classes, named *gzarot* (singular *gizra*), distinguished by the prosodic and vocalic alternations within the inflectional paradigm (tense, gender, number, and person). These sub-classes distinguish between the regular 'strong' verbs and the various types of irregular 'weak' verbs, where the classification of the irregularity depends on the locus of weakness.

The sub-classes are merely the grouping of the verbs within the binyan according to characteristics of their stem consonants and prosodic structure. There is no relation

[38] Hebrew verbs are also derivationally related to nouns (e.g. *gdilá* 'growth').

among the various sub-classes; there is no derivation and no inflection. ***The sub-class is the level that is changing and where most of the variation resides. Changes occurring at this level will not affect either derivation or inflection.*** It is the degree of similarity among the various sub-classes that will determine the extent of the change.

Similarity among word forms is expected to be greater within the specific sub-paradigm. For example, the future masculine plural paradigm of a specific binyan (that is, at the level of number), consists of an identical stem with different person prefixes, as shown for different verb types in (73) for B1 (*CaCaC*), and (74) for B2 (*niCCaC*).

(73) Future masculine singular paradigm of B1 (*CaCaC*) – Level of number

	1st person	2nd person	3rd person	Stem	
a.	eʃmór	tiʃmór	yiʃmór	ʃmor	‘guard’
b.	egdál	tigdál	yigdál	gdal	‘grow’
c.	eléx	teléx	yeléx	lex	‘walk’
d.	asím	tasím	yasím	sim	‘put’

(74) Future masculine singular paradigm of B2 (*niCCaC*) – Level of number

	1st person	2nd person	3rd person	Stem	
a.	ekanés	tikanés	yikanés	kanes	‘enter’
b.	ivaléd	tivaléd	yivaléd	valed	‘be born’
c.	esóg	tisóg	yisóg	sog	‘retreat’

Higher up in the hierarchy, we expect to find more stem diversity, simply because the higher paradigm consists of more sub-paradigms. This is shown for B1 in (75)-(77) below.

(75) Future feminine paradigm of B1 (Level of gender)

	1st person	2nd person	3rd person	Stem	
a.	egdál	tigdelí/ú	tigdál/yigdelú	gdal/gdel	'grow'
b.	eléx	telxí/ú	telex/yelxú	lex/lx	'walk'
c.	ekrá	tikreí/ú	tikrá/yikreú	kra/kre	'read'

(76) Future paradigm of B1 (Level of tense)

	Paradigm	Stem	
a.	egdál/tigdál/tigdelí/yigdál/tigdál/nigdál/tigdelú/yigdelú	gdal/gdel	'grow'
b.	eléx/teléx/telxí/yeléx/teléx/neléx/telxú/yelxú	lex/lx	'walk'
c.	ekrá/tikrá/tikreí/yikrá/tikrá/nikrá/tikreú/yikreú	kra/kre	'read'

(77) Verb paradigm of B1 (Level of category)

	Paradigm	Stem	
a.	tigdál/tigdelí/gadál/gadlá/gadél/gdelá	gdal/gdel/gadal/gadl/gadel	'grow'
b.	teléx/telxí/haláx/halxá/holéx/holéxet	lex/lx/halax/halx/holex	'walk'
c.	tikrá/tikreí/kará/kará/koré/korét	kra/kre/kara/kore	'read'

The similarity among the paradigm members decreases higher up in the hierarchy. Each level has its unique characteristics, separate from its position in the hierarchy. This consistency can be either in the stem, in the derivational affixes, or in both.

Similarity is important because it is how we group verbs into the various paradigms, and how we tell paradigms apart.[39] It is also, however, why we mix up paradigms. ***The more similar two distinct paradigms are, the more likely it is for speakers to confuse them.*** The more speakers confuse them, the more likely they are to merge. But how can we measure similarity?

Speakers can tell when two items, linguistic or otherwise, are similar. They may not always agree on the degree of similarity, but they have a sense of what is similar and what is not. Nevertheless, similarity is not a binary attribute such that two words,

[39] There may, of course, be other types of groupings of verbs, for example based on semantics, syntactic functions, etc. Such groupings are not discussed here.

or any linguistic entities, are either similar or they are not. If this were the case, it would be extremely difficult to find two words with nothing at all in common. At the very least, every word comprises consonants and vowels, so Hebrew [yad] ‘hand’ and English [hæt] ‘hat’ would be considered similar as they are both CVC strings.[40] Indeed, in this respect they are similar, but what is the meaning of this similarity? ***Similarity has meaning if it has a function.*** We must limit the domain of similarity such that not every random string of sounds that accidentally resembles another random string can be said to be similar in a way that is linguistically meaningful.

If two objects share properties, they are more similar than two objects that share no properties. The more properties they share, the more similar they are. So similarity is both comparative and gradient. But before we can determine if words are similar at all, we need to define the properties that are relevant for the comparison. For the purposes of this study, I limit the comparison to phonological properties, such as prosodic structure and segmental content. Once we have identified the phonological properties that are relevant for the comparison of any two words, we can use them to compare entire paradigms.

The similarity between paradigms will be defined as the sum of the similarity between their members. Comparing full paradigms is necessary because, I contend, ***in the Hebrew verb system, it is similar paradigms that have a tendency to merge; not individual forms***. Comparing verb paradigms should be straightforward, as most paradigms, being inflectional, have the same number of members, and the same structural hierarchy.[41] Following Hyman (1970), Zwicky (1976), Steriade (2001a, 2001b), Kenstowicz (2007), Cohen (2009), among others, I assume that feature counting is not enough to determine similarity and I proposes a model of similarity that is gradient based on relative distance between comparable paradigms.

[40] Of course, they are also both nouns, they are both monosyllabic, they both have a low vowel in the nucleus, and a coronal obstruent in the coda.

[41] Some paradigms are incomplete and do have gaps in the form of missing members. When comparing these paradigms to complete paradigms, these gaps are counted as dissimilarities and so the dissimilarity is expected to be too great to license merger.

4.3. A Model of Similarity

Similarity is at the core of Optimality Theory. With faithfulness constraints requiring identity between two levels of representation (e.g. input-output, output-output, base-reduplicant, etc.) competing against structural markedness constraints, OT sets out to explain phonological contrasts as well as similarities. Within the Hebrew verb system, where the binyan dictates similarity among its verbs through prosodic and segmental constraints on surface forms (see Chapter 2), the degree of similarity plays a major role in determining which paradigms may merge. Recall from §3.3 that most of the change in the Hebrew verb system is either through merger among the sub-classes or through substitution, and that similarity is crucial for determining which sub-classes may interact in the change process.

Membership in a specific binyan necessarily means a certain degree of similarity among its members. Nevertheless, the surface variation exhibited within each binyan, and more specifically the surface variation that is notably unattested within the binyan, suggests that ***the interaction of sub-classes in the change process is not random and not all interactions are possible***. If surface variation is attested between A and B, but not between A and C (where A, B, and C are sub-classes within the same binyan), this means, in essence, that A&B are more similar than A&C and that speakers are sensitive to the difference in similarity between the two pairs of sub-classes. Similarity, therefore, must be quantifiable.

In the context of loanword adaptation, Cohen (2009) claims that speakers adapt non-native sounds to the closest phoneme in their language and proposes a model of similarity that is gradient based on the relative distance between comparable segments and not on the basis of feeling as proposed by Hyman (1970). Cohen shows that the selection of the closest native sound is done based on the measuring of the distance between the non-native segment and the candidate native segments. The distance is measured in units named *just noticeable difference (jnd)*, the minimum amount by which the intensity of a stimulus must be changed in order to produce a noticeable variation in sensory experience (see Cohen (2009) for a full discussion). The native phoneme with the lowest distance is more likely to be selected as the replacement for the non-native sound.

In the spirit of Cohen (2009), I propose a model of similarity that is gradient based on the relative distance between comparable paradigms. As in Kenstowicz's (1996), and McCarthy's (2005) models, this model of change, too, is based on output-output relations among members of the inflectional paradigm without assuming a base (cf. Benua, 1997, Albright, 2008b). In this model, similarity is a prerequisite and identity is the end result. In McCarthy's (2005) Optimal Paradigms, candidates consist of entire inflectional paradigms. The candidate paradigms are evaluated against markedness constraints, against faithfulness constraints requiring identity between each member of the paradigm and the underlying input (IO constraints), and between each member of the paradigm (OP constraints). The winning paradigm is the one that best satisfies the constraint hierarchy. McCarthy improves on Benua's model by enabling to simultaneously evaluate paradigms that do not share the same underlying stem. Thus affixes can be levelled as well as morphological templates. But while earlier studies concentrated on intra-paradigm relations, this study focuses on inter-paradigm relations, where the comparable paradigms do not share a stem.

In the Hebrew verb system, much of the observed variations result from analogy to other verbs that do not share a common stem. These verbs are typically in the same binyan, but belong to different sub-classes. Sub-class membership determines the conjugation pattern. Of course, historically, the segmental content of the stem consonants dictated the quality of the surrounding vowels within the template of the binyan; however, synchronically, there is no telling these segments apart, and so sub-class membership for the most part is arbitrary.[42] The meaning of this arbitrariness is

[42] The writing system maintains the distinctions among the gutturals, such that each guttural is assigned a unique symbol, and therefore class membership should be straightforward. However, children acquire the verb system before they master the writing system. Moreover, the fact that variation is widespread in adult speech suggests that the knowledge gained from the writing system

that each sub-class is an independent paradigm template. So if paradigm levelling is by definition within the paradigm, how could it occur among verbs that do not share a common stem and each has an independent template? In other words, ***can inter-paradigm levelling exist alongside intra-paradigm levelling?*** In the following sections, I will argue that it is possible, but only within the binyan and only between similar sub-classes.

The loss of the gutturals discussed in previous sections, resulted in loss of cues, making the classification of verbs into the various sub-paradigms (sub-classes) within the binyan more difficult. In the more frequent verbs, the loss of the gutturals would not have been problematic as these verbs are more easily retrieved. But when faced with a less frequent verb or an altogether new verb, the speaker must invoke the grammar in order to generate it. Without the necessary cues regarding the guttural consonants, there is no reason to assume a guttural at all. The speaker can either randomly select one of the binyan's sub-paradigms or compare the verb to the other sub-paradigms and conjugate it according to the most similar familiar verb. Random selection of a sub-paradigm is possible, of course, but even then, speakers will compare the result to the other similar sub-paradigms in order to verify that their choice is the optimal one.

Similarity is thus actively invoked in the verb system in order to select a sub-paradigm for an unfamiliar verb that has lost its uniqueness due to the segmental loss of the gutturals. In the following section, I propose a mechanism for comparison.

4.3.1. The Principles of Similarity

The following is an example of how similarity between forms can be measured. Each difference between comparable forms (marked by a rectangle) counts as a single unit of dissimilarity. The dissimilarity units are summed up and the sum, marked by the Greek letter delta, Δ, determines the degree of similarity (or difference). This sum is

has little or no effect on the speakers' grammar (though degree of literacy and awareness may play a role).

compared among competing paradigm pairs, resulting in a scale of similarity. The smaller the difference, the greater the similarity.[43]

Only properties of the binyan are compared, i.e. derivational affixes, vocalic patterns, and prosodic positions. The quality of the stem consonants are ignored, although their presence or absence is compared, as this is a prosodic property. Thus, in (78a), all the stem consonant positions are filled in both verb paradigms and we ignore the fact that the consonants are not identical as they do not contribute to the shape of the paradigms. The stem consonants have, therefore, been greyed out.

(78) Quantifying similarity

a. h i t l a b é ʃ - y i t l a b é ʃ ‘dress oneself’
h i t k o f é f - y i t k o f é f ‘bend’
1 1 Δ2

b. h i t l a b é ʃ - y i t l a b é ʃ ‘dress oneself’
h i t k a s á - y i t k a s é ‘cover oneself’
1 1 1 Δ3

c. h i t k o f é f - y i t k o f é f ‘bend’
h i t k a s á - y i t k a s é ‘cover oneself’
1 1 1 1 1 Δ5

d. h i t p a l é - y i t p a l é ‘be amazed’
h i t k a s á - y i t k a s é ‘cover oneself’
1 Δ1

[43] The model would work equally well if we counted similarities rather than dissimilarities. There are, nevertheless, several advantages to counting dissimilarities: it is easier to deal with smaller numbers, and dissimilarities are more easily understood: Δ0 always means that the compared forms are identical, whereas S14 (where S stands for ‘similarity’) does not necessarily mean identity.

In the examples in (78), only the past and future 3rd pr. ms.sg. forms are compared (past-to-past and future-to-future). In (78a), the second stem vowel [é] is identical in both past forms and in both future forms. The first stem vowel, which ideally would also be identical among all past forms and also among all future forms, differs in the two comparable paradigms. Each instance of such a difference receives a dissimilarity mark (the dissimilar segments are marked by a rectangle). In this way, the comparisons count dissimilarities, which in (78a) amount to 2 dissimilarity units (∆2).

In (78b), a regular verb is compared to a weak vowel-final verb. The final stem vowel is different in the past, but not in the future, resulting in one dissimilarity mark for the vowels. But they also differ prosodically as the regular verb is consonant-final and the weak verb is vowel-final in both past and future forms. Two dissimilarity marks are added, amounting to a total of ∆3.

In this way, all types of paradigms can be compared to one another. However, it is not enough to compare only the representative past and future 3rd pr. sg. forms (as demonstrated in (78)). To get a true sense of which paradigms are more similar to which, it is necessary to compare entire paradigms. Every paradigm has 16 verb forms (8 in the past tense, and 8 in the future tense). Each member of a paradigm is compared to its parallel member in the comparable paradigm. The dissimilarity measurements from all the members of the paradigm pair are summed, resulting in the similarity measurement of the entire paradigm pair. An example of the comparison between two complete paradigms is provided in Appendix A.

4.3.2. Similarity in B5: Degree of Similarity

The active invoking of similarity in order to select a sub-paradigm is necessary for verbs that have lost the necessary cues for sub-paradigm categorisation. So, for example, the normative *hitpalʔú* ‘they were surprised’ and *hitkasú* ‘they covered themselves’ are one dissimilarity unit apart. However, in MH, the two forms are identical because of the loss of the glottal stop resulting in *hitpalú* like *hitkasú*. The mechanism of similarity is henceforth applied on the modern surface structure.

Having compared the full paradigms in (78), the comparison of the entire paradigms reveals the following similarity measurements in B5 (see example in Appendix A; every paradigm has 16 members):

(79) Whole-paradigm similarity measurements in B5

	Comparable Paradigms					Similarity
a.	hitmalé	'was filled'	cf.	hitkasá	'covered himself'	Δ2
b.	hitlabéš	'dressed himself'	cf.	hitkoféf	'bent'	Δ21
c.	hitlabéš	'dressed himself'	cf.	hitmalé	'was filled'	Δ21
d.	hitlabéš	'dressed himself'	cf.	hitkasá	'covered himself'	Δ23
e.	hitkoféf	'bent'	cf.	hitmalé	'was filled'	Δ43
f.	hitkoféf	'bent'	cf.	hitkasá	'covered himself'	Δ44

When comparing the various forms in B5, it is clear that (79a) is the most similar pair (with only two dissimilarity units from the past 3rd pr. ms.sg. and fem.sg.), and thus merging is enabled. And indeed, variation is observed in only these sub-classes of B5 forms.

