TEL AVIV UNIVERSITY אוניברסיטת תל אביב

The Lester and Sally Entin Faculty of Humanities

School of Philosophy, Linguistics and Science Studies

Hebrew Phonotactics: Frequency and Grammar

A thesis submitted for the degree of

Doctor of Philosophy

by

Hadas Yeverechyahu

Prepared under the guidance of

Prof. Outi Bat-El Foux

December 2024

Abstract

Phonotactic constraints play a fundamental role in defining permissible consonant sequences in a language, shaping speakers' segmental preferences. While their influence varies across languages, the underlying set of constraints is argued to be universal and shared across all phonological systems (Prince and Smolensky 1993/2004, 1997; Chomsky 2006; Smolensky and Legendre 2006; among others). In addition to universal constraints, language knowledge derives information from the ambient language, with frequency serving as an important component.

The present study examines how universal constraints interact with lexical frequencies to shape phonotactic restrictions within speakers' phonological systems, focusing on Modern Hebrew. Specifically, it explores three widely accepted constraints in phonological theory: the Obligatory Contour Principle (OCP; Leben 1973; Goldsmith 1976; McCarthy 1979, 1981, 1986), the Syllable Contact Law (SCL; Murray and Vennemann 1983, Vennemann 1988, Clements 1990), and the Sonority Dispersion Principle (SDP; Clements 1990).

The study combines two methodological approaches – corpus analysis and psycho-phonological experiments. The corpus analysis examines the influence of universal constraints on the lexicon, while the experiments investigate their role in shaping speakers' phonological systems. Together, these methods address whether speakers' phonological systems are influenced solely by the lexicon or also by universal constraints not reflected in the lexicon.

This dissertation comprises five chapters. Chapter 1 serves as an introduction, outlining the study's research question and goals. Chapter 2 reviews the theoretical background, addressing approaches to linguistic knowledge, the phonotactic constraints central to this study (OCP, SCL, and SDP), and details about the consonant inventory of Modern Hebrew.

Chapter 3 presents the corpus analysis conducted in this study, utilizing the *heTenTen* corpus (Jakubíček et al. 2013), a large and diverse collection of Modern Hebrew texts representing the language's lexicon. Focusing on tri-consonantal verb stems, the analysis investigated consonant co-occurrence patterns in adjacent (C_1C_2, C_2C_3) and non-adjacent (C_1VC_2, C_2VC_3) positions. The results reveal distinct differences in the impact of the phonotactic constraints: the OCP is strongly satisfied, with the lexicon largely avoiding its violation, whereas the SCL and SDP show no discernible effect and are frequently violated.

Chapter 4 describes the psycho-phonological experiments. Experiments A and B examined the OCP and SCL using two different methods. Experiment A examined the impact of OCP and SCL violations on lexical judgments of nonce verbs, and Experiment B analyzed the same violations through measures of accuracy and reaction times in identifying nonce verbs. The results of these experiments clearly indicate the strong influence of the OCP on speakers' phonological systems, whereas the SCL exhibits an intricate pattern, depending on the sonority distances between adjacent consonants.

Experiment C examined SCL violations, building on the findings of the previous two experiments. Relying on speakers' lexical judgments of nonce verbs, the experiment compared SCL satisfaction and violations across two sonority distances: pairs of two obstruents with a sonority distance of 1, and pairs of an obstruent and a sonorant with a sonority distance of 2. The findings reveal that for obstruent-sonorant pairs, participants displayed a significant preference for SCL-violating forms, aligning with the pattern observed in the lexicon. In contrast, for obstruent-obstruent pairs, participants significantly preferred forms that satisfied the SCL, even though the lexicon demonstrates a tendency toward SCL violations within two obstruents.

This divergence indicates that the SCL exerts a stronger influence in obstruent-obstruent pairs, overriding lexical tendencies. This tendency reflects the phenomenon of *The Emergence of the Unmarked* (McCarthy and Prince 1994), wherein universally less marked forms are preferred, even when lexical frequencies suggest otherwise.

Furthermore, examining two sonority distances in Experiment C facilitated an analysis of the SDP constraint. The results indicate that participants favored larger sonority distances between adjacent consonants, aligning with the predictions of the SDP, despite the lack of a clear corresponding preference in their lexicon.

Finally, Chapter 5 concludes the study by summarizing its findings, highlighting the complex interaction between universal constraints, lexical frequencies, and speakers' phonological systems. The research demonstrates that while the lexicon and the constraints it satisfies influence speakers' phonological systems, the effect of constraints with no apparent impact on the lexicon may emerge. This highlights the unique contribution of universal constraints to the shaping of phonological systems.

Acknowledgments

From the early stages of my studies, I was fascinated by the sounds of language and by the patterns that make some combinations feel natural and others not. This dissertation grew out of that curiosity and reflects years of study and research. I am thankful to the many people whose wisdom and kindness accompanied me, both in this work and beyond it.

First and foremost, I would like to thank my PhD supervisor, Prof. Outi Bat-El Foux. I met Outi in my first year of studies, in the *Beginning Phonology* course, which opened the door for me to a fascinating field that became central to my academic path. Since then, I have had the honor to continue learning from Outi in courses and seminars, and as her MA and PhD student. Outi, I am grateful for everything I have learned from you about phonology, research, and academic thinking more broadly. I greatly appreciate your guidance over these years, and the way you encouraged me to ask the right questions and to keep an open mind.

I also wish to thank Dr. Evan Cohen. From my first phonetics class, through courses, seminars, and co-supervision in my MA, your teaching and guidance enriched my path. I have greatly valued your wise advice and the positive spirit you always brought with you over the years.

I am thankful as well to the lecturers and teaching assistants in the Linguistics Department for their inspiring classes and enriching discussions. I would also like to thank the faculty of the Department of Hebrew Language, and especially Prof. Moshe Florentin, from whom I had the opportunity to learn about Biblical Hebrew phonology and morphology. I am also grateful to the School of Philosophy, Linguistics, and Science Studies, with special thanks to Inna Zilber, Graduate Studies Coordinator, and Nava Weinstock, Director of Administration, for their kindness and helpfulness throughout my studies.

My thanks also go to my colleagues in the Linguistics Department during my PhD years for the exchange of ideas and the good company: Noa Bassel, Si Berrebi, Hila Davidovitch, Alon Fishman, Noa Geller, Mor Haim, Noa Handelsman, Yuval Katz, Nicole Katzir, Stav Klein, Hezi Shabanov, Aviv Schoenfeld, Daniella Yariv, and Hadass Zaidenberg. Special thanks to Noa Karni, whose helpfulness I truly appreciated.

Among the highlights of my doctoral studies was teaching courses at Tel Aviv University and the Hebrew University of Jerusalem. I am thankful to the students I taught over the years for their thoughtful questions and discussions that deepened my thinking. I truly enjoyed witnessing their curiosity about phonology and research.

I greatly appreciate those who contributed directly to this research. My sincere thanks go to Dr. Nurit Melnik, Dr. Israela Becker, and Nicole Katzir for introducing me to the *heTenTen* corpus; to Dr. Evan Cohen, Prof. Orna Peleg, Hila Davidovitch, Hezi Shabanov, and Ofer Brodatch for their helpful input on the experimental design; to Si Berrebi and Naomi Feldheim for recording stimuli; to Dr. Boaz Tzur for valuable comments on statistics; and to all the participants in the experiments.

Throughout my PhD, I divided my time between academic work and a professional career. I am grateful to the Center for Educational Technology, my former workplace, and to Taldor Assessment, my current one, for enabling me to combine these two worlds. I especially wish to thank Dr. Meirav Arieli-Attali, head of Taldor Assessment, for her sincere interest, kind advice, and thoughtfulness.

I am deeply thankful to my friends, whose genuine interest, insightful suggestions, and steady encouragement have inspired me in countless ways. Special thanks to Yael Firer, whom I was fortunate to meet through phonology, for the conversations on research and life, and for her helpful comments and joy along the way; and to Hadas Gelbart, whose wisdom, advice, and support have been invaluable to me.