(80) Variation in B5

a. Variation

Variable Paradigm				Comparable Paradigm			Similarity
hitmalé	≈	hitmalá	'was filled'	cf.	hitkasá	'covered himself'	Δ2

b. No Variation

hitlabéš 'dressed'	cf.	hitkoféf	'bent'	Δ21
hitlabéš 'dressed'	cf.	hitmalé	'was filled'	Δ21
hitlabéš 'dressed'	cf.	hitkasá	'covered himself'	Δ23
hitkoféf 'bent'		hitmalé	'was filled'	Δ43
hitkoféf 'bent'		hitkasá	'covered himself'	Δ44

4.3.3. Similarity in B4: Directionality

In B4, more variation exists than in B5. This is because there are several distinct paradigms with very high levels of similarity, as shown in (82).

(81) Whole-paradigm similarity measurements in B4

Comparable Paradigms						Similarity
a.	milé	'filled'	cf.	nisá	'attempted'	Δ7
b.	exér	'was late'	cf.	gidél	'grew'	Δ8
c.	gerá	'excited'	cf.	nisá	'attempted'	Δ8
d.	gerá	'excited'	cf.	milé	'filled'	Δ15
e.	milé	'filled'	cf.	gidél	'grew'	Δ22
f.	exér	'was late'	cf.	milé	'filled'	Δ25

The following table shows the variation in B4, with the variable paradigm presented alongside its comparable paradigm, which triggered the variation.

(82) Variation in B4

a. Variation

Variable Paradigm				Comparable Paradigm			Similarity
milé	≈	milá	'filled'	cf.	nisá	'attempted'	Δ7 (bidirectional)
exér	≈	ixér	'was late'	cf.	gidél	'grew'	Δ8 (unidirectional)
gerá	≈	girá	'excited'	cf.	nisá	'attempted'	Δ8 (unidirectional)

b. No Variation

gerá	'excited'	cf.	milé	'filled'	Δ15
milé	'filled'	cf.	gidél	'grew'	Δ22
exér	'was late'	cf.	milé	'filled'	Δ25

The variation in B4 (82a) is more complex than in B5 (79) because in B5 it is strictly bidirectional: verbs from both paradigms migrate to the comparable paradigm. Recall that in each pair in the variation relation, the first member is the standard one and the new one arises, as I argue, due to similarity. The example in (83) shows the bidirectional nature of the change in B5. Members migrate from both paradigms to both paradigms. The normative forms, the older forms, are underlined.

(83) Bidirectional variation in B5

a. hitmalé ≈ hitmalá ‘was filled’ cf. hitkasá
b. hitkasá ≈ hitkasé ‘covered himself’ cf. hitmalé

hitmalá
Comparable paradigms: **hitmalé** – **hitkasá** ‘was filled’ – ‘covered himself’
hitkasé

Unlike B5, B4 exhibits three types of migration paths (where A, B, and C are comparable paradigm pairs). The migration paths are illustrated in (84) below, followed by their definitions.

(84) Migration paths

a. Bidirectional b. Unidirectional c. Multipath

$A \leftrightarrow B$ $A \rightarrow B$ $A \begin{smallmatrix} \nearrow B \\ \searrow C \end{smallmatrix}$

Bidirectional variation (84a): Members of A migrate to B and members of B migrate to A. From the perspective of each paradigm type, members defect and new members are admitted into the group (i.e. the migration is in both directions).[44] The same type of bidirectional migration exhibited in B5 above (83) is also exhibited in B4 (85) below.

(85) Bidirectional variation in B4

milé ≈ milá ‘filled’ cf. gilá
gilá ≈ gilé ‘discovered’ cf. milé

milá
Comparable paradigms: **milé** – **gilá** ‘he filled’ – ‘he discovered’
gilé

[44] Bidirectionality does not imply symmetry. If paradigm B is the most similar paradigm to A, this does not mean that A is also the most similar paradigm to B (Tversky 1977).

Unidirectional variation (84b): Members of A migrate to B, but members of B do not migrate to A. From the perspective of each of the paradigm types, members either defect or are admitted, but not both (i.e. the migration is in one direction only), as shown in (86) below.

(86) Unidirectional variation in B4

a.	exér ≈ ixér 'was late'	cf. gidél
	*gidél ≈ gedél 'grew'	cf. exér
b.	gerá ≈ girá 'excited'	cf. nisá
	*nisá ≈ nesá 'attempted'	cf. gerá
c.	sovév ≈ sivév 'turned'	cf. ximém
	*ximém ≈ xomém 'heated'	cf. sovév

ixér
Comparable paradigms: **exér** – **gidél** 'he was late' – 'he grew'
*

Multipath variation (84c): For the comparable paradigms A, B, and C, members from A migrate to B in certain conditions, and in other conditions members of A migrate to C. Members of B or C do not migrate to A. From the perspective of paradigms B and C, the migration is in one direction only. However, the migrating forms (the members of A), have two migration paths (B and C), as shown in (87) below. This type of variation is unique and is discussed in the next section (§4.3.7).

(87) Multipath variation in B4

bit͡séa ≈ bit͡sá 'he executed'	cf. nisá 'he attempted'
bit͡sáti ≈ bit͡séti 'I executed'	cf. miléti 'I filled'

bitsa
Comparable paradigms: **bitséa** – **nisá** 'he executed' – 'he tried'
Comparable paradigms: **bitsáti** – **miléti** 'I executed' – 'I filled'
bitseti

Variation in the course of change is due to similarity. Similarity can predict which paradigm types are analogical, but it cannot explain why for some paradigms the change is bidirectional and for others it is unidirectional or multipath. I argue that ***the direction of change, whose initial sign is variation, is greatly affected by type frequency***, i.e. the number of verbs in the sub-class. Members of the paradigm type with the lower type frequency will typically migrate to the paradigm type with the higher type frequency.

Type frequency can only predict unidirectional change as the surviving paradigm is the one with the higher type frequency, and the paradigm type with the lower type frequency becomes extinct. In bidirectional variation, change in one direction is towards the paradigm with the higher type frequency, but in the other direction it is inevitably towards the paradigm with the lower type frequency. If unidirectionality is always to the higher type frequency paradigm, then the only way for bidirectionality to be possible is if the two comparable paradigms have identical type frequencies, which is highly unlikely. The ratio between the type frequency of the two paradigms can tell us something about the difference in their type frequencies. A small ratio means that the difference in type frequencies is small, such that it is difficult to tell which of the comparable paradigm types is the larger group of verbs. ***When the ratio is small, there is a greater chance for bidirectional variation as some speakers may choose one direction and other speakers may choose the other direction.*** A large ratio means that the paradigm with the larger number of members is easily identifiable and so speakers will naturally level the smaller group accordingly and the variation will be unidirectional.

The following table shows the ratio between some of the comparable paradigms discussed in this study as well as their actual directionality of change.

(88) Groups size and directionality

Variation		**Comparable Paradigm**			**Ratio**	**Directionality**
hikír ≈ hekír	‘recognised’	cf.	hekím	‘founded’	1.65	Bidirectional
heví ≈ hiví	‘brought’	cf.	hisí	‘married’	3.00	Bidirectional
<u>milé</u> ≈ milá	‘filled’	cf.	gilá	‘revealed’	4.25	Bidirectional
hitmalé ≈ hitmalá	‘was filled’	cf.	hitkasá	‘was covered’	5.12	Bidirectional
<u>exér</u> ≈ ixér	‘was late’	cf.	gidél	‘grew’	17.21	Unidirectional
sovév ≈ sivév	‘turned’	cf.	ximém	‘heated’	17.21	Unidirectional
gerá ≈ girá	‘teased’	cf.	nisá	‘attempted’	51.00	Unidirectional

The ratio between the most similar paradigms within the binyan enables to determine which of the paradigm types is the larger group. Further study is required in order to determine the threshold below which groups are perceived as small and above which groups are perceived as large. This threshold is probably affected by the number of group members to which speakers are exposed and may, therefore, vary among speakers.

Group size may not be the only factor influencing directionality, though. All the variation in B3 is bidirectional, regardless of group size. In the other binyanim, while all bidirectional change exhibit a small type frequency ratio between the paradigms involved, not all small type frequency ratios result in bidirectionality. For example, *nexkár–yexakér* ‘be investigated B2’ has a variant paradigm *nixkár–yixakér* (cf. *nixnás–yikanés* ‘enter’). The variation is unidirectional (as there is no attested variant paradigm *nexnás–yekanés*), even though the ratio between the group sizes is only 4.61.

Bidirectionality is not the optimal migration path. It does not result in fewer patterns, or in simplifying the verb system. Other factors must be at play in order to block bidirectionality. Further research is required in order to determine what these factors may be.

4.3.4. *Similarity in B3*

In B3, there seems to be a general tendency of change from initial vowel /i/ to [e], causing the entire binyan to become *hefíl,* instead of *hifíl*. It appears as though the [e] in initial position in forms where it is governed by constraints has randomly spread to all other *hifíl* verbs, including the regular verbs. However, this is not the whole story. In fact, it is not a general tendency of the binyan at all. Although [e] in initial position is increasingly surfacing, it is not doing so randomly. The binyan is not changing to *hefíl*.

Although B3 exhibits much more variation than B4 or B5, the variations exhibited are of the same nature as in the other binyanim. The reason for the increased number of surface variation results from the elevated number of sub-classes.

The variations exhibited in B3 are as follows (similarity measurements are for the entire paradigm pair, as described in Appendix A):

(89) Variation in B3 (bidirectional)

a.	higdíl	≈	hegdíl	'enlarged'	cf.	hexlít	≈	hixlít	'decided'	Δ8
b.	himt͡sí	≈	hemt͡sí	'invented'	cf.	hexbí	≈	hixbí	'hid'	Δ8
c.	hikír	≈	hekír	'recognised'	cf.	hekím	≈	hikím	'founded'	Δ8
d.	hisí	≈	hesí	'married'	cf.	heví	≈	hiví	'brought'	Δ8
e.	hirʃá	≈	herʃá	'allowed'	cf.	hexná[45]	≈	hixná	'parked'	Δ8
f.	hiká	≈	heká	'hit'	cf.	herá[46]	≈	hirá	'showed'	Δ9

For other forms which appear to be similar but for which there is no evidence of surface variation, their similarity is predicted to be less prominent. Consider, for example, the comparable paradigms in (90). They appear to be similar enough (e.g. Δ12), and yet there is no surface variation to suggest that they are merging. This means that each of these sub-classes has a more similar sub-class with which to merge.

---

[45] Formally, *hexená*. I submit here the colloquial form in which epenthesis does not occur.
[46] The comparable form here is again the colloquial form in which the glottal stop is deleted.

(90) No Variation in B3

a.	himt͡sí	'invented'	cf.	hiršá	'allowed'	Δ12
b.	hisí	'married'	cf.	hiká	'hit'	Δ12
c.	hexbí	'hid'	cf.	hexná	'parked'	Δ12
d.	heví	'brought'	cf.	herá	'showed'	Δ12
e.	higdíl	'enlarged'	cf.	hikír	'recognised'	Δ16
f.	hexlít	'decided'	cf.	hexbí	'hid'	Δ21

A similarity of Δ12, as in (90a), for example, could probably yield variation (*himt͡sí* ≈ **himt͡sá*). Nevertheless, there is no sign of variation here. In fact, we would not expect there to be any variation as long as there is another paradigm with a higher similarity, i.e. *himt͡sí* ≈ *hexbí* (89b) with a similarity of Δ8.

4.3.5. Similarity in B2: Non-alternating Paradigms

In B2, the regular sub-class of verbs has an {ia} vowel pattern in the past tense, and an {ea} vowel pattern in the future tense, with an external *Ci-* prefix.

(91) B2 regular Past-Future paradigms (Normative)

Past	**Future**	
nixnás	yikanés	'enter'
niʃbár	yiʃavér	'be broken'
nivdák	yibadék	'be examined'
nigrám	yigarém	'be caused'

In a previous stage of the language, as is in the normative language, the high vowel is lowered when followed by a guttural-initial or *r*-initial stem.

(92) B2 vowel lowering before gutturals and *r* (Normative)

Past	**Future**	
neʕetsár	yeʕatsér	'be stopped'
neħkár	yeħakér	'be interrogated'
nehená	yehané	'enjoy'

In some cases with *r*-initial stems, lowering occurs only in the future tense.

(93) B2 partial application of vowel lowering (Normative)

Past	Future	
nirdám	yeradém	‘fall asleep’
nirtáv	yeratév	‘get wet’
nirgáʕ	yeragáʕ	‘calm down’

Thus, B2 normative verbs either exhibit no alternation such that verbs that have a high vowel in the past tense also have a high vowel in the prefix of the future tense (91), and verbs that have a mid vowel in the past tense also have a mid vowel in the prefix of the future tense (92), or they do exhibit alternation between a high stem initial vowel in the past tense and a mid vowel in the prefix of the future tense (93). The type frequency of these three groups and an additional group found in B2 are given below.

(94) B2 normative alternation patterns

	Alternation			Type Frequency
a.	No alternation [i]	nixnás–yikanés	‘enter’	210
b.	No alternation [e]	nexkár–yexakér	‘be investigated’	69
c.	Alternation [i]~[e]	nirdám-yeradém	‘fall asleep’	17
d.	Other	nolád–yivaléd	‘be born’	13

There is a clear preference for non-alternating paradigms over alternating ones, and also for paradigms with a high initial vowel over a mid initial vowel.

Synchronically, though, with the gutturals out of the picture, the environment for vowel lowering is gone, and with it, so are the alternating paradigms. Speakers are getting rid of the alternation within the paradigm, aligning them with the most frequent pattern, that with the high vowel.

(95) Variation in B2

yeradém	≈	yiradém	‘fall asleep’	cf.	yikanés	‘enter’	Δ7
yeragá	≈	yiragá	‘calm down’	cf.	yiʃamá	‘be heard’	Δ7
nexkár	≈	nixkár	‘be investigated’	cf.	nixnás	‘enter’	Δ15
yexakér	≈	yixakér	‘be investigated’	cf.	yikanes	‘enter’	Δ15

The non-alternating paradigms with initial [e], are being levelled with the more frequent non-alternating paradigms with [i], leaving non-alternating [e] paradigms only in cases of a following vowel, as in the two last paradigms in (96) below:

(96) Synchronic alternation in B2

nirdám	–	yiradém	'fall asleep'	cf.	yikanés	'enter'	Δ7
nirgá	–	yiragá	'calm down'	cf.	yiʃamá	'be heard'	Δ7
nixkár	–	yixakér	'be investigated'	cf.	nixnás	'enter'	Δ15
neetsar	–	yeatser	'be arrested'				
neena	–	yeane	'enjoy'				

Note that normative *nexkár* 'was investigated' in (95) is merging with the regular paradigm *nixnás* 'entered' (Δ15), even though the normative *nirdám* 'fell asleep' paradigm is more similar, with only Δ8 on the similarity scale. However, merging with *nirdám–yeradém* will result in an alternating paradigm, accentuating the problem, rather than fixing it. In B2, the cause for reducing the number of paradigm types has joined forces with another requirement for non-alternation within the paradigm (an OP constraint) blocking the merger with the most similar paradigm, when the result is an alternation within the paradigm.

4.3.6. Similarity in B1

In B1, things are even more complicated as the binyan has 44 distinct sub-classes, i.e. paradigm types, collectively hosting 666 verbs (see Appendix B for a complete list).[47] Out of these 44 paradigm types, 15 account for 91% of the verbs. Many of the B1 paradigms host a single unique verb, and many verbs are no longer in use or have a very low frequency (e.g. *gadáʃ* 'overfill,' *hamám* 'stun' *zaváx* 'sacrifice'). I will discuss only the paradigm types that are being replaced by another paradigm type (97), typically the paradigm with a higher type frequency that is most similar to it.

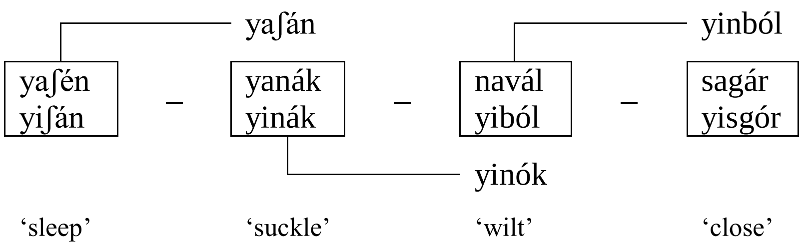
[47] Tarmon and Uval (1998) list 64 distinct B1 paradigms. However, if we consider the past and future tenses only (ignoring the present, imperative, and infinitive), these can be reduced to 44 distinct types.

(97) B1 Variation

	Normative Paradigm	Colloquial Variant	Type Freq.		Comparable Paradigm	Type Freq.	Similarity
a.	yaʃén–yiʃán	yaʃán–yiʃán	1	cf.	yanák–yinák	7	Δ2
b.	yanák–yinák	yanák–yinók	7	cf.	nafál–yipól	17	Δ6
c.	navál–yiból	navál–yinból	17	cf.	sagár–yisgór	226	Δ10
d.	nazál–yinzál	nazál–yinzól	3	cf.	sagár–yisgór	226	Δ13
e.	katón–yiktán	katán–yiktán	1	cf.	lamád–yilmád	90	Δ5

Note the daisy chain in (97) above: verbs in the normative paradigm in (97a) (*yaʃén–yiʃán* 'sleep') are migrating to the normative paradigm in (97b) (*yanák–yinák* 'suckle'), verbs in the normative paradigm in (97b) are migrating to the normative paradigm in (97c) (*nafál–yipól* 'fall') and verbs from the normative paradigm in (97c) are migrating to the regular paradigm (*sagár–yisgór* 'close').

(98) Migration daisy chain



The verbs in (97a) and (97b) cannot migrate directly to the regular paradigm (*sagár–yisgór*), though, because of the low degree of similarity.

(99) Variation in B1

a. Variation

yaʃén ≈ yaʃán	cf.	yanák	Δ2
yinák ≈ yinók	cf.	yipól	Δ6

b. No Variation

yaʃén	cf.	yipól	Δ6
yaʃén	cf.	yisgór	Δ17
yinák	cf.	yisgór	Δ17

Although verbs cannot skip a step to go directly to the paradigm with the highest frequency, as indicated in (99), once they have climbed a step to the most similar paradigm, there is nothing stopping them from climbing another step. This type of similarity effect reduces the number of paradigm types from 44 to 30 (30% reduction), predicting further reduction in the future.

4.3.7. Multipath Variation

As mentioned in §4.3.3, there is a unique case of multipath change in B4, in which the older normative form has two possible migration paths, but each path is unidirectional. The members of the comparable paradigm, if any, will not migrate to this group.

(100) Multipath variation in B4

bi͡tséa ≈ bi͡tsá ‘he executed’	cf. nisá ‘he attempted’	Δ6
bi͡tsáti ≈ bi͡tséti ‘I executed’	cf. miléti ‘I filled’	Δ6

Recall that in B4, when a *x*-final stem originating from the historical *ħ* or a vowel-final stem originating from a historical *ʕ* appears word finally, lowering occurs and a diphthong emerges as a result; e.g. *ni͡tséax* (histrocially *ni͡tséaħ*) ‘he won’, *bi͡tséa* (historically *bi͡tséaʕ*) ‘he executed’). In §3.3.1, I claimed that the three paradigms (*milé* ‘he filled,’ *nisá* ‘he attempted’, and *bi͡tséa* ‘he executed’) were analogous as exemplified by the levelling effects exhibited (*milá*, *nisá*, and *bi͡tsá*). However, *bi͡tsá* and *nisá* have not really merged, although they are equally similar to the *bi͡tsá* and *milé* pair (Δ6 for both paradigm pairs). That *bi͡tsá* and *nisá* are not merging is evidenced by other members of the paradigm (e.g. *bi͡tsáti* ‘I executed’ vs. *nisíti* ‘I attempted’, but **bi͡tsíti*). The new *bi͡tsá* paradigm is not the result of merger with any existing paradigm, and is thus a unique type of pseudo levelling. This new paradigm results from the simplification of the diphthong. The surface realisation of a diphthong in final position is opaque, and so the new paradigm eliminates it. The diphthong comprises two vowels, an initial *e* and a final *a*. Speakers have a choice which member to keep and which to discard. The mid vowel is comparable to the glottal final sub-class paradigm (*milé* ‘he filled’). The low vowel is not like any other paradigm, but it is the trademark of the group of verbs with a historical pharyngeal in final position. The option with the low vowel (*bit͡sá* ‘he executed’) has a higher token frequency than the one with the mid vowel (*bit͡séti* ‘I executed’ – only two tokens in the recorded database). But because *bit͡sá* is limited to formal speech whereas *bit͡séti* follows the regular unidirectional merging path, *bit͡séti* is more likely to increase in token frequency in regular speech because of its similarity to an existing sub-class paradigm.

4.3.8. Thresholds vs. scale

When comparing forms or paradigms in order to determine their degree of similarity, one of the things that we need to determine is whether or not there is a threshold beyond which the degree of similarity between forms or paradigms is no longer effective and merging will not occur. When comparing paradigms, such a threshold may very well exist, but paradigms will merge only with the most similar paradigm, even if another paradigm is well within that threshold. In (82b), the difference between the two paradigms *sovév* ‘turned’ and *ximém* ‘heated’ is Δ16, and merging occurs; but in (90e), where the difference between the two paradigms, *higdíl* ‘enlarged’ and *hikír* ‘knew,’ is also Δ16, no merging can occur (yielding either **hidíl* or **hinkír*, depending on the direction). The absolute number indicating the distance between two paradigms (Δ16 in the above example) is therefore not the most crucial indicator; rather, whether or not a more similar paradigm exists will determine whether or not the two paradigms can merge.

A threshold, therefore, need not be assumed to prevent paradigms from merging. If their level of dissimilarity is too high, then the probability of their merging will be low enough to never emerge.

4.3.9. *Qualitative Similarity*

Returning to Kuryłowicz's (1949) idea that some distinctions are more important than others (§4.1.1), it is conceivable that the more important distinctions have a greater effect on the degree of similarity than the less important distinctions. If so, then these distinctions should carry more weight during the comparison. If, for example, prosodic structure is more important than segmental quality, then differences in the existence or absence of a consonant between comparable paradigms should be granted more dissimilarity units (e.g. two dissimilarity units) than differences in the vocalic pattern (one dissimilarity unit). This will cause the two paradigms to be even more dissimilar, forcing merging paradigms to be prosodically identical.

In the examples cited above, (82b) and (90e), the qualitative similarity measurements granting two dissimilarity units for differences in the prosodic structure and only one dissimilarity unit for each difference in the segmental quality would be as follows:

(101) Qualitative Similarity

a.	sovév	cf.	ximém	Δ16
b.	higdíl	cf.	hikír	Δ32

The qualitative distinction does not change the empirical observation that merging occurs in (101a), but not in (101b), but while the quantitative distinctions were identical (Δ16 for both pairs), the qualitative similarity shows that the two pairs are not equally similar. The pair in (101a) differ only in one of the vowels throughout the paradigm, whereas the pair in (101b) differ in the number of stem consonants that appear throughout the paradigm. As this is the more important distinction, it is granted two dissimilarity units for each member of the paradigm, resulting in much greater dissimilarity.

Taking this idea even further, distinctions can be made according to position, i.e. coda vs. onset, where differences in the onset are granted more weight than differences in the coda. In this way, a difference in the existence or absence of a coda would be granted two dissimilarity units as before, whereas a difference in the existence or absence of an onset would be granted three dissimilarity units. Such positional distinctions can measure similarity more accurately, to the point that we can rely on it to group verbs into paradigms. Δ0 means that the two paradigms are identical, a difference of up to Δ10 would mean that the paradigms are similar enough to merge, etc.

The proposed model of similarity can be further refined by allowing different types of dissimilarities more influence than repeating dissimilarities. For example, the Δ16 dissimilarity units attributed to the *sovév* – *ximém* paradigm pair results from a single difference in the initial stem vowel that is repeated throughout the paradigm and therefore counted in all members of the paradigm. In this case, there are many instances of a single unique difference. In the paradigm pair *kaná* 'buy' – *kará* 'read,' on the other hand, the Δ14 dissimilarity units result from four distinct difference types: *i* vs. *a* (e.g. *kan**í**ti* vs. *kar**á**ti* – 1st. sg. past), *e* vs. *a* (e.g. *yikn**é*** vs. *yikr**á*** – 3rd. sg. future), *t* vs. Ø (e.g. *kan**t**á* vs. *kará* – 3rd. fem. sg. past), and *e* vs. Ø (e.g. *tikní* vs. *tikr**e**í* – 2nd. fem. sg. future). If we attribute a full dissimilarity unit to teach difference type and a partial dissimilarity unit for each repetition, say 0.1, then the similarity measurements would be as follows:

(102) Qualitative Similarity

a.	sovév	cf.	ximém	Δ2.5
b.	kaná	cf.	kará	Δ5.0

By attributing different weights to types of differences and tokens of the same difference, we can now explain why merging is more likely in (102a) than in (102b).

The refinements to the similarity model proposed in this section are not necessary for the Hebrew data and should be tested on languages in which such distinctions are crucial. The point made in this section is that the proposed model of similarity is scalable and can accommodate any types of distinctions that may be required for the language being investigated.[48]

4.4. Conclusion

When faced with a new, unfamiliar, or even a familiar but less frequent verb, speakers rely on similarity to group this verb with an existing verb sub-class (A). Once grouping is done, conjugation is done according to the pattern dictated by sub-class A. When two sub-classes A and B are similar enough, the new verb can be grouped with either sub-class, thus starting the process of merging of the two sub-classes. The reason why this is happening now in the Hebrew verb system is that the loss of the gutturals has caused these sub-classes to become more similar than before, with less dissimilarity cues, causing uncertainty as to which sub-class a verb belongs.

The model of similarity proposed in this chapter explains why some verb sub-classes merge and others do not. The merger is only between the most similar sub-classes in the binyan. Within a stochastic framework, where variation is probabilistic, it could be argued that the difference in similarity between comparable pairs denote the probability of their merging. Thus, a paradigm pair with similarity $\Delta 10$ is more likely to merge than a paradigm pair with similarity $\Delta 21$, but that is not to say that the least similar paradigms cannot merge. It could also be argued that the difference in similarity between comparable pairs predicts the order in which they will merge. Thus, a paradigm pair with similarity $\Delta 10$ is predicted to merge before a paradigm pair with similarity $\Delta 21$. Indeed, the proposed model makes this prediction, but only within the binyan, as illustrated in Chapter 6 in Figure 7, where merging is between the most similar paradigms. When the merging is complete, there is nothing to stop the paradigm from merging with the next most similar paradigm within the binyan. Whether or not a comparable pair with similarity $\Delta 10$ in one binyan is likely to occur

48 As noted by one of the reviewers, language-particular adjustments should be limited. A cross-language study is required to determine the limits of refinement.

before a comparable pair with similarity Δ21 in another binyan requires further research.

This model was shown to be scalable and accommodate any distinction required for the language being investigated.

CHAPTER 5. PREDICTING THE FUTURE OF CHANGE

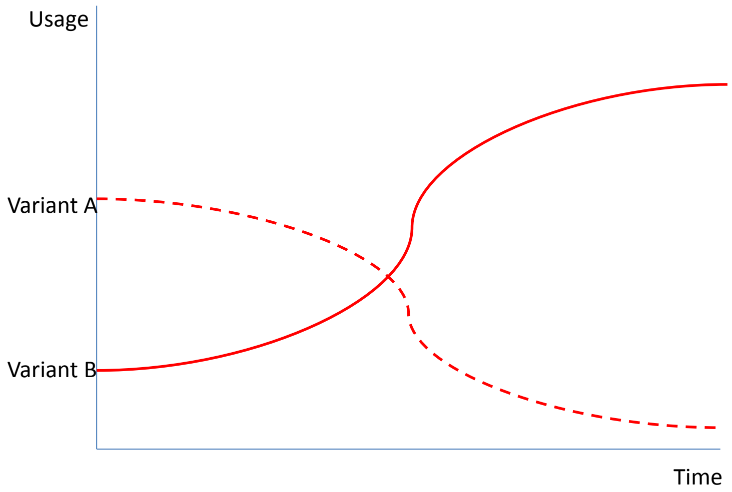
In the previous chapters, I showed that Hebrew has essentially two types of change patterns: unidirectional and bidirectional. Synchronically, these two patterns have the same effect, since in both change patterns the two variants, i.e. the older normative form and the newer colloquial form, co-exist to a varying degree of frequency. However, mapping the direction of each change revealed that in some cases, the migration is from one group to another, but not in the other direction, and in other cases the migration occurs in both directions.

For example, we saw in §3.4 that in the case of the future first person singular prefix, the migration is towards the third person prefix, but not from the third person to the first person. Both prefixes are exhibited on the surface, in varying degrees of frequency. In this case, the change is unidirectional; the first person prefix is being replaced by the third person prefix. We also saw in §3.3.2 cases of bidirectional change, in the merging of paradigms, which is regulated by similarity. *mileti* 'I filled' and *nisiti* 'I tried' have been shown to be similar enough to merge. And when merging occurs, verbs from each class move to the comparable class. In this case, too, the only manifestation of the change is the surface variation where both variants co-exist.

Variation can persist a very long time and the process causing the variation slows down as it nears the end. Modern Hebrew is in the midst of change. None of the verbs have completely transformed, and variation is still widespread. But while variation may stick around for a while, it is not the optimal state. Variation, particularly intra-speaker variation, means that speakers are not sure which variant to use and so they alternate. This state of affairs will resolve itself somehow. Eventually speakers will pick a variant and stick to it. Whether all speakers will pick the same variant, as in the English past tense, or different groups will choose different variants, forming sociolects, remains to be seen. In this study, I did not look at any sociolinguistic parameter that may affect variant choice. This requires further study. Nevertheless, looking at the language as a whole, it is possible to make predictions on the future of the change, based on the current frequency of its resulting variants.