Above all, I would like to thank my parents and my brother Doron, and express my deepest gratitude for their constant support and encouragement throughout this journey.

Table of Contents

Chapter 1

Introduction

Words are composed of sequences of speech sounds, but not all sequences occur with equal likelihood. The likelihood is scalar, ranging from preferred sequences to prohibited ones. This distinction between preferred and prohibited patterns is supported by typological studies, which demonstrate that certain linguistic properties and phenomena are consistently observed across most languages, while others are less frequent (Greenberg 1978, Maddieson 1984).

Different theoretical approaches attempt to explain these patterns. One approach holds that universal constraints govern phonotactic preferences, with abstract principles forming an internal source of knowledge shared by all speakers (Prince and Smolensky 1993/2004, 1997; Chomsky 2006; Smolensky and Legendre 2006; among others). The alternative approach, known as the Usage-Based approach (Elman et al. 1996, Tomasello 1998, Bybee 2006, among others), attributes these preferences to an external source, emerging from speakers' exposure to patterns in their language's lexicon and usage-based generalizations.

The present study examines the role of universal constraints, focusing on how they interact with lexical frequencies to shape phonotactic restrictions within speakers' phonological systems. While the lexicon plays a significant role in shaping speakers' phonological systems, it may not fully account for all patterns their phonological system permits. This study explores whether universal constraints can explain patterns in phonological systems that operate beyond the influence of the lexicon.

In line with this goal, the study examines three phonotactic constraints extensively discussed in the phonological literature: the Obligatory Contour Principle (OCP), the Syllable Contact Law (SCL), and the Sonority Dispersion Principle (SDP). These constraints reflect different dimensions of phonological organization, including feature similarity and sonority relations between adjacent segments.

The OCP (Leben 1973; Goldsmith 1976; McCarthy 1979, 1981, 1986) limits the occurrence of identical elements within a melodic level. Initially proposed for tonal systems, it was later extended to consonantal systems, where it shapes phonological structures by restricting the co-occurrence of consonants with shared features. The SCL (Murray and Vennemann 1983; Vennemann 1988; Clements 1990) governs the contact between adjacent syllables, favoring a coda more sonorous than the onset of the following syllable. The SDP (Clements 1990) prioritizes greater sonority distance between adjacent segments over a smaller distance.

The analysis focuses on Modern Hebrew and uses it as a test case for examining whether the OCP, SCL, and SDP operate within speakers' phonological systems independently of supporting lexical patterns. Accordingly, the study combines corpus analysis with psycho-phonological experiments, providing a comprehensive account of how these universal constraints shape the phonotactic system of the language. Such an investigation contributes to broader discussions on the role of universal constraints in phonological structure.

The corpus analysis utilized the *heTenTen* corpus (Jakubíček et al. 2013), a large and diverse collection of Modern Hebrew texts representing the language's lexicon. The analysis focused on tri-consonantal verb stems to investigate consonant co-occurrence patterns in adjacent (C_1C_2, C_2C_3) and non-adjacent (C_1VC_2, C_2VC_3) positions. This approach enabled a detailed examination of co-occurrence restrictions across various phonological environments and established an empirical basis for evaluating the activity of the tested constraints in the lexicon.

The psycho-phonological experiments investigated the role of phonotactic constraints within the phonological systems of native Hebrew speakers. Specifically, they examined whether these constraints influence speakers' linguistic behavior and whether their phonological system activity corresponds to their lexicon patterns. Designed to measure sensitivity to these constraints, the experiments evaluated their impact on phonological preferences.

This study contributes to understanding how universal constraints interact with language-specific phonological systems. By analyzing their activity in both the lexicon and speakers' phonological systems, the research highlights their role in shaping phonotactic preferences and restrictions. Additionally, the findings contribute to cross-linguistic discussions on the universality of phonological constraints and their interplay with language-specific factors, offering insights that extend beyond the scope of Hebrew.

The dissertation is organized into five chapters. Chapter 2 reviews the theoretical background, including approaches to linguistic knowledge, the phonotactic constraints (OCP, SCL, and SDP), and the consonant inventory of Modern Hebrew. Chapter 3 presents the corpus analysis, examining the activity of universal constraints in the Hebrew lexicon. Chapter 4 describes three psycho-phonological experiments designed to evaluate the role of these constraints in native speakers' phonological systems. Finally, Chapter 5 synthesizes the findings, discusses their theoretical implications, and offers directions for future research.

Chapter 2

Theoretical Background

This chapter provides the relevant theoretical background for the study, addressing two main approaches to linguistic knowledge (§2.1) and focusing on three phonotactic constraints (§2.2) in the context of Modern Hebrew consonantal system (§2.3).

2.1. Approaches to Linguistic Knowledge

Phonotactic asymmetries, where some sound sequences being favored across languages while others are dispreferred, have been examined from two broad theoretical perspectives. Constraint-based approaches maintain that universal constraints govern phonotactic preferences. These constraints are part of speakers' linguistic competence, serving as an internal source of knowledge (Prince and Smolensky 1993/2004, 1997; Chomsky 2006; Smolensky and Legendre 2006; among others). Usage-Based approaches, in contrast, locate the source of these preferences outside the pure competence system, proposing that preferences arise from speakers' exposure to lexical distributions and usage-based generalizations (Elman et al. 1996, Tomasello 1998, Bybee 2006, among others). The following section develops these two approaches and considers their implications for phonotactic theory.

The universal framework assumes that phonological preferences are governed by a set of abstract principles shared across languages, forming part of Universal Grammar (UG). According to this approach, universal constraints represent inherent properties of the human language faculty, are present in all grammars, and account for cross-linguistic similarities. These constraints operate universally, but not all are active in every language.

This constraint-based perspective is formalized and expanded in Optimality Theory (OT), introduced by Prince and Smolensky (1993/2004). OT builds on the idea of universal constraints through their interaction and ranking. These include faithfulness constraints, which require preserving input structure, and markedness constraints, which favor unmarked or simpler structures. According to this framework, all languages share the same set of universal constraints. However, differences arise from how these constraints are ranked within each language's grammar. Multiple output candidates are generated and evaluated against the ranked constraints for a given input. Candidates that incur violations of higher-ranked constraints are excluded, and the candidate that best satisfies the constraint hierarchy is selected as optimal.

Building on this perspective, phonetically based phonology proposes that markedness constraints are grounded in phonetic knowledge, namely speakers' implicit sensitivity to the physical conditions of speech production and perception (Hayes et al., 2004 among others). On this view, the recurrence of unmarked patterns across languages reflects shared articulatory and perceptual pressures rather than arbitrary grammatical stipulations. Such an approach situates markedness constraints as components of grammar while explaining their universality through phonetic grounding.

Within OT, the mechanism of constraint ranking explains how universal constraints account for cross-linguistic variation while preserving their universality. For example, there is a universal preference for *gla* over *lga* (assuming each sequence appears within a single syllable), as sonority tends to increase toward the syllable nucleus – a pattern known as the Sonority Sequence Generalization (SSG; Selkirk 1984, Clements 1990). While this preference holds across languages, differences in ranking explain why structures like *lga* are permissible, for instance, in Russian but not Hebrew or English.

The preference for certain sound sequences over others is also evident in psycholinguistic experiments across various languages (e.g., Berent and Shimron 1997, Berent et al. 2001 for Hebrew; Frisch and Zawaydeh 2001 for Arabic; Dupoux et al. 1999 for Japanese; Dupoux et al. 2011 for Japanese and Portuguese). As discussed, OT posits that UG constraints are present in all phonological systems, regardless of specific structures in a given language. Psycholinguistic evidence supports this claim; for instance, Berent et al. (2008) demonstrated that Korean speakers are sensitive to the SSG even though Korean prohibits clusters entirely, meaning speakers lack the opportunity to learn the SSG from the lexicon. Similarly, language acquisition research shows that children are attuned to UG constraints even without frequency-based support from their language. For example, Hebrew-acquiring children initially exhibit penultimate stress, a universally preferred pattern, despite the predominance of final stress (Adam and Bat-El 2009).