In the unidirectional type of change, an older normative form is replaced by a newer colloquial form. The replacement is not immediate, and for an extended period of time both the old and the new form appear on the surface. But the two variants differ in the frequency of use. At first, the old form has a higher frequency as the new form appears. Then slowly the frequency of the new form increases as that of the old form decreases, until finally they both reach a plateau where the old form is hardly ever used (i.e. its frequency approaches zero) and the new form reaches the frequency of its old predecessor. An example of a graph for this type of change is presented below (103).

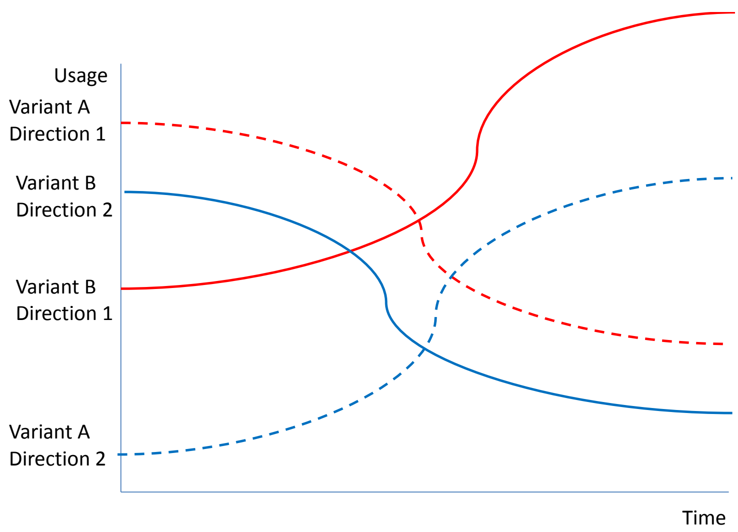
(103) Unidirectional Change Path



The frequency of usage of variant B, the newer form, increases over time at the expense of the competing variant A, the older form, whose usage decreases over time, until it ceases to exist. A similar change pattern is expected for the paradigm types, as more and more forms migrate from the A paradigm type to the B paradigm type.

In the bidirectional type of change where two groups merge, members of both groups migrate to the other group. While the migration is observed in both directions, the migration in one direction may be more dominant than in the other direction. That is, the frequency of the variants from one group may be higher than that from the other group. However, note that for the specific sub-class paradigm, the migration is unidirectional. Its members can only migrate to the sub-class paradigm most similar to them. And so the graph for the migration path of the specific class paradigm (i.e. in each direction) is the same as that in (103). The bidirectionality is reflected by the number of variants (i.e. the number of lines in the graph) and the token frequency of each variant locates the line in a specific place on the x axis (the usage axis), as illustrated in (104).

(104) Bidirectional Change Path



In this case, both paradigm types are expected to survive (the survivors of each direction).

I turn to learning algorithms in order to test this claim.

5.1. Why Learning Algorithms and Why GLA?

Learning algorithms are essentially computational models that simulate language learning. These models can be used and indeed have been used to test proposed grammars and prove that they are learnable. A learning algorithm computes the correct grammar provided it is supplied with suitable training data and it converges if it yields a result on every training set. Any generative linguistic framework must have an associated learning mechanism that describes how the grammars couched within that framework can be learned. Focusing on OT, perhaps the most well-known learning algorithm proposed for the learning of optimality-theoretic grammars is the Constraint Demotion Algorithm (1996, Tesar and Smolensky, 1998, Tesar and Smolensky, 2001).[49] Using the Constraint Demotion Algorithm, Tesar and Smolensky (1998) show that grammars with ordered constraint hierarchies are in fact learnable.

The algorithm is not designed to handle either free variation or gradient grammaticality, which means that grammars with multiple winners, such as the one discussed in the present study, cannot be shown to be learnable. Boersma and Hayes' (2001) modified version of OT, namely Stochastic OT and its associated learning algorithm, the Gradual Learning Algorithm (GLA) improve upon standard OT and the Constraint Demotion Algorithm in dealing with both free variation and gradient grammaticality.

In §3.4.1, I showed how stochastic OT deals with free variation by assigning a range of application for each constraint. When constraints overlap, free variation can occur. The degree of overlap reflects the frequency of occurrence of the competing rankings, thus reflecting the frequency of use of the variants.

Gradient grammaticality refers to speakers' judgement calls on the grammaticality of linguistic elements. Studies have shown that speakers can reliably make gradient well-formedness distinctions in morphology and phonology (Hayes and MacEachern, 1998, Hayes, 2000, Keller and Alexopoulou, 2001) and also in syntax (Bard et al., 1996, Keller, 2000). Therefore, gradient well-formedness is assumed to be part of native speakers' knowledge of language and as such should be accounted for by the linguistic theory. In stochastic OT and in GLA, gradient grammaticality is dealt with using the same tools applied for free variation. In this sense, the theoretical framework

[49] Other proposals include Pulleybland and Turkel's (2000) Genetic Algorithms and also Hale and Reiss (1998) algorithm, which posits an initial ranking of Faith constraints above Markedness constraints.

as well as the learning algorithm associated with it treat gradience as frequency. That is, the degree of acceptability of a linguistic element is reflected by the frequency of its use. Completely unacceptable forms are deemed ungrammatical and should be disallowed by the grammar. Acceptable forms should be allowed by the grammar, but their frequency depends on their degree of acceptability. The more acceptable the variant is, the higher its frequency will be, compared to that of the competing variant.

The idea that gradient grammaticality and corpus frequency are related and can be treated within the same probabilistic model is not uncontroversial. As language consists of an infinite set of structures, there will always be structures that are grammatical, but have a very low frequency or will not appear at all in a finite corpus. The absence of a structure from the corpus cannot serve as evidence of its ungrammaticality. Therefore, probability of occurrence within a corpus (i.e. frequency of production) and well-formedness (or the degree of grammaticality) must be treated separately (Abney, 1996, Culy, 1998, Keller, 2000). This is the distinction made in language acquisition between competence (the knowledge about the grammar) and performance (the actual production). Based on experiments on the first stages of phonological acquisition, Hayes (2004) concludes that children show evidence of knowledge of phonotactics even before the first productions emerge. He therefore proposes a separate grammar for each. Pater (2004) proposes to incorporate competence and performance within a single grammar, but with a duplicate set of constraints: one set for Input-Output correspondence, and another for Output-Input correspondence.

In this study, I abstract away from gradient grammaticality associated with competence. I assume that if both variants of the same verb appear within the corpus, it means that both rankings responsible for the variation are accessible to speakers, even if they do not produce it themselves. Within the realm of language change, studies on intra-speaker variation may shed light on whether or not the frequency of usage within finite corpora reflects the degree of the variants' acceptability. I leave aside the controversy surrounding the competence/performance disparity and return to the various applications of learning algorithms.

Learning algorithms such as the Gradual Learning Algorithm can also be used to simulate language development. If the running of the algorithm simulates a child's learning process, and we assume that the grammar that the algorithm is fed is the language's only grammar, then the running of the algorithm simulates a generation's learning process. If so, then running the algorithm numerous times can simulate the learning process of multiple generations, assuming also that the output of one generation is the input for the next generation.

The input for GLA is a grammar consisting of an underlying representation, a set of constraints, a set of candidates, the frequency of each candidate in the language and markings of each candidate's violations of the constraints. The only thing that is not fed into the algorithm is the assumed ranking of the constraints. The algorithm assumes an initial non-ranking where all constraints are equal, and the actual ranking is derived based on the output and its frequency in the language.

5.2. Applying GLA

To simulate language development in the course of time, I apply evolOT (Jäger, 2002b), a software implementation of the Gradual Learning Algorithm and its variant, the bidirectional GLA (Jäger, 2002a) for Stochastic OT. In evolOT, a frequency distribution is defined over GEN, and the actual training corpus is generated by a random generator interpreting the relative frequencies as probabilities. The software enables the application of Jäger's version of the algorithm, which assumes bidirectional learning. Bidirectional GLA (BiGLA) both generates the optimal output for the observed input, and the optimal input for the observed output. The grammar that is acquired from a sample corpus that is used for another run of GLA/BiGLA may differ from the previously learned language. The absolute frequencies of the different inputs are kept constant in each learning cycle ('generation'). What may change from generation to generation are the relative frequencies of the different outputs for each input.

The following are two evolOT charts generated based on the relative frequencies of the variants in the recorded database. Figure 5 illustrates the evolution of the unidirectional change from the 1st pr.sg. prefix to the 3rd pr.ms.sg. prefix. At present, the constraints responsible for the surfacing of the two prefixes (*ʔe*- and *yi*-) are close, predicting the existing variation. At some point, the ranking of the constraints will cross and they will grow further and further apart, until the probability of the 1st pr.sg. prefix ever emerging will be quite low.

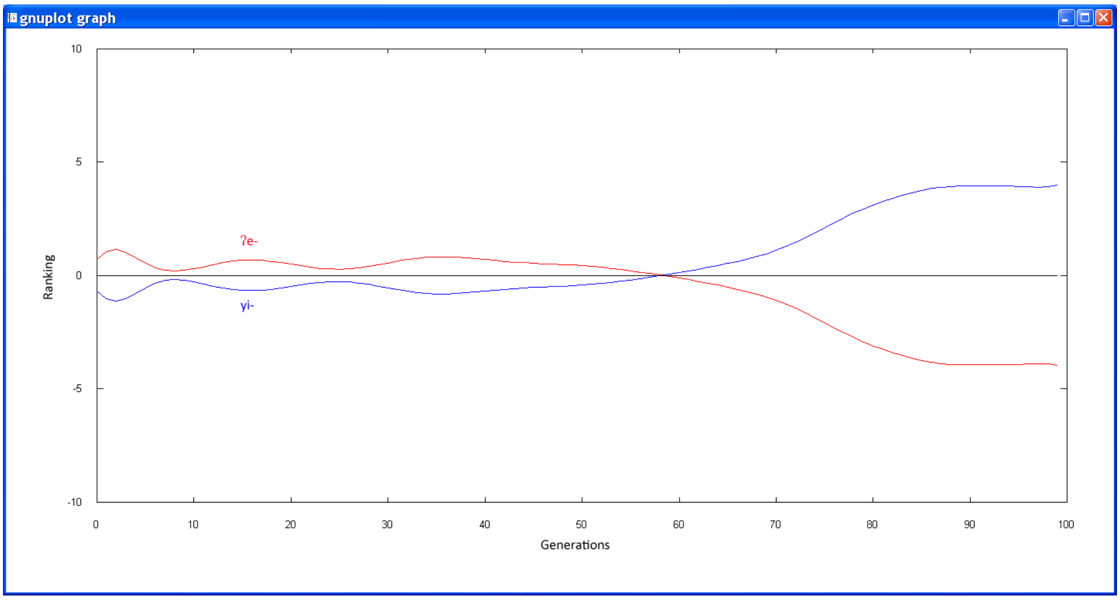


Figure 5: Prefix Change from 1st Person (*ʔe*-) to 3rd Person (*yi*-)

A similar pattern emerges in Figure 6, illustrating the bidirectional change of the B3 regular verbs from the *higdíl* ‘enlarged’ paradigm type to *hegdíl* and in the opposite direction, from the *hextím* ‘signed someone’ paradigm type to *hixtím* (§3.4.2).

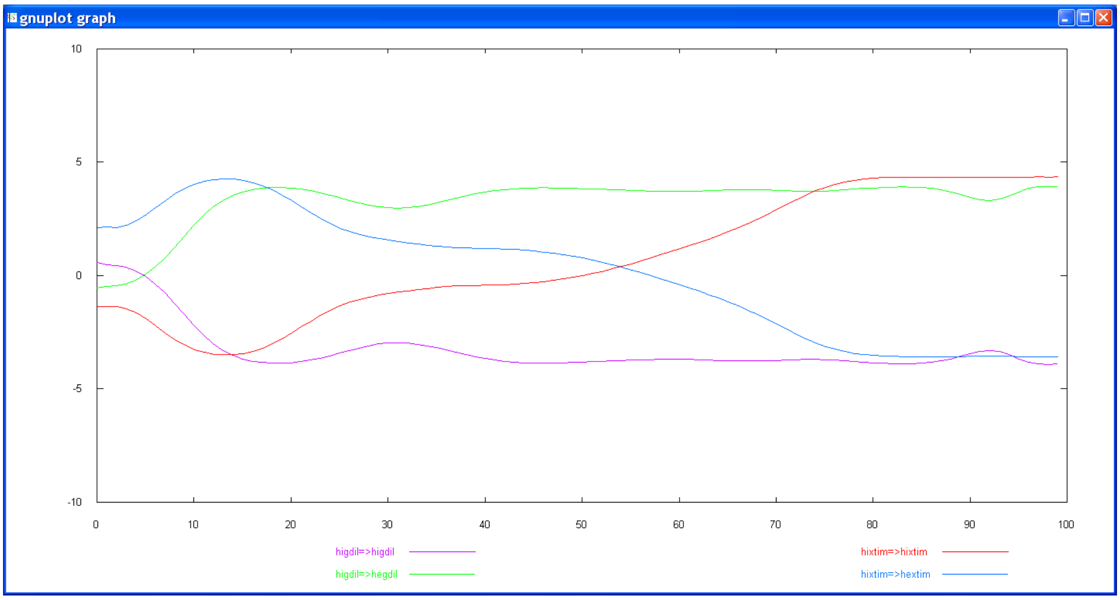


Figure 6: Bidirectional Change in B3

The same end state is predicted for both pairs: at some point, the ranking of the constraints responsible for the emergence of the normative forms (indicated by the purple and blue lines) will reverse and the constraints will grow sufficiently apart such that the probability of occurrence of the current normative forms will be extremely low. Note the difference in the distance between the constraint pairs, i.e. between the green and purple constraints (responsible for the variation between normative *higdíl* and colloquial *hegdíl*) and between the blue and the red constraints (responsible for the variation between normative *hextíim* and colloquial *hixtím*). The distance reflects the frequencies of the colloquial variants, *hegdíl* (37%) and *hixtím* (10%) relative to their normative counterparts, *higdíl* and *hextím* in the recorded database. Where the distance is small, the overlap of the constraints is great, resulting in more variation. Conversely, a greater distance reflects a smaller overlap between the constraints, resulting in less variation.

The degree of overlap (the distance between the overlapping constraints which reflects the degree of variation) reflects the progression of the change. Where the distance is small, the crossing point where the two constraints completely overlap, is predicted to occur sooner than where the distance is greater. This means that the process in one direction is predicted to end sooner than in the opposite direction.

CHAPTER 6. SUMMARY AND CONCLUSION

The Hebrew verb system is changing. This dissertation set out to describe the nature of the change, what triggered it, what its driving forces are, and where it is going.

Hebrew has lost a group of segments known as the gutturals (for most speakers). This loss is not specific to the verb system, but it so happens that this is the group of consonants that determines class membership, according to which we conjugate verbs. In addition, the gutturals are notorious for triggering changes in their surroundings, such as vowel lowering and vowel epenthesis, to break a cluster that would otherwise cause them to be in the coda.

(105) Loss of the gutturals

Phase I		**Phase II**		
Past	**Future**	**Past**	**Future**	
ʕavád	yaʕavód	avád	yaavód	‘work’
ʔasáf	yeʔesóf	asáf	yaasóf	‘collect’
ʔaráx	yeʔeráx	aráx	Ø	‘last’

The loss of the gutturals eliminated important cues for proper conjugation and rendered certain vowel changes opaque (e.g. *ʕavád–yaʕavód* ‘work’ cf. *gadál–yigdál* ‘grow’). In the absence of a guttural, it is no longer clear why epenthesis or lowering from *i* to *a* should occur. Therefore, the conjugation patterns are today distinct templates that do not result from any phonological rule and therefore must be memorised along with the verbs that follow them.

Because the classes are found in all verb structures, i.e. binyanim, to speak ‘normatively’ as prescribed in the grammar books, much memorisation is required. However, most speakers are not bound by normative prescriptivism and so rather than memorising class memberships, they attempt to arrive at new generalisations. In the example in (105), such a generalisation (Phase II) would be that if the past tense form of two paradigms looks the same, so should the future tense forms, resulting in the levelling of the future paradigm.

But this is an oversimplification of the data, as some verbs survive this levelling and some do not. In (105), the *asáf* ‘collect’ paradigm is levelled according to the *avád* ‘work’ paradigm. Both paradigms survive; one through substitution and one remains intact. The paradigm of *aráx* ‘lasted’, on the other hand, does not survive and it loses its future tense. To complicate matters further, while some levelling is to a single surviving pattern, as in (105), some is bidirectional, as in (106). Bidirectionality is more costly as verbs move from both classes to both comparable classes, resulting in much movement, but no reduction in the number of templates.

(106) Symmetrical variation in B5 (the normative form is underlined)

<u>hitmalé</u> ≈ hitmalá	‘was filled’	cf. hitkasá
<u>hitkasá</u> ≈ hitkasé	‘covered himself’	cf. himalé

Change is not instantaneous. It is a long process during which members migrate at a varying pace, at which time both the old ‘normative’ form and the new ‘colloquial’ form coexist. This surface variation is the inevitable consequence of change. The regular alternations as well as the surface variations are described in detail in Chapters 2 and 3. Taking into account that the classes of weak verbs exhibit more alternations normatively, and as a result much more surface variation than regular verbs, combined with the observation that, generally speaking, weak verbs have a much higher token frequency, the picture that is painted is of a chaotic system. However, it is not at all chaotic.

Except for spirantisation, I argue in Chapter 4 that all the variation observed in the verb system is triggered by similarity due to the loss of the gutturals. This loss caused the increased similarity between once distinct paradigms and the increase in similarity triggered the migration among the sub-classes within the binyan. But the migration is not chaotic; verbs do not migrate to any class paradigm, only to the one most similar to them.

To show this, I proposed a model of similarity that enables to compare and to quantify similarity between two full paradigms. Within the proposed model, only the templatic content is compared while ignoring the verb specific stem consonants. Each difference is assigned a dissimilarity unit and the dissimilarity units from all pairs are tallied. The result is the similarity value between the compared paradigm pairs. In OT terms, this can be viewed as a variation of an OP (Optimal Paradigms) constraint that penalises every difference between comparable forms (Kenstowicz, 1996, McCarthy, 2005). The candidate pair with the lowest value, or with the least violation marks is the most similar pair.

This model of similarity can predict which paradigm pair might exhibit variation and which might not. It cannot, however, predict which paradigm will survive and which will become extinct. The measured similarity between any two paradigms A and B means that A is similar to B as B is to A. If they are equally similar, then both paradigms have an equal probability of surviving by virtue of similarity alone. This would explain the bidirectional change (referred to in this study as ‘merger’), but not the frequency of the usage of the variants, which typically is not 50% for each variant.

The surface variation is not always bidirectional. More often than not, it is unidirectional, where one class takes over as the other becomes extinct. Type frequency is responsible for the direction of the change. The paradigm type with the higher type frequency will typically survive and the one with the lower type frequency will be replaced. It stands to reason that this should be the case, because fewer verbs undergo change while still reducing the number of paradigm types. Because the smaller group is the one to change, fewer verbs exhibit surface variation, achieving maximum impact at a minimal cost.

But then how can bidirectional change exist? In the bidirectional migration, the change is to the pattern with the higher type frequency in only one of the directions. In the opposite direction, it is inevitably to the pattern with the lower type frequency. If migration to the class with the lower frequency is possible, what blocks it in the unidirectional change? There must be an additional condition regulating the directionality. If unidirectionality is always to the higher type frequency paradigm, then the only way for bidirectionality to be possible is if the two comparable paradigms had identical type frequencies. The chance of this happening is miniscule, but the ratio between the type frequency of the two paradigms can help predict if migration will be in one direction or in both directions. A small ratio means that the difference in type frequencies is small, such that it is difficult to tell which paradigm is the larger group of verbs and so some speakers may choose one direction and other speakers may choose the other direction. A large ratio means that the paradigm type with the larger number of members is easily identifiable and so speakers will naturally level the smaller group accordingly. Additional research on inter- vs. intra-speaker variation is necessary to test this claim.

The following diagram summarises the flow of change in the verb system.

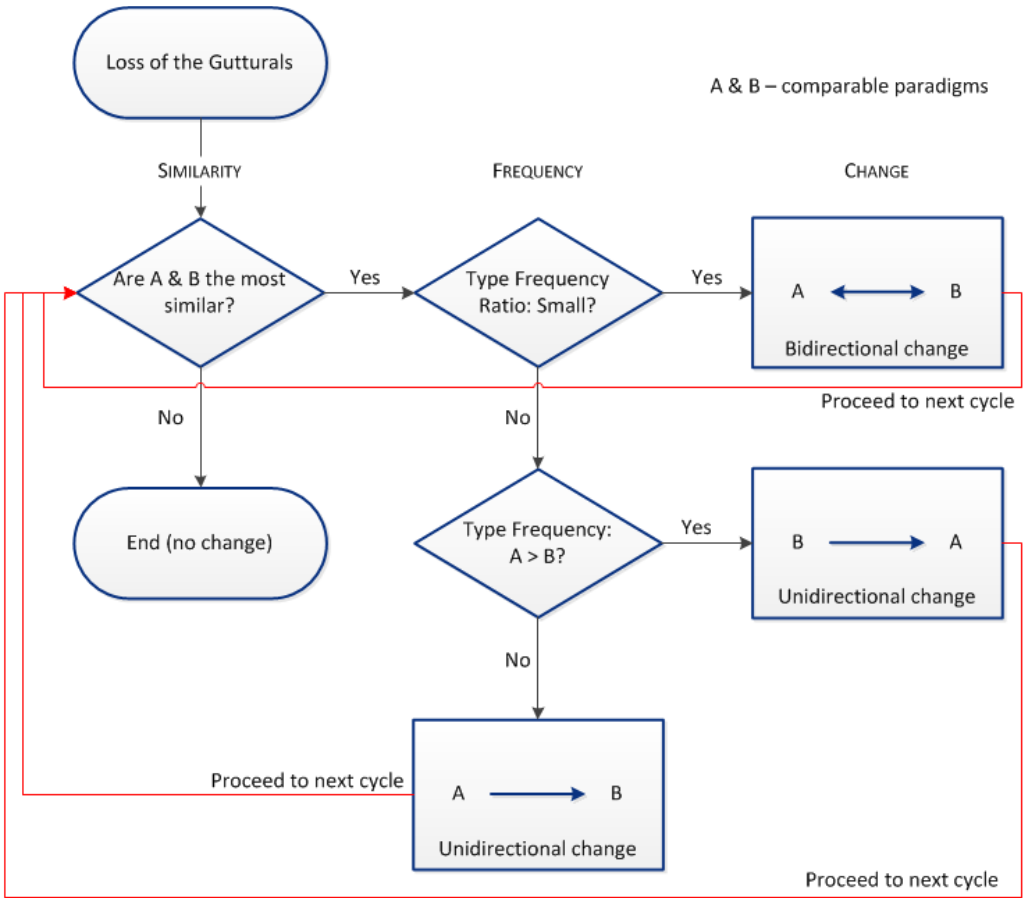


Figure 7: Change Flow in the Hebrew Verb System

As a side remark, bidirectionality is, for the most part, an illusion. From the point of view of the sub-class, its members can migrate to only one sub-class, the most similar one. So bidirectionality is only from the view point of the comparable pair. The only ‘true’ bidirectional migration is found in the B4 class of guttural finals (see §4.3.3 and §4.3.7). This class exhibits a diphthong whenever the historical guttural is word-final, as in *nits**éa**x–nitsáxti* ‘he–I won’. But without a surface guttural word-finally, the surface realisation of the diphthong is opaque (cf. *gidél–gidálti* ‘he–I raised’). Speakers have a choice which member of the diphthong to keep and which to discard: the mid vowel, which is comparable to the glottal-final (thus *bit͡sé–bit͡séti*), or the low vowel, which is the trademark of the pharyngeal-finals (thus *bit͡sá–bit͡sáti*). The latter option (*bit͡sá–bit͡sáti* ‘he–I executed’) has a higher token frequency than *bit͡sé–bit͡séti* ‘he–I executed’ (with an *e*), but it is a new class formation that is not comparable to any other paradigm type, and it is limited to formal speech. *bit͡sé–bit͡séti* ‘he–I executed’ has therefore the potential to increase in token frequency in regular speech as it is similar to an existing class, *milé–miléti* ‘he–I filled’. Time will tell.

At the selection point, the actual variant that surfaces is selected stochastically. I provided a general description of Stochastic OT in §1.3.2 and a more detailed description in §3.4. Stochastic OT reflects the relative token frequency of the variants by sometimes selecting one variant as the optimal output and sometimes the other variant. The relative frequencies of the two variants results from the relative distance of the constraints responsible for their selection. As the constraints move closer together, increasing their overlap, the higher the token frequency becomes. In time, if nothing disturbs the process, the ranking responsible for the newer form takes over and the constraints involved move further apart.

The possible resolution of the observed variation is discussed in Chapter 5, where I applied evolOT (Jäger, 2002b) to simulate the evolution of the change specific to the Hebrew verb system. Based on the analysis described in this study and based on the token frequencies of the variants in the corpus it is fed, evolOT mirrored the predictions that given enough time, the overlapping constraints will grow apart, reversing their order relative to the previous non-variable state. Thus, in unidirectional variation, the new form will substitute the older form, thus reducing the number of sub-classes. In bidirectional variation, the same will occur, but both sub-classes existing today will continue to co-exist.

APPENDIX A. Example of Full Paradigm Comparison

The following is an example of how two complete paradigms are compared. In this example, I compare the B5 *hitlabéʃ* paradigm type to the *hitkoféf* paradigm type. Recall from example (79b) in §4.3.1 that the similarity between these two paradigm types is Δ21. Every paradigm has 16 forms (8 in the past tense and 8 in the future tense).

Person	**Past**	**Future**	
1^{st} sg.	h i t l a b á ʃ t i	- e t l a b é ʃ	
	h i t k o f á f t i	- e t k o f é f	
	1	1	Δ2
2^{nd} ms.sg.	h i t l a b á ʃ t a	- t i t l a b é ʃ	
	h i t k o f á f t a	- t i t k o f é f	
	1	1	Δ2
2^{nd} fem.sg.	h i t l a b á ʃ t	- t i t l a b ʃ í	
	h i t k o f á f t	- t i t k o f e f í	
	1	1 1	Δ3
3^{rd} ms.sg.	h i t l a b é ʃ	- y i t l a b é ʃ	
	h i t k o f é f	- y i t k o f é f	
	1	1	Δ2
3^{rd} fem.sg.	h i t l a b ʃ á	- t i t l a b é ʃ	
	h i t k o f e f á	- t i t k o f é f	
	1 1	1	Δ3
1^{st} pl.	h i t l a b á ʃ n u	- n i t l a b é ʃ	
	h i t k o f á f n u	- n i t k o f é f	
	1	1	Δ2
2^{nd} pl.	h i t l a b á ʃ t e m-	t i t l a b ʃ ú	
	h i t k o f á f t e m-	t i t k o f e f ú	
	1	1 1	Δ3

Person	Past	Future	
3rd pl.	h i t l a b ʃ ú	- y i t l a b ʃ ú	
	h i t k o f e f ú	- y i t k o f e f ú	
	1 1	1 1	Δ4
		Total:	**Δ21**

APPENDIX B. THE DATABASE

The following tables list the paradigm types in each binyan. The past and future forms are provided for each paradigm type in their normative form, with modern pronunciation. Note the following:

- **Types**: The number of verb types counted from Tarmon and Uval (1998). Where the type is empty, it is counted with the paradigm type in the row above (only in the normative).
- **Tokens**: The number of tokens counted from the recorded database (see §1.4 for information on the data sources).
- **Variants**: Deviations from the normative form found either in the recorded database or the sporadic database. A minus (-) sign means that there are no variants for the specific paradigm type. Where the token count is 0, and there is a variant, the variant is from the sporadic database. The error rate for these variants could not be calculated.
- **Error rates** are calculated from the tokens that match the variation criteria. For example, In B1, the *aláx-yeléx* 'go' paradigm type has 72 tokens in the recorded database. The variant *iléx* is expected to surface only in the future tense, so a token in the past tense provides no information on paradigm type. Out of the 72 tokens, only 11 are in the future tense and out of these 5 are variants with an *i*. The error rate is thus 45.45% (5 out of 11).