The Usage-Based approach challenges the UG perspective, arguing that speakers can learn the phonotactic restrictions of their language from input and the lexicon without requiring universal constraints (Elman et al. 1996; Tomasello 1998, 2003; MacWhinney 1998; Ellis 2002; Bybee 2006, among others). By utilizing general cognitive abilities, such as pattern recognition and generalization, speakers extract phonotactic restrictions directly from language use. What appear to be language universals, according to this view, are instead statistical tendencies shaped by auditory and motor constraints on language evolution (Blevins 2004). For instance, words beginning with *lb* tend to decline in use compared to those beginning with *bl*, as they are more likely to be misperceived or mispronounced. This approach, therefore, positions the lexicon as a heuristic source, emphasizing the constraints that define valid combinations in the language.

This approach aligns with computational models of language learning, such as connectionist modeling, probabilistic grammars, and statistical learning (e.g., Redington et al. 1998, Mintz et al. 2002, Newport et al. 2004). These models process input based solely on a lexicon (e.g., words or nonce-words) to "learn" relevant linguistic constraints (See Hayes and Wilson 2008, Hayes et al. 2009, Adriaans and Kager 2010, Daland et al. 2011, Rebuschat and Williams 2012, Becker and Gouskova 2016, Rasin and Katzir 2016, among others). For example, the model by Hayes and Wilson (2008), developed with English data, successfully captured various phonotactic phenomena in Hebrew, demonstrating the potential of statistical learning to derive phonotactic limitations. However, it failed to account for speakers' sensitivity to restrictions involving non-native consonants (Berent et al. 2012).

Building on the above theoretical perspectives, the current study examines whether the lexicon is sufficient to account for the phonological system in speakers' minds or whether the speakers' minds reflect patterns that can be attributed solely to universal constraints. The lexicon is a major component in the development of speakers' phonological systems and is, therefore, expected to influence these systems. However, the lexicon is often "stained" by the effects of historical changes that are not universally natural. Thus, the effect of universal constraints not supported by the lexicon may emerge.

For this purpose, three universal constraints are examined in the context of Modern Hebrew lexicon (Chapter 3) and speakers' phonological system (Chapter 4): the Obligatory Contour Principle (OCP), the Syllable Contact Law (SCL), and the Sonority Dispersion Principle (SDP). The following section provides an overview of these constraints.

2.2. Phonotactic Constraints

Phonotactic constraints refer to preferences and restrictions governing permissible combinations of segmental sequences. This study focuses on three widely discussed constraints, as outlined below.

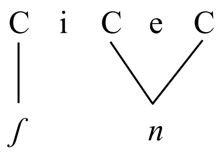
The Obligatory Contour Principle (OCP) restricts identical elements at the melodic level. Initially proposed for tonal systems (Leben 1973, Goldsmith 1976), it was later extended to segments (McCarthy 1979, 1981, 1986), features, syllables, and morphemes (Yip 1988). For features, the OCP applies to feature groups rather than individual features, incorporating non-linear representation through the Feature Geometry hierarchy (Clements 1985, Sagey 1986, Kaisse 1988, McCarthy 1988, Padgett 1995). Within an autosegmental framework (Goldsmith 1976), the OCP explains restrictions on similar consonants by attributing violations to shared features

(Pierrehumbert 1993, Frisch et al. 2004 for Arabic; Yeverechyahu 2014, 2019 for Hebrew).

The scope of the OCP extends beyond adjacent consonants to nonadjacent ones, depending on their proximity. For instance, consonants separated by vowels (CVC) may also violate the OCP, though to a lesser extent than adjacent consonants (CC), due to their greater distance (Rose 2000, Rose and Walker 2004).

Hebrew, like other Semitic languages, provides evidence for the OCP through co-occurrence restrictions on homorganic consonants within stems (Greenberg 1950; McCarthy 1979, 1981, 1986) as well as on similar, non-homorganic consonants (Yeverechyahu 2014, 2019).[1] Notably, the second and third stem consonants (C_2 and C_3) are allowed to be identical, as in ***ʦitet*** 'quote PAST.3MSG', ***hilel*** 'praise PAST.3MSG', and ***ʃinen*** 'memorize PAST.3MSG'. Assuming a non-linear representation, McCarthy (1981) proposes that there is no OCP violation in such cases, as these verbs contain only two underlying consonants, with the second consonant occupying two prosodic positions, as illustrated in (1).

(1) Non-violation of the OCP by identical consonants at the right edge



Hebrew also avoids morphological operations resulting OCP violations; for example, *dike* 'make depressed PAST.3MSG' does not undergo the valence-changing operation to **hitdake* 'get depressed PAST.3MSG', as this would create the OCP-violating sequence *t-d* (Laks 2011).

Psycholinguistic evidence supports Hebrew speakers' sensitivity to the OCP and similarity-based restrictions. Experiments reveal a correlation between similarity

[1] There are a few exceptional verbs in Hebrew where C_1 and C_2 are identical: *mimen* 'finance PAST.3MSG', *mimeʃ* 'realize PAST.3MSG', and *hitmameʃ* 'be realized PAST.3MSG'.

factors and speakers' judgments of nonce verbs and their reaction times in lexical decision tasks (Yeverechyahu 2014). Similar sensitivity has been observed in Arabic speakers (Frisch and Zawaydeh 2001).

The other two constraints examined in this study, the **Syllable Contact Law (SCL)** and the **Sonority Dispersion Principle (SDP)**, restrict co-occurrence based on the sonority distance between adjacent consonants. These constraints assume segments are arranged along a sonority scale (or hierarchy). While various proposals for sonority scales exist (Parker 2002, Albert 2023, among others), they generally classify segments according to their manner of articulation (MoA) and perceived loudness, with stops being the least sonorous and vowels the most sonorous (Foley 1972), as illustrated in (2).

(2) The sonority scale

vowels > glides > liquids > nasals > fricatives > stops

The Syllable Contact Law (SCL) restricts the contact between adjacent syllables, specifically between the coda of one syllable and the onset of the immediately following syllable (Murray and Vennemann 1983, Vennemann 1988, Clements 1990). According to the SCL, if α and β are segments in adjacent syllables, a syllable contact α.β is preferred when α is more sonorous than β, with greater sonority distances resulting in better contact (in this study, the latter is treated separately as the Sonority Dispersion Principle, SDP; see below).

The Hebrew lexicon appears to disregard the SCL, as evidenced by minimal pairs like *sag.ʁa* 'close PAST.3FSG' and *saʁ.ga* 'knit PAST.3FSG'. Not only do both sequences occur in the lexicon, but the SCL-violating sequence *gʁ* occurs more frequently than the sequence satisfying the SCL, *ʁg* (77 vs. 63, respectively, in Bolozky and Becker's *Living Lexicon of Hebrew Nouns* 2006).[2] Further support is provided in §3.4.2.

[2] I adopt the syllabification often assumed for Hebrew, i.e., VC.CV; under different syllabification (*sa.gʁa*), there is no SCL violation (Albert 2014), and see further discussion in §4.4.3.2.2.

The Sonority Dispersion Principle (SDP) states that greater sonority distances between adjacent segments are preferred over smaller ones (Clements, 1990). This principle identifies the optimal sonority gap as occurring between a consonant and a vowel within a syllable, particularly between a stop and a vowel (e.g., *ta, at*), as such combinations maximize the sonority difference. Consequently, those with greater sonority distances are considered more favorable in consonant sequences. While the constraint is typically discussed concerning consonant clusters, it can be extended to encompass any adjacent consonants, not necessarily within the same syllable.

The SDP can explain why only certain consonant clusters are permitted in some languages. For instance, English allows only obstruent-liquid clusters (see below on clusters with *s*), whereas Modern Hebrew permits clusters with smaller sonority distances between consonants (e.g., stop-nasal, as in *tmuna* 'picture') and even sonority plateaus (e.g., two stops, as in *bgadim* 'clothes').