B1

	Past	Future		Types	Tokens	Variant	Error rate
1.	sagár	yisgór	'close'	226	107	-	0%
2.	ʃaál	yiʃál	'ask'	87	25	-	0%
3.	lamád	yilmád	'learn'	85	85	-	0%
4.	lakáx	yikáx	'take'	1	22	-	0%
5.	kaná	yikné	'buy'	47	164	-	0%
6.	avád	yaavód	'work'	44	90	-	0%
	xazár	yaxazór	'return'		24	yaxzór	100%
7.	xakár	yaxkór	'investigate'	42	38	-	0%
8.	kam	yakúm	'rise'	40	25	-	0%
	af	yaúf	'fly'		3	-	0%
9.	nafál	yipól	'fall'	17	3	-	0%
10.	raá	yiré	'see'	17	101	-	0%
11.	asáf	yeesóf	'collect'	15	3	yaasóf	100%
12.	azál	yeezál	'run out'	2	1	yaazól	100%
13.	nasá	yisá	'travel'	9	10	insá	-

	Past	Future		Types	Tokens	Variant	Error rate
14.	kará	yikrá	‘read’	9	17	-	0%
15.	xasár	yexsár	‘evade’	7	2	ixsár	100%
16.	yanák	yinák	‘suckle’	7	0	inók	-
17.	asá	yaasé	‘do’	7	148	-	0%
18.	xalá	yexelé	‘become ill’	6	0	yaxlé	-
19.	axál	yoxál	‘eat’	5	102	-	0%
	aáv	yoáv	‘love’		40	-	0%
20.	samáx	yismáx	‘rejoice’	5	38	-	0%
21.	yaʃáv	yeʃév	‘sit’	4	19	iʃév	83.33%
	aláx	yeléx	‘go’		72	iléx	45.45%
22.	t͡samé	yit͡smá	‘thirst’	4	0	-	-
23.	nax	yanúax	‘rest’	4	0	-	-
24.	nazál	yizál	‘drip’	3	0	inzól	-
25.	sam	yasím	‘put’	3	28	-	0%
26.	natá	yité	‘lean’	2	0	-	0%
27.	agá	yeegé	‘pronounce’	2	0	-	0%
28.	ʃamán	yiʃmán	‘become fat’	2	0	-	0%
29.	xatá	yexetá	‘sin’	1	0	-	0%
30.	xarád	yexerád	‘fear’	1	0	-	0%
31.	yaʃén	yiʃán	‘sleep’	1	0	-	0%
32.	yat͡sá	yet͡sé	‘exit’	1	47	itsé	70%
33.	yare	yirá	‘fear’	1	0	-	0%
34.	yadá	yedá	‘know’	1	87	idá	42.86%
35.	yagá	yigá	‘toil’	1	0	-	0%
36.	afá	yofé	‘bake’	1	0	-	0%
37.	ayá	yiyé	‘be’	1	243	-	0%
38.	boʃ	yevóʃ	‘feel ashamed’	1	0	-	-
39.	ba	yavó	‘come’	1	80	-	0%
40.	yaxól	yuxál	‘be able to’	1	1	yexól	100%
41.	katón	yiktán	‘decrease’	1	0	-	-
42.	natán	yitén	‘give’	1	29	yetén	-
43.	met	yamút	‘die’	1	13	-	0%
44.	nigáʃ	yigáʃ	‘approach’	1	0	-	0%
45.	xanán	yaxón	‘pardon’	1	0	-	0%

B2

	Past	Future		Types	Tokens	Variant	Error rate
1.	nixnás	yikanés	‘enter’	120	59	-	0%
2.	nidxá	yidaxé	‘be postponed’	1	0	-	-
3.	nit͡sál	yinat͡sél	‘be saved’	7	0	-	-
4.	nisá	yinasé	‘marry’	2	0	-	-
5.	nirdám	yeradém	‘fall asleep’	13	0	yiradém	-
6.	nirá	yeraé	‘be seen’	1	56	yiraé	100%
7.	nivál	yibaél	‘be frightened’	35	26	-	0%
8.	nikrá	yikaré	‘be read’	7	10	-	0%
9.	nivná	yibané	‘be built’	17	1	-	0%
10.	neet͡sár	yeat͡sér	‘be stopped’	40	0	-	-
11.	nexkár	yexakér	‘be investigated’	26	2	nixkár	100%
12.	neená	yehané	‘enjoy’	3	7	yeené	100%
13.	naasá	yeasé	‘be done’	2	0	neesá	-
14.	nirgá	yeragá	‘calm down’	3	6	yiragá	60%
15.	niʃmá	yiʃamá	‘be heard’	30	5	-	0%
16.	nolád	yivaléd	‘be born’	5	4	-	0%
17.	noád	yivaéd	‘destined’	3	2	-	0%
18.	nodá	yivadá	‘be known’	3	0	-	-
19.	norá	yiyaré	‘be shot’	1	0	-	-
20.	nidón	yidón	‘be discussed’	4	0	-	-
21.	nasóg	yisóg	‘retreat’	4	0	-	-
22.	namás	yimás	‘be melted’	1	0	-	-

B3

	Past	Future		Types	Tokens	Variant	Error rate
1.	idlík	yadlík	‘light up’	237	194	edlík	34.48%
2.	irvíax	yarvíax	‘earn’	48	31	ervíax	16.66%
3.	igbía	yagbía	‘elevate’	1	31	egbía	20%
4.	imt͡sí	yamt͡sí	‘invent’	10	1	emt͡sí	0%
5.	ikír	yakír	‘recognise’	26	109	ekír	29.17%
6.	isí	yasí	‘marry’	1	0	esí	-
7.	igía	yagía	‘arrive’	9	41	egía	25.93%
8.	eexíl	yaaxíl	‘feed’	48	36	ixíl	0%

	Past	Future		Types	Tokens	Variant	Error rate
	exelíf	yaxalíf	‘change’		10	ixlíf	0%
9.	exlít	yaxlít	‘decide’	17	7	ixlít	50%
10.	exetí	yaxatí	‘miss’	2	0	ixtí	-
11.	exbí	yaxbí	‘hide’	1	0	ixbí	-
12.	ekím	yakím	‘establish’	43	64	ikím	22.92%
13.	eríax	yaríax	‘smell’	7	2	iríax	0%
14.	eví	yaví	‘bring’	3	20	iví	10%
15.	emit	yamít	‘kill’	2	0	-	-
16.	irʃá	yarʃé	‘allow’	29	3	erʃá	0%
17.	iká	yaké	‘hit’	3	0	eká	-
18.	erá	yare	‘show’	2	6	irá	0%
19.	exená	yaxané	‘park’	4	1	ixná	0%
20.	oríd	yoríd	‘lower’	15	15	-	0%
21.	ot͡sí	yot͡sí	‘bring out’	1	8	-	0%
22.	odía	yodía	‘announce’	5	2	-	0%
23.	odá	yodé	‘thank’	3	1	-	0%
24.	ekél	yakél	‘ease’	9	0	-	-
25.	eréa	yaréa	‘harm’	1	0	-	-

B4

	Past	Future		Types	Tokens	Variant	Error rate
1.	dibér	yedabér	‘talk’	654	232	-	0%
2.	nit͡séax	yenat͡séax	‘win’	64	8	nit͡sáx	12.5%
	bit͡séa	yevat͡séa	‘execute’		6	bit͡séti	-
3.	milé	yemalé	‘fill’	12	2	milá	-
4.	nisá	yenasé	‘attempt’	51	69	nisé	6.66%
5.	exér	yeaxér	‘be late’	38	5	ixér	20%
6.	eréax	yearéax	‘host’	1	0	iréax	-
7.	sovév	yesovév	‘turn’	38	7	sivév	0%
8.	soxéax	yesoxéax	‘converse’	1	0	soxáx	-
9.	gerá	yegaré	‘tease’	1	1	girá	100%
10.	erá	yeerá	‘occur’	1	0	-	-

B5

	Past	Future		Types	Tokens	Variant	Error rate
1.	itlabéʃ	yitlabéʃ	‘dress up’	369	151	-	0%
2.	itgaléax	yitgaléax	‘shave’	46	13	-	0%
3.	iʃtaamém	yiʃtaamém	‘be bored’	2	2	-	0%
4.	itpalé	yitpalé	‘wonder’	8	3	itpalá	100%
5.	itkoféf	yitkoféf	‘bend’	47	30	-	0%
6.	itnoéa	yitnoéa	‘sway’	2	0	-	0%
7.	itkasá	yitkasé	‘cover oneself’	41	8	itkasé	66.66%

APPENDIX C.Z-TABLE

z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0..6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	00.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

REFERENCES

Abney, Steven. 1996. "Statistical Methods and Linguistics." In J. Klavans and P. Resnik (eds.), *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. Cambridge, MA: MIT Press. 1-26.

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

Albright, Adam. 2002a. *The Identification of Bases in Morphological Paradigms*. Ph.D Dissertation, University of California, Los Angeles.

Albright, Adam. 2002b. Base Selection in Analogical Change in Yiddish. *Proceedings of the Proceedings of the 28th annual meeting of the berkeley linguistics society (BLS 28),* 1-13.

Albright, Adam. 2006a. "Base-driven Leveling in Yiddish Verb Paradigms." *Natural Language & Linguistic Theory* 28:475-537.

Albright, Adam. 2006b. Lexical and Morphological Conditioning of Paradigm Gaps. *Proceedings of the Modelling Ungrammaticality in OT workshop,* Oslo, 22-23.

Albright, Adam. 2008a. "Inflectional Paradigms Have Bases Too: Arguments from Yiddish." In A. Bachrach and A. Nevins (eds.), *The Bases of Inflectional Identity*. Oxford: Oxford University Press.

Albright, Adam. 2008b. "Explaining Universal Tendencies and Language Particulars in Analogical Change." *Linguistic Universals and Language Change*:144-184.

Anderson, Stephen R. 1985. *Phonology in the Twentieth Century: Theories of Rules and Theories of Representations*. Chicago: The University of Chicago Press.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Antilla, Raimo. 1977. *Analogy*. The Hague Mouton

Anttila, Arto. 1997a. "Deriving Variation from Grammar." In F. Hinskens, R. Van Hout and L. Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam: Benjamins. 35-68.

Anttila, Arto. 1997b. *Variation in Finnish Phonology and Morphology*. Ph.D Dissertation, Stanford University, Stanford, CA.

Anttila, Arto, and Young-mee Yu Cho. 1998. "Variation and Change in Optimality Theory." *Lingua* 104:31-56.

Anttila, Arto. 2002a. "Morphologically Conditioned Phonological Alternations." *Natural Language and Linguistic Theory* 20:1-42.

Anttila, Arto. 2002b. "Variation and phonological theory." *The handbook of language variation and change*:206-243.

Anttila, Arto. 2006. "Variation and Opacity." *Natural Language and Linguistic Theory* 24:893-944.

Anttila, Arto. 2007. "Variation and Optionality." In P. de Lacy (ed.), *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press. 519-536.

Anttila, Raimo. 1972. *An introduction to historical and comparative linguistics*. New York: Macmillan.

Aronoff, Mark. 1994. *Morphology by Itself. Stems and inflectional classes: Binyanim, as Inflectional Classes*. Cambridge, MA: MIT Press.

Bailey, Charles-James. 1973. *Variation and Linguistic Theory.* Washington: Center for Applied Linguistics.

Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace. 1996. "Magnitude Estimation of Linguistic Acceptability." *Language* 72(1):32-68.

Barkai, M. 1975. “On phonological representations, rules, and opacity.” *Lingua* 37:363-376.

Bat-El, Outi. 1994. “Stem Modification and Cluster Transfer in Modern Hebrew.” *Natural Language and Linguistic Theory* 12:571-596.

Bat-El, Outi. 2002a. “Semitic verb structure within a universal perspective.” In J. Shimron (ed.), *Language Processing and Acquisition in Languages of Semitic, Root-based, Morphology*. Amsterdam: John Benjamins. 29-59.

Bat-El, Outi. 2002b. “True Truncation in Colloquial Hebrew Imperatives.”. *Language* 78:651-683.

Bat-El, Outi. 2003. “The Fate of the Consonantal Root and the Binyan in Optimality Theory.” *Recherches Linguistiques de Vincennes* 32:31-60.

Bat-El, Outi. 2005. “Competing Principles of Paradigm Uniformity: Evidence from Hebrew Imperative Paradigm.” In L. J. Downing, T. A. Hall and R. Raffelsiefen (eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press. 44-64.

Bat-El, Outi. 2008. “Morphologically Conditioned V-0 Alternations in Hebrew: Distinction among Nouns, Adjectives & Participles, and Verbs.” In S. Armon Lotem, G. Danon and S. Rothstein (eds.), *Current Issues in Generative Hebrew Linguistics*. Amsterdam: John Benjamins. 27-60.

Bat-El, Outi. 2011. “Semitic Templates.” In M. van Oostendrop, C. Ewen, E. Hume and K. Rice (eds.), *The Blackwell Companion to Phonology*. Malden, MA: Wiley-Blackwell. 2586-2608.

Benua, Laura. 1995. “Identity Effects in Morphological Truncation.” In J. Beckman, S. Urbanczyk and L. Walsh (eds.), *Papers in Optimality Theory: University of Massachusetts Occasional Papers in Linguistics*. Amherst, Massachusetts: GLSA. 77-136.

Benua, Laura. 1997. *Transderivational Identity: Phonological Relations Between Words*. Ph.D., University of Massachusetts Amherst.

Berman, Ruth. 1978. *Modern Hebrew Structure*. Tel-Aviv: University Publishing Projects.

Berman, Ruth. 1981a. “Language development and language knowledge: Evidence from the acquisition of Hebrew morphophonology.” *Journal of Child Language* 8:609-626.

Berman, Ruth. 1981b. “Regularity vs. anomaly: The acquisition of Hebrew inflectional morphology.” *Journal of Child Language* 8:265-282.

Bloomfield, Leonard. 1933. *Language*. London: Unwin.

Bod, Rens, Jennifer Hay, and Stefanie Jannedy. 2003. *Probabilistic Linguistics*. Cambridge: MIT Press.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives* Doctoral Dissertation, University of Amsterdam, Amsterdam.

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

Bolozky, Shmuel. 1978a. “Word Formation Strategies in the Hebrew Verb System: Denominative Verbs.” *Afroasiatic Linguistics* 5:1-26.

Bolozky, Shmuel. 1978b. “Some Aspects of Modern Hebrew Phonology.” In R. Berman (ed.), *Modern Hebrew Structure*. Tel-Aviv: University Publishing Projects.

Bolozky, Shmuel. 1979. “On the New Imperative in Colloquial Hebrew.” *Hebrew Annual Review* 3:17-24.

Bolozky, Shmuel. 1980a. “On the Autonomy of the Tense Paradigm: The Modern Hebrew Evidence.” *Lingua* 51:115-123.

Bolozky, Shmuel. 1980b. "Paradigm Coherence: Evidence from Modern Hebrew." *Monographic Journals of the Near East: Afroasiatic Linguistics* 7:103-126.

Bolozky, Shmuel. 1996. *501 Hebrew Verbs: fully conjugated in all the tenses in a new easy-to-follow format alphabetically arranged by root*. Barrons Educational Series Incorporated.

Bolozky, Shmuel. 1997. "Israeli Hebrew Phonology." In A. S. Kaye (ed.), *Phonologies of Asia and Africa*. Winona Lake, Ind.: Eisenbrauns. 287-311.

Bolozky, Shmuel. 1999. "On the Special Status of the Vowels a and e in Israeli Hebrew." *Hebrew Studies* 40:233.

Bolozky, Shmuel. 2003a. "Phonological and Morphological Variation in Spoken Hebrew." In B. H. Hary (ed.), *Corpus Linguistics and Modern Hebrew: Towards the Compilation of the Corpus of Spoken Hebrew (CoSIH)*. Tel Aviv: Tel Aviv University Press. 119-156.

Bolozky, Shmuel. 2003b. "Tofaot lešoniyot tiviyot, hamešutafot laivrit hayisre'elit hameduberet velaivrit hamikra'it (Natural linguistic phenomena found in both colloquial Israeli Hebrew and Biblical Hebrew)." *Hadoar* 82:30-36.

Bolozky, Shmuel, and Rina Kreitman. 2007. Uvulars in Modern 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, 2007.

Burzio, Luigi. 1998. "Multiple Correspondence." *Lingua* 104:79-109.

Bybee, Joan. 1984. *Morphology: A Study of the Relation between Meaning and Form*. Philadelphia: Benjamins.

Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Philadelphia: John Benjamins.

Bybee, Joan L., Revere Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Bybee, Joan L. 2001. *Phonology and Language Use*. Cambridge Studies in Linguistics 94. Cambridge UK: Cambridge University Press.