The SDP effect in Hebrew is exemplified in blends (Bat-El 1996). For instance, *ʁam.kol* ('speaker', derived from *ʁam* 'loud' + *kol* 'voice') is preferred over *kol.ʁam*, as the sonority distance between *m* and *k* (a nasal and a stop) is greater than that between *l* and *ʁ* (two liquids). This preference reflects *The Emergence of the Unmarked* phenomenon (McCarthy and Prince 1994), in which less marked structures are favored over more marked ones, even in languages where the relevant constraint is typically inactive.

The SCL and the SDP can be interpreted as a hierarchy of constraints (Gouskova 2001, 2002). According to this approach, the differences between languages regarding which sequences are permissible are derived from the specific cutoff points along the hierarchy. However, in the current study, the SCL is treated as a single, binary constraint, while the SDP captures differences in sonority distance between consonants.

It should be noted that stridents often exhibit distinct phonological behavior concerning sonority. One well-documented example of this is the behavior of s-stop clusters (sC clusters), which involve a strident fricative, typically *s*, preceding a stop consonant, as in English words like ***stap*** 'stop', ***skaɪ*** 'sky', and ***spɔɹt*** 'sport' (Fudge 1969, Kenstowicz 1994, Goad 2011, Albert 2023). Although such clusters are relatively common in languages that allow consonant sequences (Steriade 1999, Morelli 2003), their sonority profile is classified as ill-formed in word-initial position, contrary to their actual distribution. This discrepancy highlights the unique behavior of stridents in sonority-based analyses. One common analysis posits that the *s* in such clusters is *extrametrical*, excluding it from the following consonant's syllable structure. By treating the *s* as external to the syllable, the sequence avoids violations of sonority-based constraints (Steriade 1982, among others). Due to these complexities, the behavior of stridents in sonority-related constraints is beyond the scope of the current study.

2.3. Overview of Modern Hebrew Consonants

The current study examines the role of phonotactic universal constraints in Modern Hebrew, drawing on a corpus analysis and psycholinguistic experiments with native speakers. This subsection focuses on the consonantal system of Modern Hebrew, with particular attention to place of articulation and sonority – features central to the universal constraints examined in this study. Specifically, place of articulation is relevant to the OCP, while sonority underpins the SCL and SDP.

The consonant inventory of Modern Hebrew is presented in Table (3).

(3) Modern Hebrew consonant inventory (Asherov and Cohen 2019)

	Bilabial	Labio-Dental	Alveolar/ Dental	Post-alveolar	Palatal	Velar	Uvular	Glottal
Plosives	p b		t d			k g		ʔ
Affricates			ʦ	ʧ ʤ				
Fricatives		f v	s z	ʃ ʒ			χ	h
Nasals	m		n					
Lateral approx.			l					
Central approx.					j	w	ʁ	

Notes: (a) In gray are consonants not included in the current study – glottals and consonants appearing in loanwords; (b) *w* is labio-velar central approximant.

Table (3) presents the consonant inventory of Modern Hebrew, displaying consonants across multiple places and manners of articulation. The following notes highlight aspects of this inventory.

- **Loan Consonants:** The consonants *ʧ*, *ʤ*, *ʒ*, and *w* are loan consonants in Modern Hebrew (Asherov and Cohen 2019, Cohen 2019). Historically, *ʃ* was the sole post-alveolar consonant in Hebrew, but significant borrowing from other languages has likely led to the phonemic adoption of *ʒ*, *ʧ*, and *ʤ*. Another addition to the phonemic inventory is *w*. These loan consonants appear in borrowed words (e.g., *ʧips* 'chips'; *ʤip* 'jeep'; *ʒurnal* 'journal'; *wala* 'really; indeed', originally from Arabic, *walla*) and in denominative verbs, primarily in slang derived from English nouns (e.g., *leʧotet* 'to chat', *leʤamʤem* 'to jam'). These loan consonants are not included in the current study.

- **Pharyngeals and Glottals:** The pharyngeals *ʕ* and *ħ* (orthographic ע and ח, respectively) are not included in the table above. Although they were part of the consonant inventory of Biblical Hebrew, they are absent for most Modern Hebrew speakers, with *ħ* merging into *χ* and *ʕ* merging into *ʔ* or omitted (Berman 1978, Laufer 1990, Gafter 2014, Asherov and Cohen 2019, Faust 2019, among others). However, some speakers of the *Mizrahi Israeli* variety preserve the distinctions between *χ* and *ħ* and between *ʔ* and *ʕ*. The phonological status of the historical pharyngeals in the underlying representations of speakers without surface pharyngeals remains a topic of debate (see Bolozky 1978; Matras and Schiff 2005; Faust 2005, 2019; Laks et al. 2016; Enguehard and Faust 2018; Berrebi and al. 2022).

  The glottals *ʔ* and *h* are considered part of the Modern Hebrew consonant inventory. However, they are frequently omitted (Gafter 2014, Asherov and Cohen 2019). In the onset position, *ʔ* can occasionally be realized, and *h* in this position may sometimes surface as a *ʔ* instead of being omitted (Faust 2019). These glottals are not included in the current study.

- **The Hebrew rhotic:** The Modern Hebrew rhotic is a uvular approximant, though it may also be pronounced as a fricative, trill, tap, or even a plosive in specific phonological environments (Bolozky and Kreitman 2007, Cohen et al. 2019). Historically, the exact place of articulation of the Biblical rhotic remains uncertain. It has been classified either as a guttural back consonant, interpreted as velar or uvular, as a coronal dental consonant, or as a consonant with both coronal and dorsal variants (Gesenius 1910, Blau 2010, Meloni 2021, among others). However, nowadays the vast majority of speakers of Modern Hebrew, including almost all young native speakers and most adults, use only a dorsal rhotic (Yaeger-Dror 1988, Cohen et al. 2019, Berrebi 2021).

Consonants are categorized into **natural classes** primarily based on their **place of articulation**, which is particularly relevant for the OCP. The classification adopted in this study is based on Frisch et al. (2004) proposal for Arabic, as presented in (4):[3]

(4) Natural Classes

- Labials (LAB) = {p, b, f, v, m}
- Coronal Obstruents (CORo) = {t, d, ʦ, s, z, ʃ}
- Coronal Sonorants (CORs) = {n, l, j}
- Dorsals (DOR) = {k, g}
- Gutturals (GUT) = {χ, ʁ}

This classification organizes consonants into broad articulatory categories. Bilabial and labio-dental consonants are grouped under Labials (LAB). Alveolar/dental, post-alveolar, and palatal consonants are grouped under Coronals (COR), with a distinction between obstruents (plosives, affricates, fricatives) and sonorants (nasals, lateral approximants, central approximants), following McCarthy (1988, 1994) and Frisch et

[3] The classification includes only consonants that are part of the study. The following consonants, which are excluded from the study, are classified as follows: the loan consonants *ʧ*, *ʤ*, and *ʒ* belong to the CORo group; the consonant *w* is labio-velar and could theoretically belong to the LAB group, the DOR group, or both; the glottal consonants *ʔ* and *h* are usually considered lack a place of articulation (Sagey 1986, Bessell 1992).

al. (2004). Velars are grouped under Dorsals (DOR), while uvulars are grouped under Gutturals (GUT).

The uvular consonants *χ* and *ʁ* are categorized as gutturals in this study, although they can also be classified as dorsals or as belonging to both groups (see Greenberg 1950, McCarthy 1994, Frisch et al. 2004, Asherov and Cohen 2019). Their classification as gutturals is primarily supported by their phonological behavior in Hebrew (Bolozky 1978, Blau 2010, Faust 2019, among others). *χ* which corresponds to the letter ח (and not כ) triggers the insertion of an epenthetic vowel in specific environments, such as word-finally after a non-low vowel (e.g., *ʃole**a**χ* 'send PRESENT.MSG', cf. *kotev* 'write PRESENT.MSG') or within consonant clusters where *χ* would otherwise lack a following vowel (e.g., *χefets–χ**a**fatsim* 'object–objects', *χodeʃ–χ**o**daʃim* 'month–months', cf. *kelev–klavim* 'dog–dogs', *boker–bkarim* 'morning–mornings').[4] *χ* also influences vowel lowering, as seen in forms like *ʃoleaχ–ʃol**a**χ**a**t* 'send PRESENT.MSG–FSG' (cf. *kotev–kot**e**v**e**t* 'write PRESENT.MSG–FSG').

ʁ, while exhibiting fewer synchronic phonological effects typically associated with back consonants, still shares important characteristics with *χ*, justifying their grouping in this study. For instance, *ʁ* contributes to vowel lowering in forms such as ***p**e**ʁ**ek* 'dismantle PAST.3MSG' and *s**e**ʁev* 'refuse PAST.3MSG' (cf. *b**i**ʃel* 'cook PAST.3MSG'). Further support comes from Hebrew child language, where *ʁ* can be replaced by *χ*, as in *oχ* for *oʁ* 'light' (Ben-David and Bat-El 2016). These shared phonological behavior underscore their classification as gutturals within the phonological system of Hebrew.

The consonants can also be positioned along a **sonority scale** based on their manner of articulation, as illustrated in (5).

---

[4] The consonant *χ* which corresponds to the letter כ does not exhibit the phonological behavior typically associated with back consonants. For example, forms such as *holeχ–hol**e**χ**e**t* 'walk PRESENT.MSG–FSG' lack the phonological effects seen in *ʃoleaχ–ʃol**a**χ**a**t* 'send PRESENT.MSG–FSG'. This phenomenon is rooted in historical phonological distinctions: while in Biblical Hebrew ח was a pharyngeal consonant (*ħ*), כ was a velar consonant (*x*) (Blau 2010, among others). Although this distinction has not been preserved in Modern Hebrew, as discussed in this chapter, the phonological differences persist.

(5) Sonority scale of Modern Hebrew consonants[5]

sonorants							obstruents		
glides	>	**liquids**	>	**nasals**	>	**fricatives**	>	**stops**
j		l		m		f		p
		ʁ		n		v		b
						s		t
						z		d
						ʃ		ʦ
						χ		k
								g

The classification of consonants along the scale reflects their sonority levels, shaped by their manner of articulation. Plosives and affricates, characterized as [-son, -cont], are classified as stops. Fricatives and nasals are grouped as their names suggest. The liquid category includes *l* and *ʁ*, while the glide category consists of *j*. For further details on the classification of consonants based on sonority, see, for example, Albert (2023).

The **Sonority Distance** (SonD) between adjacent categories on the sonority scale is one unit. For example, the SonD between stops and fricatives is 1, while the SonD between stops and nasals is 2. SonD values can also be assigned a positive or negative profile. A positive profile indicates a rise in sonority, as seen in the sequence *tm* (stop-nasal), where SonD is +2. A negative profile indicates a fall in sonority, as in the sequence *mt* (nasal-stop), where SonD is -2.

This classification is particularly relevant to the study of sonority-related constraints, such as the Syllable Contact Law (SCL) and the Sonority Dispersion Principle (SDP).

[5] As in (4), the classification includes only consonants that are part of the study. The following consonants, which are excluded from the study, are classified as follows: *ʧ*, *ʤ*, and *ʔ* belong to the stop group; *ʒ* and *h* belong to the fricative group; and *w* belongs to the glide group.

Chapter 3

Corpus Analysis

3.1. Introduction

The first part of the study examines the status of phonological constraints in the Hebrew lexicon, as reflected in verbs extracted from a Modern Hebrew corpus. The constraints in question are the **OCP**, **SCL**, and **SDP**, as detailed in §2.2.

The analysis focuses on verb stems with three consonants extracted from the *heTenTen* corpus (see §3.2.1 for corpus details). Verbs were analyzed based on the stems within their inflectional paradigms (see §3.2.2.4). For example, in the paradigm of *hiχtiv* 'dictate PAST.3MSG', two types of stems are observed: *χtiv* (as in *hi**χtiv**a* 'dictate PAST.3FSG', *ja**χtiv*** 'dictate FUT.3MSG', etc.) and *χtav* (as in *hi**χtav**ti* 'dictate PAST.1SG', *hi**χtav**t* 'dictate PAST.2FSG', etc.). All in all, **1833** inflectional paradigms containing **6892** unique stems were analyzed, with an average of 3.76 stems per paradigm, ranging from 2 to 6 stems.

Four phonological environments were analyzed:

a. **C_1C_2:** adjacent first and second stem consonants (e.g., ***χt**iv,* as in *hi**χtiv*** 'dictate PAST.3MSG');

b. **C_2C_3:** adjacent second and third stem consonants (e.g., *ka**tv***, as in ***katv**a* 'write PAST.3FSG');

c. **C_1VC_2:** nonadjacent first and second stem consonants (e.g., ***kat**av*, as in ***katav*** 'write PAST.3MSG');

d. **C_2VC_3:** nonadjacent second and third stem consonants (e.g., *k**atav***, as in ***katav*** 'write PAST.3MSG').

The corpus analysis has two main goals. First, it aims to provide a detailed description of the role of the three constraints in Hebrew. As discussed in §2.1, universal constraints are not necessarily satisfied in all languages and can be violated. This study examines whether these constraints are satisfied in the Hebrew lexicon and, if so, to what extent. Second, the corpus analysis serves as the foundation for the psycho-phonological experiments (see Chapter 4), which explore the status of these constraints within speakers' phonological systems. By comparing the constraints' status in the lexicon with the experimental results, this study aims to determine whether the patterns observed in the corpus are mirrored in speakers' phonological systems and how speakers respond to violations of constraints that are not active in their lexicon.

3.2. Methodology

The following subsections introduce the *heTenTen* corpus, which was selected for the current analysis, and describe the data preparation process. Key methodological decisions are outlined and discussed in detail.

3.2.1. heTenTen corpus

The corpus selected for the analysis is the Hebrew Web Corpus (*heTenTen*) by *Sketch Engine*, a Modern Hebrew corpus consisting of texts collected from the internet. These include blogs, forums, news sites, and various other online platforms. As such, the corpus represents written Modern Hebrew, though often in informal or conversational registers that reflect spoken usage. It is part of the *TenTen* corpus family (Jakubíček et al. 2013), a set of web corpora in over 40 languages, designed with a target size exceeding 10 billion words.

For this research, the *Hebrew Web 2021* version (*heTenTen21*), processed by *Yet Another (natural language) Parser* (Sima'an et al. 2001), was used. The corpus was compiled using the *SpiderLing* web crawler (Suchomel and Pomikálek 2012), with data collected in November–December 2019, November–December 2020, and January 2021. It contains over 2.7 billion words sourced from more than 43,000 web domains. The corpus is tagged by part of speech and includes morphological annotations for features such as gender, number, affixes, and others.

The *heTenTen* corpus offers an extensive and diverse representation of Modern Hebrew, making it an ideal resource for this analysis. Reflecting contemporary language use, it draws from a wide range of websites and captures different language registers. Unlike dictionaries, which often mix archaic and contemporary vocabulary, or smaller corpora that focus on specific registers such as formal newspaper language, the *heTenTen* corpus provides a comprehensive view of Modern Hebrew as it is currently used. Furthermore, its detailed part-of-speech tagging facilitates the extraction of verbs, which are the focus of this study.

3.2.2. Methodological decisions

The analysis focuses on Modern Hebrew tri-consonantal verb stems. Several methodological decisions were made during the corpus preparation process, as outlined in the following subsections.

3.2.2.1. Focusing on verbs

The current study focuses on verbs, which constitute a distinct class in Hebrew and function as a sub-lexicon (see Becker and Guskova, 2016). This distinctiveness is evident in both their morphological and phonological behavior.

From a morphological perspective, Hebrew verbs are organized into a closed set of fixed templates, known as *binyanim* (see Berman 1978; Bolozky 1978; Aronoff 1994; Bat-El 1989, 2003; Doron 2003, among others). Each template is characterized by a specific prosodic structure and vocalic pattern, with some templates also including prefixes. The verb templates are outlined in Table (6).