Bybee, Joan L. 2007. *Frequency of Use and the Organization of Language*. New York: Oxford University Press.

Cohen, Evan-Gary. 2009. *The Role of Similarity in Phonology: Evidence from Loanword Adaptation in Hebrew*. Dissertation, Tel-Aviv University, Tel-Aviv.

Culy, Christopher 1998. "Statistical Distribution and the Grammatical/Ungrammatical Distinction." *Grammars* 1(1):1-19.

Daugherty, Kim.G, and Mark. S Seidenberg. 1994. *Beyond Rules and Exception: A Connectionist Modeling Approach to Inflectional Morhplogy. The Reality of Linguistic Rules*. Amsterdam: John Benjamins.

Downing, Laura J., T.A. Hall, and Renate Raffelsiefen. 2005. "The Role of Paradigms in Phonological Theory." In L. J. Downing, T. A. Hall and R. Raffelsiefen (eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press. 1-16.

Even-Shoshan, Avraham. 1982. *hamilón hexadásh (The New Dictionary)*. 14th edition. Jerusalem: Kiryat Sefer. (In Hebrew)

Faust, Noam. 2005. *The Fate of Hebrew Gutturals*. MA, Tel-Aviv University, Tel-Aviv.

Ferguson, Charles A., and Carol B. Farwell. 1975. "Words and Sounds in Early Language Acquisition." *Language* 51:419-439.

Gesenius, Wilhelm. 1910. *Gesenius' Hebrew Grammar*. Oxford: Clarendon (2nd English edition by K. Kautzsch, revised in accordance with the 26th German edition by A.E. Cowley. 1st German edition: 1813).

Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.

Greenberg, Joseph Harold ed. 1966. *Universals of Language*. Cambridge, MA: MIT Press.

Haiman, John. 1994. "Ritualization and the Development of Language." In W. Pagliuca (ed.), *Perspectives on Grammaticalization*. Amsterdam: Benjamins. 3-28.

Hale, Mark, and Charles Reiss. 1998. "Formal and Empirical Arguments Concerning Phonological Acquisition." *Linguistic Inquiry* 29:656-683.

Hare, Mary, and Jeffrey L. Elman. 1995. "Learning and Morphological Change." *Cognition* 56:61-98.

Hayes, Bruce, and Margaret MacEachern. 1998. "Quatrain Form in English Folk Verse." *Language* 74(3):473-507.

Hayes, Bruce. 2000. "Gradient well-formedness in Optimality Theory." In J. Dekkers, F. van der Leeuw and J. van de Weijer (eds.), *Optimality Theory: Phonology, syntax, and acquisition*. Oxford: Oxford University Press. 88-120.

Hayes, Bruce. 2004. "Phonological Acquisition in Optimality Theory: The Early Stages." In R. Kager, J. Pater and W. Zonneveld (eds.), *Fixing Priority: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press. 158-203.

Hyman, Larry. 1970. "The Role of Borrowing in the Justification of Phonological Grammars." *Studies in African Linguistics* 1:1-48.

Jäger, Gerhard. 2002a. evolOT: Software for Simulating Language Evolution Using Stochastic Optimality Theory User's Manual.

Jäger, Gerhard. 2002b. "Learning Constraint Sub-hierarchies. The Bidirectional Gradual Learning Algorithm." *ROA-544 http://roa.rutgers.edu/*.

Keller, Frank. 2000. *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. Ph.D Thesis, University of Edinburgh.

Keller, Frank, and Theodora Alexopoulou. 2001. "Phonology Competes with Syntax: Experimental Evidence for the Interaction of Word Order and Accent Placement in the Realization of Information Structure." *Cognition* 79:301-372.

Kenstowicz, Michael. 1996. "Base-Identity and Uniform Exponence: Alternatives to Cyclicity." In J. Durand and B. Laks (eds.), *Current Trends in Phonology: Models and Methods*: University of Salford Publications.

Kenstowicz, Michael. 2007. "Salience and similarity in loanword adaptation: a case study from Fijian." *Language Sciences* 29:316-340.

Kiparsky, Paul. 1968. "Linguistic universals and linguistic change." In E. B. a. R. Harms (ed.), *Universals in Linguistic Theory*. New York: Holt, Reinhart & Winston. 171-202.

Kiparsky, Paul. 1982. *Explanation in Phonology*. Dordrecht: Foris.

Kiparsky, Paul. 1988. "Phonological Change." In F. Newmeyer (ed.), *Linguistics: The Cambridge Survey*. Cambridge: Cambridge University Press.

Kiparsky, Paul. 1993. Variable rules. Handout of presentation at Rutgers Optimality Workshop.

Kiparsky, Paul. 1995. "The Phonological Bases of Sound Change." In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Oxford: Blackwell. 640-670.

Kiparsky, Paul. 2000. "Analogy as Optimization: 'Exceptions' to Sievers' Law in Gothic." In A. Lahiri (ed.), *Analogy, Levelling, Markedness: Principles of Change in Phonology and Morphology*. New York: Mouton de Gruyter.

Kroch, Anthony S. 1989. "Reflexes of Grammar in Patterns of Language Change." *Language Variation and Change* 1:199-244.

Kuryłowicz, Jerzy. 1949. "La nature des proces dits 'analogiques'." *Acta Linguistica* 5:15-37. Reprinted in E.P. Hamp, EW. Householder, and R. Austerlitz (eds.), (1966), *Readings in Linguistics II*, 157-174. Chicago: University of Chicago Press. .

Lahiri, Aditi. 2000. *Analogy, Levelling, Markedness: Principles of Change in Phonology and Morphology*. New York: Mouton de Gruyter.

Laks, Lior. 2006. *The Morphology-Lexicon and Morphology-Syntax Interfaces: Thematic Operations in Semitic Languages* MA Thesis, Tel-Aviv University, Tel-Aviv.

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

Laufer, Asher. 1990. "Hebrew." *Journal of the International Phonetic Association* 20:40-43.

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang, and Martin A. Nowak. 2007. "Quantifying the evolutionary dynamics of language." *Nature* 449:713-716.

Lindblom, Björn. 1992. "Phonological Units as Adaptive Emergents of Lexical Development." In C. A. Ferguson, L. M. Menn and C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press. 131-163.

Lowenstamm, Jean, and Jonathan Kaye. 1986. "Compensatory lengthening in Tiberian Hebrew." In L. Wetzels and E. Sezer (eds.), *Studies in Compensatory Lengthening*. Dordrecht: Foris. 97-132.

Mańczak, W. 1958. "Tendences générale des changements analogiques." *Lingua* 7:298-325, 387-420.

McCarthy, John J. 1981. "A prosodic theory of nonconcatenative morphology." *Linguistic Inquiry* 12:373-418.

McCarthy, John J. 1985. *Formal problems in Semitic phonology and morphology*. New York: Garland.

McCarthy, John J., and Alan S. Prince. 1993. "Generalized Alignment." *Yearbook of Morphology*.

McCarthy, John J. 1994. "The phonetics and phonology of Semitic pharyngeals." In P. Keating (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*. Cambridge: Cambridge University Press. 191-233.

McCarthy, John J., and Alan S. Prince. 1995. "Faithfulness and Reduplicative Identity *Papers in Optimality Theory: University of Massachussetts Occasional Papers*. Amherst Mass.: Graduate Linguistic Student Association. 249-384.

McCarthy, John J. 2005. "Optimal Paradigms." In L. J. Downing, T.A. Hall, and Renate Raffelsiefen (ed.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press. 170-207.

Most, Tova, Ofer Amir, and Yishai Tobin. 2000. "The Hebrew Vowel System: Raw and Normalized Acoustic Data." *Language and Speech* 43:295.

Newmeyer, Frederick. 2003. "Grammar Is Grammar and Usage Is Usage." Language 79:682-707.

Pariente, Itsik. 2006. *Pharyngeal Effects in Sephardic Modern Hebrew*. MA Thesis, Tel Aviv University.

Pariente, Itsik. 2010. "Pharyngeal Related Non-lexical Vowels in Sephardic Modern Hebrew." *Linguistics in Amsterdam* 3:1-19.

http://dpc.uba.uva.nl/aclc/03/nr01/a01

Pater, Joe. 2004. "From Phonological Typology to the Development of Receptive and Productive Phonological Competence: Applications of Minimal Violation." In

R. Kager, J. Pater and W. Zonneveld (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press.

Paul, Hermann. 1891. *Prinzipien der Sprachgeschichte*. Tubingen: Niemeyer. (English trans. 1891. Principles of the History of Language. London: Longmans, Green & Co.).

Prince, Alan S., and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. [Also Report RuCCS-TR-2 Rutgers University Cognitive Science Center ROA 537, http://roa.rutgers.edu].

Pulleyblank, Douglas, and William J. Turkel. 2000. "Learning Phonology: Genetic Algorithms and Yoruba Tongue-root Harmony." In J. Dekkers, F. van der Leeuw and J. van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*. Oxford: Oxford University Press. 554-591.

Rabin, Haim. 1972. *History of the Hebrew Language*. Jerusalem: World Zionist Association. [in Hebrew].

Raffelsiefen, Renate. 1995. *Conditions for Stability: The Case of Schwa in German*. Düsseldorf: Arbeiten des Sonderforschungsbereich 282.

Raffelsiefen, Renate. 2004. "Absolute ill-formedness and other morphophonological effects." *Phonology* 21:91-142.

Ravid, Dorit. 1995. *Language Change in Child and Adult Hebrew: A Psycholinguistic Perspective*. New York: Oxford University Press.

Rebrus, Péter, and Miklós Törkenczy. 2005. "Uniformity and Contrast in the Hungarian Verbal Paradigm." In L. J. Downing, T. A. Hall and R. Raffelsiefen (eds.), *Paradigms in Phonological Theory*. Oxford: Oxford University Press. 263-295.

Reiss, Charles. 1997. Explaining Analogy. In *ROA-199, http://roa.rutgers.edu*.

Reynolds, William T. 1994. *Variation and Phonological Theory*. Ph.D Dissertation, University of Pennsylvania, Philadelphia.

Rice, Curt. 2006. "Optimal gaps in optimal paradigms." *Catalan journal of Linguistics* 4:155-170.

Rice, Curt. 2007. "Gaps and repairs at the phonology-morphology interface." *Journal of linguistics* 43:197.

Rosén, Haim. 1973. "Mivne ha'ivrit haisra'elit (The Structue of Israeli Hebrew)." In U. Ornan (ed.), *Mikraa le-Torat ha-Hege*. Jerusalem: Hebrew University. [in Hebrew] 253-138.

Rumelhart, David E., and James. L McClelland. 1986. "On Learning the Past Tenses of English Verbs: Implicit Rules or Prallel Distributed Processing?" In J. L. McClelland, D. E. Rumelhart and t. P. R. Group (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press. 216-271.

Sapir, E. 1921. *Language*. New York: Harourt, Brace, and World Inc.

Saussure, Ferdinand de. 1916, 1959. *Course in Modern Linguistics*. New York: The Philosophical Library, Inc. Course de linguistique générale. W. Baskin.

Schwartzwald, Ora Rodrigue. 1985. "Mivtao šel ha-dor ha-cair be-Israel (The Pronunciation of the Young Generation in Israel)." *ševet vaam* 10:[in Hebrew] 66-75.

Schwarzwald, O.R. 1973. Lexical Representations, Phonological Processes, and Morphological Patterns in Hebrew: By Ora Rodrigue Schwarzwald.

Schwarzwald, Ora. 2001. *Modern Hebrew*. vol. 127: Languages of the World/Materials. Munich: LINCOM Europa.

Shin-ichiro, Sano. 2009. *The Role of Internal and External Factors and the Mechanism of Analogical Levelling: Variationist and Probabilistic OT*

Approach to Ongoing Language Change in Japanese Voice System. MA Thesis, Sophia University, Tokyo.

Steriade, Donca. 1999. Lexical Conservatism in French Adjectival Liaison. In M. Authier, B. Bullock and L. Reed (eds.) *Proceedings of the 25th Linguistic Colloquium on Romance Languages* Amsterdam: John Benjamin's. 243-270.

Steriade, Donca. 2001a. "The phonology of perceptibility effects: the P-map and its consequences for constraint organization." *Ms., UCLA*. http://www.linguistics.ucla.edu/people/steriade/steriade.htm

Steriade, Donca. 2001b. "Directional asymmetries in place assimilation." *The role of speech perception in phonology*: 219-250.

http://www.linguistics.ucla.edu/people/steriade/steriade.htm

Sumner, Meghan Marie. 2002. The Psycholinguistic Reality of Abstract Representations. *Proceedings of the West Coast Conference on Formal Linguistics*, Somerville, MA: Cascadilla Press, 429-442.

Sumner, Meghan Marie. 2003. *Testing the Abstractness of Phonological Representations in Modern Hebrew Weak Verbs*, Stony Brook University.

Tarmon, Asher, and Ezri Uval. 1998. Hebrew Verb Tables. Jerusalem: Tamir Publishers.

Tesar, Bruce, and Paul Smolensky. 1996. *Learnability in Optimality Theory (Long Version)*. In Technical Report JHU-CogSci-96-4, Department of Cognitive Science, Johns Hopkins University, Baltimore.

Tesar, Bruce, and Paul Smolensky. 1998. "Learnability in Optimality Theory." *Linguistic Inquiry* 29(2):229-268.

Tesar, Bruce, and Paul Smolensky. 2001. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Trubetzkoy, Nicolai S. 1939. *Grundzüge der Phonologie. Travaux du Cercle le Linguistique de Prague 7*.

Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84:327.

Ussishkin, Adam. 1999. "The Inadequacy of the Consonantal Root: Modern Hebrew Denominal Verbs and Output-Output Correspondence." *Phonology* 16:401-442.

van Marle, Jaap. 1985. *On the Paradigmatic Dimension of Morphological Creativity*. vol. 18. Dordrecht: Foris.

van Oostendorp, Marc. 2004. Variation in Phonology Theory and Evidence. *CASTL*, Tromsø. October 4-8, 2004.

Wang, William. 1969. "Competing Changes as a Cause of Residue." *Language* 45:9-25.

Weinreich, Uriel., William. Labov, and Marvin I. Herzog. 1968. *Empirical Foundations for a Theory of Language Change*. University of Texas Press.

Wexler, Paul. 1990. *The Schizoid Nature of Modern Hebrew: A Slavic Language in Search of a Semitic Past*. Wiesbaden: Otto Harrassowitz.

Wurzel, Wolfgang Ullrich. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer Academic.

Zuckermann, Ghil'ad. 2008. *Israelit Safa Yafa (Israeli, a Beautiful Language: Hebrew)*. Tel-Aviv: Am Oved.

Zwicky, Arnold. 1976. Well, This Rock and Roll Has Got to Stop, Junior's Head is Hard as a Rock. *Proceedings of the Proceedings of Chicago Linguistic Society 12*, University of Chicago, 676-697.