(6) Verb templates: *binyanim* (examples in PAST.3MSG; *k-χ* alternation due to spirantization)[6]

Template:	qal (pa'al)[7]	nif'al	hif'il	huf'al	pi'el	pu'al	hitpa'el
Example:	*katav*	*niχtav*	*hiχtiv*	*huχtav*	*kitev*	*kutav*	*hitkatev*

The templates *huf'al* and *pu'al* (in gray) represent the passive forms of *hif'il* and *pi'el*, respectively. Lexical-syntactic and phonological evidence suggests that these templates should not be considered independent but rather dependent on their active counterparts. This will be further discussed in §3.2.2.3.

The strict alignment of verbs to fixed templates contrasts sharply with the greater morphological flexibility of other parts of speech, such as nouns. Unlike verbs, which must conform to one of the established *binyanim*, nouns do not adhere to rigid patterns. This distinction is particularly evident with loanwords: nouns can be borrowed into Hebrew without adapting to specific morphological templates (e.g., *katalog* from 'catalog', *esemes* from 'SMS'). In contrast, loan verbs must integrate into the verbal system by adapting to one of the *binyanim* (e.g., *lekatleg* from 'catalog', *lesames* from 'SMS', both in *pi'el*; see Berman 1978, Bat-El 1994, Ussishkin 1999, Schwarzwald 2002, among others).

---

[6] In Modern Hebrew, several paradigms exhibit alternations between the stops *p*, *b*, and *k* and their corresponding fricatives *f*, *v*, and *χ*. These patterns are traces of a broader phonological process in Biblical Hebrew, where most stop consonants were spirantized in post-vocalic positions. However, historical developments beyond the scope of this study have resulted in numerous exceptions, such as stop variants following vowels vowel (e.g., *sa**p**a* 'couch') and fricative variants occurring outside post-vocalic contexts (e.g., ***χ**atul* 'cat'). Consequently, it is widely acknowledged that stop-fricative alternations in Modern Hebrew are primarily governed by morphological and lexical rules (Adam 2002; Albert 2014, 2019; among others).

[7] *Binyan pa'al* is commonly referred to as *binyan qal* ("*the simple binyan*"). The term *qal* (Modern Hebrew: *kal*) means "light" in Hebrew, reflecting that the *binyan* is "light in additions" as it lacks prefixes and templatic gemination (Blau 2010). From this point onward, *binyan qal (pa'al)* will be referred to as *binyan qal*.

Additionally, verb inflection for person and tense is highly systematic. This systematic behavior in verbs contrasts with the irregularity observed in nouns, as demonstrated by plural endings. In present-tense verbs (and adjectives), the forms are consistent: the suffix *-im* always marks masculine (e.g., *kotvim* 'write PRESENT.MPL', *meχabkim* 'hug PRESENT.MPL'), and the suffix *-ot* always marks feminine (e.g., *kotvot* 'write PRESENT.FPL', *meχabkot* 'hug PRESENT.FPL'). In nouns, however, this regularity is not maintained: the suffix *-im*, typically masculine, can also be used for feminine nouns (e.g., *milim* 'words'), and the suffix *-ot*, typically feminine, can also be used for masculine nouns (e.g., *χalomot* 'dreams').

In addition to their morphological distinctions, verbs also exhibit unique phonological behaviors that distinguish them from nouns. One such behavior is the V~Ø alternation in CVCVC stems followed by V-initial suffixes (Bat-El 2008). In verbs, this alternation occurs in the second stem vowel, as in *gamal–gamla* 'reward, PAST.3MS–FM'. In nouns, however, the alternation is less systematic: it may occur in the first stem vowel when the vowel is *a* (e.g., *gamal–gmalim* 'camel(s)') or may not occur at all (e.g., *gamad–gamadim* 'dwarf(s)'). Although historical factors explain these differences, synchronically, V~Ø alternation illustrates how two CVCVC stems with identical prosodic structures can exhibit distinct phonological behaviors based on their lexical category.

The distinction between verbs and nouns is also evident in the acquisition of Hebrew, as shown by Handelsman et al. (2021). For instance, during certain phases of acquisition, differences in production length between nouns and verbs were observed, regardless of syllable count or stress patterns in the target words. Additionally, higher accuracy was noted in the production of final codas in verbs compared to nouns during a specific phase.

Therefore, it is plausible to predict that verbs will exhibit uniform patterns in relation to phonotactic restrictions, given their predictable and systematic nature. This study focuses on verbs precisely because their structured system provides a clear framework for researching phonotactic constraints. The findings of this research may serve as a foundation for future studies exploring whether similar constraints apply to other lexical categories, such as nouns and adjectives.

3.2.2.2. Focusing on regular verbs

The current analysis focuses on regular verbs, defined as those in which all three stem consonants surface phonetically throughout the templates of the inflectional paradigm (Zadok 2012). For example, the verb ***jaʃav*** 'sit PAST.3MSG' is considered irregular since its first stem consonant does not surface consistently across all forms of its inflectional paradigm, such as in ***eʃev*** 'sit FUT.1SG'. This definition deviates from the traditional one (*shlemim*). For instance, the verb ***ʃama(ʕ)*** 'hear PAST.3MSG', which historically ended with ʕ, is traditionally considered regular. However, in the current state of the language, the final ʕ has no phonetic realization, and synchronically only two consonants appear throughout the paradigm. Following Zadok (2012), such verbs are not classified as regular and were excluded from the analysis. It is worth noting, however, that traditionally defined *shlemim* verbs constitute the largest verbal group in Hebrew, comprising approximately 75% of the system.[8] Thus, the study targets the most representative patterns of the verbal system.

Also omitted from the analysis are verbs with bi- or quadri-consonantal stems (e.g., ***bana*** 'build PAST.3MSG', ***pitpet*** 'chatting PAST.3MSG'), as these may introduce additional morphological and phonological variables, such as deletion or reduplication, which could obscure the patterns under investigation. Verbs exhibiting consonant metathesis in the *hitpa'el* template were also excluded (e.g., *histagel* 'adapt PAST.3MSG'). This metathesis, triggered when the stem begins with a strident, blurs the distinction between the template's prefix *hit-* and the stem, potentially complicating the analysis.

Taken together, these considerations show that focusing on regular tri-consonantal verbs provides greater control over confounding factors while capturing the canonical

---

[8] According to R. Gadish, Academy of the Hebrew Language (personal communication, August 2025).

and most frequent verbal pattern in Hebrew. This defined scope establishes a solid empirical and theoretical foundation for future studies that may expand the investigation to the full range of Hebrew verb structures.

3.2.2.3. Filter out the passive

Verbs from five templates (*binyanim*) were included in the analysis: *qal* (e.g., *katav*), *nif'al* (e.g., *niχtav*), *hif'il* (e.g., *hiχtiv*), *pi'el* (e.g., *kitev*), and *hitpa'el* (e.g., *hitkatev*) (see §3.2.2.1). The two passive templates, *huf'al* (e.g., *huχtav*) and *pu'al* (e.g., *kutav*), were excluded from the analysis. This exclusion is supported by lexical-syntactic and phonological evidence, as discussed below.

From a lexical-syntactic perspective, passivization in Hebrew is a syntactic process. In other words, verbal passives are formed post-lexically; their derivation occurs after the verb's insertion into the syntax, and thus, they are not part of the speaker's mental lexicon (Horvath and Siloni 2008, Laks 2011). This contrasts with other thematic operations that occur within the lexicon itself (Reinhart and Siloni 2005).

Phonologically, Hebrew passive forms are characterized by a distinct vocalic pattern {u, a}, independent of prosodic structure. This suggests that the difference between the active form (listed in the lexicon) and the passive form results from an ablaut process (Bat-El 2003). The passive form of hiCCiC verbs (*hif'il*) is consistently huCCaC (*huf'al*), e.g., *hilbiʃ–hulbaʃ* 'dress PAST.3MSG–was dressed PAST.3MSG', *hiʁgiʃ–huʁgaʃ* 'feel PAST.3MSG–was felt PAST.3MSG'. Similarly, the passive form of CiCeC verbs (*pi'el*) is consistently CuCaC (*pu'al*), e.g., *supeʁ–supaʁ* 'tell PAST.3MSG–was told PAST.3MSG', *χibeʁ–χubaʁ* 'connect PAST.3MSG–was connected PAST.3MSG'.

Typically, Hebrew verbs in *qal* and *hitpa'el* do not undergo this ablaut process to form passives. Nevertheless, evidence of the process can also be found in these *binyanim*, reinforcing the argument for the phonological process underlying the formation of passive forms. For instance, sporadic colloquial forms in *hitpa'el* adopt a passive *hitpu'al* structure, such as *hitnudav* 'to be forced to volunteer PAST.3MSG' from *hitnadev* 'volunteer PAST.3MSG' and *hitputaʁ* 'to be forced to resign PAST.3MSG' from *hitpateʁ* 'resign PAST.3MSG' (Bat-El 2003). In *qal*, evidence of passive forms with a CuCaC pattern is found in Biblical Hebrew, e.g., *luqqaħ* 'was taken PAST.3MSG' (Genesis 3:23) from *laqaħ* 'take PAST.3MSG' (Schwarzwald, 2008, 2019; Blau 2010).[9] However, this formation no longer applies to Modern Hebrew.

Binyan nif'al also hosts passive verbs. However, several differences distinguish it from other passives, indicating that it is lexically encoded similarly to the other *binyanim*. First, its vocalic pattern differs from the {u, a} pattern found in the other passives. Second, *nif'al* verbs also carry non-passive meanings (Berman 1978; Doron 2003; Arad 2005; Schwarzwald 2008, 2019), such as *niχnas* 'enter PAST.3MSG', *niʁdam* 'fall asleep PAST.3MSG', and *nilχam* 'fight PAST.3MSG'. Therefore, verbs in *nif'al* were included in the analysis.

3.2.2.4. Analyzing paradigms

The analysis examines verb stems across the entire paradigm, providing a detailed understanding of consonantal behavior within the stems. This approach captures various distances between the consonants, including adjacent ones (e.g., *t* and *v* in *ka**tv**a* 'write PAST.3FSG') and non-adjacent ones (e.g., *t* and *v* in *ka**t**a**v*** 'write PAST.3MSG'). It also accounts for stop-fricative alternations due to spirantization, as seen in *k-χ* alternation in ***k**atav* 'write PAST.3MSG' vs. *ji**χ**tov* 'write FUT.3MSG'. Analyzing only stem consonants and focusing on a representative form (typically the third-person singular past) would fail to capture the full scope of phonological alternations across the paradigm.

Table (7) presents the paradigm of *binyan hif'il* as an example. Complete paradigms for all verb templates (*binyanim*) are available in Appendix A.

[9] Note that the gemination of the *q* in *luqqaħ* is secondary and not an inherent part of the template, unlike the gemination in *pi'el* forms, where it is integral to the template.

(7) *Binyan hif'il* paradigm, illustrated by the verb *hiχtiv* 'dictate PAST.3MSG'; stems are in bold

Past		Present[10]		Future	
1SG	*hi**χtav**ti*	MSG	*ma**χtiv***	1SG	*ʔa**χtiv***
2MSG	*hi**χtav**ta*	FSG	*ma**χtiv**a*	2MSG	*ta**χtiv***
2FSG	*hi**χtav**t*	MPL	*ma**χtiv**im*	2FSG	*ta**χtiv**i*
3MSG	*hi**χtiv***	FPL	*ma**χtiv**ot*	3MSG	*ja**χtiv***
3FSG	*hi**χtiv**a*			3FSG	*ta**χtiv***
1PL	*hi**χtav**nu*			1PL	*na**χtiv***
2MPL	*hi**χtav**tem*			2PL	*ta**χtiv**u*
2FPL	*hi**χtav**ten*			3PL	*ja**χtiv**u*
3PL	*hi**χtiv**u*			Infinitive	*leha**χtiv***

From the paradigm above, it can be observed that in *binyan hif'il,* there are two distinct stems: CCaC, as in *χtav* (e.g., *hi**χtav**ti*, *hi**χtav**nu*), and CCiC, as in *χtiv* (e.g., *hi**χtiv***, *ma**χtiv**a*).

Imperative forms (e.g., *haχtev*) were excluded from the analysis, as they are not commonly used in Modern Hebrew; future forms often replace the imperative, particularly in spoken language (Avinery 1965, Peretz 1971).[11] The historical (Biblical) plural feminine form in the second and third person (*taχtevna*) was also excluded, as it is rarely used nowadays. In both spoken and formal-normative usage, masculine forms are often employed for the feminine contexts as well (see Avinery 1965, among others).

3.2.2.5. Looking from a synchronic perspective

The analysis adopts a synchronic perspective, focusing on the contemporary pronunciation of the consonants as spoken by (most) Modern Hebrew speakers (see §2.3). This approach aims to capture the most accurate representations of present-day

---

[10] Present forms are not specified for person, only for number and gender.

[11] Some imperative forms (e.g., *bo* 'come IMP.2MSG', *kχi* 'take IMP.2FSG', *leχu* 'go IMP.2PL') are commonly used in spoken Hebrew. However, all these frequent imperatives belong to weak verb classes rather than regular verbs and are therefore beyond the scope of the present analysis.

Hebrew phonology, regardless of historical sound changes. The synchronic perspective reflects the current stage of the language and facilitates a comparison between the corpus analysis and the experimental results, which reflect the speakers' phonological system (see Chapter 4).

Along this line, the selection of regular verbs was based on their synchronic status (see §3.2.2.2), with normative or historical forms replaced by their spoken counterparts.

- Consonants were analyzed based on their synchronic pronunciation, as spoken by most Modern Hebrew speakers. Thus, the historical $t^{ʕ}$ (ט) is treated as *t*, the historical *q* (ק) as *k*, and the historical *ħ* (ח) as *x* (see Bolozky 1997).
- In some verb forms, *p* was replaced by *f*, reflecting the more common pronunciation of the verb (see Adam 2002; Albert 2014, 2019). For example, *jit**f**os* 'to catch FUT.3MSG' replaced the normative *jit**p**os*, and *nit**f**as* 'to be caught PAST.3MSG' replaced the normative *nit**p**as*.
- The same applies to denominative verbs, where speakers aim to maintain maximum faithfulness between the noun and the verb (Bat-El 1994). For example, *ki**x**ev* 'star PAST.3MSG' (expected normative form: *ki**k**ev*), derived from the noun *ko**χ**av* 'star'; *hin**fi**ʃ* 'animate PAST.3MSG' (expected normative form: *hin**pi**ʃ*), derived from the noun *ne**f**eʃ* 'spirit'; ***f**iʃel* 'screw up (slang) PAST.3MSG', from the noun ***f**aʃla* 'flop (slang)', originally from Arabic.
- Normatively, the letter ח (historical *ħ*, now *χ*) is followed by a vowel in certain paradigm forms (e.g., *ʦa**χa**ka* 'laugh PAST.3FSG'). However, most speakers do not pronounce this vowel in regular speech. Therefore, this vowel was excluded from the analyzed forms (e.g., *ʦa**χ**ka*).
- Normatively, verbs in *qal*, past tense, second person plural, have no vowel after the first consonant (e.g., ***k**tavtem* 'write PAST.2MPL'). However, almost all speakers pronounce a vowel after the first consonant in regular speech. Therefore, this vowel was included in the analyzed forms (e.g., ***ka**tavtem*).