אוניברסיטת תל-אביב
הפקולטה למדעי הרוח ע״ש לסטר וסאלי אנטין
בית הספר למדעי התרבות ע״ש שירלי ולסלי פורטר

דמיון, שוֹנוּת ושינוי:
אי-יציבות בפעלים בעברית

חיבור לשם קבלת התואר ״דוקטור לפילוסופיה״

מאת

גילה צדוק

הוגש לסנאט של אוניברסיטת תל-אביב

דצמבר, 2012

עבודה זו נעשתה בהדרכת

פרופ' אותי בת-אל

תוכן העניינים

נספחים

תקציר

מערכת הפועל של העברית המודרנית עוברת שינוי. עדות לכך היא ריבוי השׁוֹנוּת (variation) הקיימת במערכת, במיוחד בפעלים "העלולים" (weak verbs). פעלים אלו מכילים מספר רב של סוגי פרדיגמה (אלו הן הגזרות, sub-classes) המוגדרים על פי הסגמנט החלש ומיקומו בגזע הפועל (stem). מידת השונות שאנו עדים לה גורמת למערכת הפועל כולה להראות כאוטית. עבודה זו יוצאת לחקור את טיבו של השינוי, מה עורר אותו, מה הם הכוחות המניעים שלו, ולאן הוא מוביל.

1. הגורמים לשינוי

מערכת הפועל מכילה חמישה בניינים הקובעים את צורתו הכללית של הפועל (ובכלל זה המוספיות והתנועות). כל בניין מכיל סוגים שונים של פרדיגמה, כך שסך כל הפרדיגמות הייחודיות הינו גבוה. האופי הכאוטי לכאורה של הפעלים העלולים, המתבטא בשונות הרבה, נובע בעיקר מהשינוי ההסטורי במצאי הסגמנטלי של השפה, כלומר באובדנם של העיצורים הגרוניים, *א,ע,ה*, ו-*ח* (*ʔ, ʕ, h, ħ*). השפעתו של שינוי זה במצאי העיצורים הינה כפולה:

א. הצורך בהבחנה בין פרדיגמות אבדה: אובדנם של העיצורים הגרוניים, הידועים בהשפעתם על התנועות שסביבם, גרם למידה רבה של עמימות (opacity) ביחס להמצאותן של תנועות מסויימות על פני השטח. השוו, למשל, את הזוג המינימלי כמעט *maxár* (מכר) ו-*maxák* (מחק). זהות כמעט מוחלטת בין פעלים כמו אלו בדרך כלל מרמזת על פרדיגמות נטייתיות זהות. זהו לא המקרה עם שני הפעלים הללו, שנבדלים במספר ההברות של צורות העתיד: *maxrú* (מכרו) לעומת *maxakú* (מחקו). הסיבה לשוני זה נעוצה במקורו של העיצור *x*. ה-*x* ב-*maxák* (מחק), מקורו בעיצור הגרוני *ח*, אשר התמזג עם *x*. לעומתו, מקורו של העיצור *x* ב-maxár (מכר) הינו ב-*כ*. כלומר, ההבחנה בין הפרדיגמות נגזרה בעבר ממוטיבציה פונולוגית: ה-*ח* הגרונית דרשה אחריה תנועה נמוכה, ואילו ל-*כ* שאינה גרונית לא היתה דרישה כזו. בעקבות מיזוגם של ה-*ח* וה-*כ*, המצאותה של התנועה הנמוכה אחרי ה-*x* ב-*maxakú* (מחקו) הפכה לעמומה.

ב. העדויות הנדרשות להבחנה בין פרדיגמות אבדו. האובדן של העיצורים הגרוניים מקשה מאוד על ההבחנה בין פרדיגמות בהן העיצור הגרוני הוא המבחין העיקרי. השווה את שלושת הפעלים בצורתם הנורמטיבית, *milʔú* (מילאו), *nisú* (ניסו), ו-*bit͡sʕú* (ביצעו). פעלים אלו

נבדלים בסגמנט האחרון: *א* ב-*milʔú* (מילאו), תנועה ב-*nisú* (ניסו), ו-*ע* ב-*bit͡sʕú* (ביצעו). בגוף ראשון בזמן עבר, פעלים אלו נבדלים בתנועה המוטעמת (בצורתם הנורמטיבית), דבר המעיד על כך שאלו הן שלוש פרדיגמות שונות: *mil**é**ti* (מילאתי), *nis**í**ti* (ניסיתי), ו-*bit͡s**á**ʕti* (ביצעתי). בחירת התנועה תלויה רק בעיצור האחרון: אם העיצור האחרון הינו ʔ, הרי שהתנועה בצורות הנטיה בעבר תהיה *e*; אם הפועל מסתיים בתנועה, התנועה בצורות הנטיה בעבר תהיה *i*, ואם הפועל מסתיים בעיצור גרוני, תהיה התנועה בצורות הנטיה בעבר *a*. אובדנם של העיצורים הגרוניים גרם לזהות בחלק מצורות הפועל: *milú* → *milʔú* (מילאו), *nisú* (ניסו), ו-*bit͡sú* → *bit͡sʕú*. משאבדו העדויות להבחנות בין הפרדיגמות, יש לצפות לזהות בין יתר הצורות בפרדיגמה.

כלומר, בעקבות אובדנם של העיצורים הגרוניים אבד הצורך בהבחנה בין סוגי פרדיגמה מסוימים, ואם צורך כזה בכל זאת קיים, הרי שבמקרים רבים אין עוד עדויות כיצד להבחין בין סוגי הפרדיגמה. לכן, השינוי הנובע מאובדן זה הוא בלתי נמנע.

העמימות שנוצרה בעקבות השינוי במצאי העיצורים הובילה לשונות (variation). השונות בתורה, מובילה לשינוי נוסף. ניתן להבחין בין שני סוגי שונות: החלפה ומיזוג.

א. **החלפה** הינה חד-כיוונית (סעיף 3.2.1). לדוגמא, את הצורה הנורמטיבית *yeesóf* (יאסוף) מחליפה הצורה *yaasóf*, על פי הצורה הנורמטיבית של *yaavód* (יעבוד). החלפה מסוג זה גורמת לאובדן של התבנית *CeeCóC*.

ב. **מיזוג** הינו דו-כיווני (3.2.2). לדוגמא, לצורה הנורמטיבית *miléti* (מילאתי) יש צורה נוספת מדוברת *milíti* על פי הצורה הנורמטיבית של *nisíti* (ניסיתי). בנוסף, לצורה הנורמטיבית *nisíti* יש צורה נוספת מדוברת *niséti* על פי הצורה הנורמטיבית של *miléti*. דבר זה מעיד על כך שפעלים עם התנועה *e* נודדים לתבנית עם התנועה *i*, ופעלים עם התנועה *i* נודדים לתבנית עם התנועה *e*. הנדידה, אם כן, היא דו-כיוונית, ושתי התבניות שורדות.

ההבחנה בין החלפה ומיזוג מושתתת על כיווניות בלבד, אולם עבור הפעלים שעוברים שינוי, התנועה היא חד-כיוונית בלבד. לכן, תהליך השינוי המתבטא בהחלפה ובמיזוג הינו אחד, ואילו הכיווניות מוסדרת ע״י גורמים נוספים.

2. הכוחות המניעים של השינוי

תהליך השינוי בו נמצאת מערכת הפועל מונע ע״י דמיון (similarity), אך לא די בכך ששתי פרדיגמות יהיו דומות כדי למזגן. כל פרדיגמה תתמזג אך ורק עם הפרדיגמה הדומה לה ביותר בבניין.

על מנת לבחון זאת, אני מציעה מודל המכמת דמיון בין פרדיגמות ומאפשר לצפות אילו פרדיגמות יתמזגו (אם באמצעות מיזוג דו-כיווני או החלפה חד-כיוונית) ואילו לא (§4.3).

על פי המודל המוצע, כל צורות הנטיה של פרדיגמה אחת מושוות לצורות המקבילות בפרדיגמה הדומה לה ביותר באותו הבניין. המודל בודק את מרכיבי הבניין בלבד (תנועות, מוספיות, והמצאותם או העדרם של סגמנטים) ומתעלם מעיצורי השורש הייחודיים לכל פועל. כל הבדל בין שתי צורות פרדיגמה מקבילות נספר וסכום ההבדלים קובע את מידת הדמיון בין שתי הפרדיגמות. כך, ניתן לקבוע איזו פרדיגמה הכי דומה לכל אחד מסוגי הפרדיגמות בבניין. מודל זה מאפשר לחזות אילו פרדיגמות יתמזגו, אך אינו יכול לקבוע מי תהיה הפרדיגמה שתשרוד.

עבור כל זוג פרדיגמות, כיווניות השינוי מוסדרת ע״י שכיחות הסוג (type frequency): פעלים דומים השייכים לתבנית עם השכיחות הנמוכה יותר ינדדו לתבנית עם השכיחות הגבוהה יותר. כך, הנדידה היא לכיוון התבנית המאכלסת מספר רב יותר של פעלים, כלומר מהקבוצה הקטנה לקבוצה הגדולה. אך לא תמיד ניתן לקבוע בקלות איזוהי הקבוצה הגדולה. היחס בין השכיחויות של שתי התבניות הדומות יקבע האם התנועה תהיה חד-כיוונית או דו-כיוונית. יחס קטן בין השכיחויות יקשה על הדוברים לקבוע בוודאות איזו תבנית שכיחה יותר ולפיכך התנועה (כלומר השונות) תהיה דו-כיוונית. יחס גדול בין השכיחויות מראה כי ניתן לקבוע בקלות איזוהי התבנית השכיחה ולכן התנועה (כלומר השונות) תהיה חד-כיוונית.

השונות הנידונה במחקר זה כרוכה בתהליך של שינוי שפה (language change). גם נושא של קבילות הדרגתית (gradient acceptability) בהקשר של שונות נידון כתוצר לוואי (§1.3.6). נושאים אלו הקשורים בתהליך דיאכרוני מהווים בעיה למודלים תיאורטיים דטרמיניסטיים שמטרתם לספק ניתוח סינכרוני של ידע לשוני. המודל אותו אני מציעה לטיפול בשונות במערכת הפועל בעברית מעוגן במסגרת ״תאוריית האופטימליות״ (Prince and Smolensky 1993), ליתר דיוק במודל הסטוכסטי שלו, Stochastic OT (Boersma and Hayes 2001). תאוריית האופטימליות מייצגת את הדקדוק במונחים של אילוצים (constraints) מדורגים הניתנים

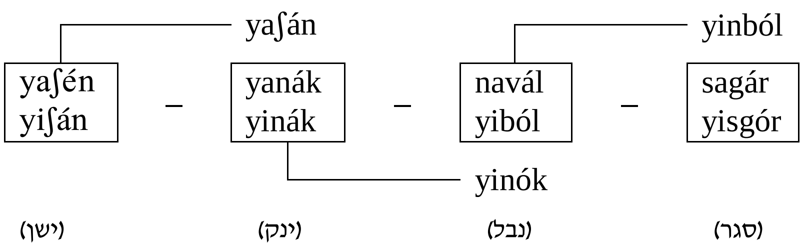
להפרה, אולם ההפרה תהיה מינימלית כך שהפלט האופטימלי הוא זה שמפר באופן מינימלי את האילוצים המדורגים גבוה יותר.

המודל הסטוכסטי מניח אילוצים מדורגים כמו בתאוריית האופטימליות הסטנדרטית, אולם הוא מניח טווח לכל אילוץ והתפלגות נורמלית של נקודת הבחירה (selection point) של האילוץ. על פי מודל זה, ככל ששני אילוצים קרובים יותר, כך גוברת ההסתברות ששני האילוצים המדורגים זה ביחס לזה יניבו שני פלטים שונים. המרחק בין שני האילוצים הוא תוצר של השכיחות של שני הפלטים השונים (variants). מודל זה מוסבר ביתר פירוט בסעיפים 1.3.2 ו-3.3.1.

3. עתיד השינוי

פרק 5 מנסה לצפות לאן מוביל השינוי במערכת הפועל. שונות אינה המצב האופטימלי בשפה, והצפיה היא ששונות הנובעת משינוי תפתר כך שרק פלט אחד ישרוד. בעברית, משמעו של פתרון זה הוא צמצום מספר סוגי הפרדיגמה הקיימים. הפרדיקציה היא שבשני סוגי השונות (ההחלפה החד-כיוונית והמיזוג הדו-כיווני) הצורה החדשה יותר, הלא נורמטיבית, תחליף את הצורה הנורמטיבית. פרדיקציה זו נבדקת באמצעות evolOT (Jäger 2002a), תוכנה המדמה התפתחות של שפה ע״י יישום מודל הלמידה הסטוכסטי Gradual Learning Algorithm (GLA) (Boersma and Hayes 2001), המאפשר לבחון האם הדקדוק המוצע ניתן ללמידה (learnable). זאת תחת ההנחה שדקדוק שנלמד ע״י דור אחד מהווה קלט ללמידה של הדור הבא. ההדמיה באמצעות evolOT נבדקה הן על שונות חד-כיוונית והן על שונות דו-כיוונית והיא מאששת את הפרדיקציה לגבי הפרדיגמות שישרדו. כמו כן, ההדמיה מראה את הקשר שבין תדירות השימוש (frequency of use) בכל אחד מהווריאנטים המעידה על הקרבה בין האילוצים האחראים להם, לבין התקדמות השינוי. כששני האילוצים קרובים, גובר השימוש בווריאנט הלא נורמטיבי על חשבון השימוש בווריאנט הנורמטיבי וככל שהאילוצים קרובים יותר, כך תהליך השינוי של שתי הפרדיגמות הדומות מתקדם יותר וקרוב יותר לסיום.

בסיומו של המיזוג, אין דבר שימנע מהפרדיגמה השורדת להתמזג עם פרדיגמה נוספת, זו שבאותה העת היא הדומה לה ביותר. עדות לכך מגיעה משרשרת הנדידה הבאה בבניין קל (CaCáC):



הפועל *yaʃén* (ישן) נודד לקבוצה של *yanák* (ינק) ומניב בכך את הווריאנט *yaʃán*. באותה העת, הפועל *yanák* נודד לקבוצה של *navál* (נבל) ומניב בכך את הווריאנט החדש *yinók*, וכך הלאה. דבר זה מעיד על כך שגם הפרדיגמה החדשה (*yaʃán-yiʃán*, כמו הפרדיגמה הנורמטיבית *yanák-yinák*) יכולה לנדוד בבוא העת לקבוצה של *navál-yiból* ולהניב את הפרדיגמה החדשה עוד יותר, *yaʃán-yiʃón*. אך עקב הדרישה למיזוג עם הפרדיגמה הדומה ביותר, *yaʃán-yiʃón* לא יכול להתמזג ישירות עם *navál-yiból*, ובכך לצמצם עוד יותר את מספר התבניות בשפה. הסבב הנוכחי של השינוי, המתבטא בשונות שאנו עדים לה כיום, יפחית את מספר הפרדיגמות הייחודיות בבניין קל בלבד בכ-30% (מ-44 ל-30). הפחתה נוספת צפויה, כאמור, בסבב הבא.

4. סיכום

התרשים הבא מסכם את תהליך השינוי במערכת הפועל, החל מאובדן העיצורים הגרוניים, דרך מציאת הפרדיגמה הדומה ביותר, ועד בחירת כיוון השינוי. בתום השינוי, התהליך חוזר לשלב הראשון של מציאת הפרדיגמה הבאה הדומה ביותר.