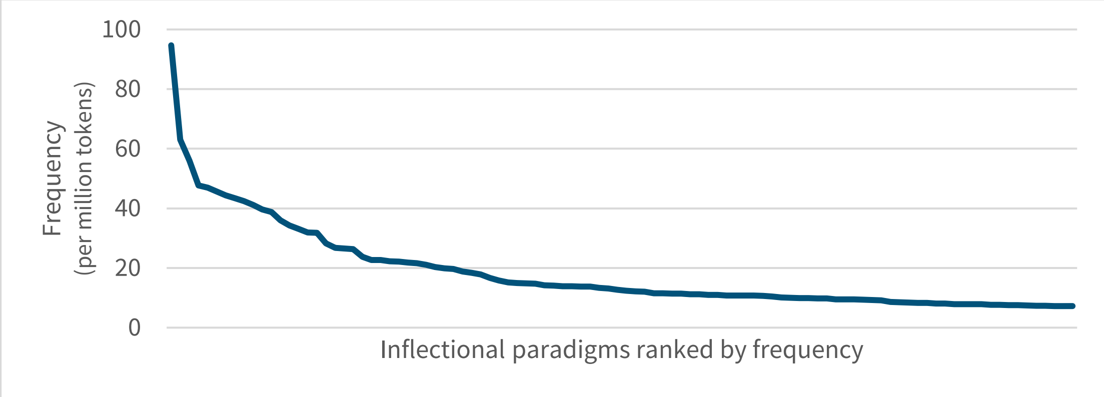
3.2.2.6. Types or tokens?

A key methodological issue in corpus analysis is whether to rely on type or token frequency. In a type-frequency approach, all forms are treated equally, while in a token-frequency approach, each form is weighted according to its frequency in the corpus.

The current study adopts a type-frequency analysis. Previous studies have argued that pattern strength, defined as how strongly a pattern is entrenched in a speaker's mental lexicon, is determined by type frequency rather than token frequency (Bybee 1995, 2001; Pierrehumbert 2001; Hay et al. 2004, among others). Computational evidence also supports this view. Albright and Hayes (2003) showed that models based on type frequency align more closely with experimental data than those based on tokens. Goldwater et al. (2006) further justified the role of type frequencies in formal analyses of natural language, and more recent work emphasizes both theoretical and practical advantages of type-based approaches (Pimentel et al. 2020).

In addition to previous claims and findings from the literature, the corpus data used in the current study highlights the problem in token-frequency analysis. The most common inflectional paradigm is that of the verb *kibel* 'receive PAST.3MSG', with 102,402 occurrences (94.69 per million). The second most common is *χaʃav* 'think PAST.3MSG', with 68,157 occurrences (63.03 per million), showing a considerable drop in frequency. The graph in (8) illustrates the frequencies of the 100 most frequent inflectional paradigms in the corpus.

(8) Frequencies of the top 100 inflectional paradigms in the corpus



The sharp drop in word frequencies aligns with Zipf's Law (Zipf 1936, 1949), which describes a common pattern in language data where frequency decreases rapidly as rank increases. If the analysis were based on token frequency, it would disproportionately emphasize a small group of high-frequency stems. This tendency is evident in the present corpus, where a few verbs dominate the tokens and obscure the overall distribution of phonotactic patterns across the lexicon. Accordingly, the analysis relies on type frequency, which offers a more balanced representation of the data.

3.2.3. Preparing the corpus for analysis

Preparing the data for analysis involved several steps, including extracting and coding the relevant verbs from the corpus. First, all verb forms were extracted from the corpus (see §3.2.2.1), and the list of verbs was manually checked to ensure that only genuine verbs were included. Recall that the dataset includes only regular verbs from five *binyanim*, excluding the passive templates (see §3.2.2.2, §3.2.2.3).

Since the corpus data was in Hebrew orthography, the next involved transcribing all extracted forms into IPA, following the decisions outlined in §3.2.2.5. The transcription process was automated, followed by manual corrections to ensure accuracy.

For each inflectional paradigm, the complete set of stem forms was compiled based on the paradigm of its *binyan* (see §3.2.2.4 and Appendix A). All in all, **1833** inflectional paradigms containing **6892** stems were analyzed. A detailed classification of the stems by *binyanim* is presented in Table (9).

(9) Number of inflectional paradigms and stems by *binyanim*

		Inflectional Paradigms		Stems	
binyan	Example	Occurrences	%	Occurrences	%
qal	*katav*	413	22.53%	2568	37.26%
nif'al	*niχtav*	228	12.44%	899	13.04%
hif'il	*hiχtiv*	274	14.95%	571	8.28%
pi'el	*kitev*	619	33.77%	2137	31.01%
hitpa'el	*hitkatev*	299	16.31%	717	10.40%
Total		**1833**	100%	**6892**	100%

Next, the 6892 stem forms were divided into four subsets by phonological environments:

(10) Phonological environments

Adjacent CC	a.	C_1C_2	Adjacent first and second stem consonants (e.g., ***χt**iv,* as in *hi**χt**iv* 'dictate PAST.3MSG');
	b.	C_2C_3	Adjacent second and third stem consonants (e.g., *ka**tv***, as in *ka**tv**a* 'write PAST.3FSG');
Nonadjacent CC	c.	C_1VC_2	Nonadjacent first and second stem consonants (e.g., ***kat**av*, as in ***kat**av* 'write PAST.3MSG')
	d.	C_2VC_3	Nonadjacent second and third stem consonants (e.g., *ka**tav***, as in *ka**tav*** 'write PAST.3MSG').

Some stems appeared in more than one data subset. For example, *χtiv* was listed in group (4a) for the adjacent pair *χt*, and in group (4d) for the nonadjacent pair *tiv*. Within each subset, consonant-pair frequency was calculated by summing occurrences across all stem forms. This yielded frequency tables of consonant co-occurrences in each tested phonological environment, which then served for the statistical analyses detailed in the following subsection.

3.3. Results

The results section provides an analysis of the distributions of consonant pairs in the corpus. First, statistical tests were conducted to examine whether significant differences exist among the four phonological environments. Results from paired-sample t-tests indicated significant differences based on adjacency between consonants (C_1C_2 vs. C_1VC_2: $t(323) = 13.19$, $p < 0.0001$; C_2C_3 vs. C_2VC_3: $t(323) = 13.45$, $p < 0.0001$). However, no significant differences were found based on word position, regardless of adjacency (C_1C_2 vs. C_2C_3: $t(323) = 0.27$, $p = 0.78$; C_1VC_2 vs. C_2VC_3: $t(323) = 0.12$, $p = 0.90$). Recall that Hebrew generally prohibits identical consonants in the C_1VC_2 position but allows them in the C_2VC_3 position (Greenberg 1950; McCarthy 1979, 1981, 1986, see also §2.2). Nevertheless, no significant difference was found between C_1VC_2 and C_2VC_3.

Hence, the first part of the analysis considered adjacency between consonants but not word position; that is, C_1C_2 and C_2C_3 were combined into CC, and C_1VC_2 and C_2VC_3 were combined into CVC. In total, the analysis included **3872** CC pairs and **9912** CVC combinations. Note that the experiments (see Chapter 4) tested each phonological environment separately to provide more specific observations. Accordingly, the sections of the corpus analysis focusing on phonological constraints are also divided into more specific phonological environments.

The results section is organized as follows: §3.3.1 presents a statistical analysis comparing the observed and expected frequencies; §3.3.2 discusses the most frequent pairs found in the corpus, while §3.3.3 addresses pairs that are entirely absent; §3.3.4 and §3.3.5 provide a detailed examination of the relevant co-occurrence constraints: §3.3.4 focuses on the OCP, and §3.3.5 explores the SCL and SDP.

3.3.1. Observed and expected frequencies

First, a comparison between the observed and expected frequencies was conducted to confirm the assumption that co-occurrence restrictions exist on consonant pairs in the Hebrew lexicon. Previous studies on Hebrew consonantal roots (Greenberg 1950; Frisch et al. 2004, among others) and verbal stems (Yeverechyahu 2014, 2019) provide a solid foundation for expecting such restrictions to emerge in the current study.

Chi-square tests were conducted separately for CC and CVC, considering the frequency of each tested consonant in the corpus. The expected frequencies represent what would occur if single consonants were randomly combined into pairs. The results reveal highly significant differences between the observed and expected frequencies in both cases (CC: $\chi^2(289) = 3914$, $p < 0.0001$; CVC: $\chi^2(289) = 8328$, $p < 0.0001$), consistent with a lexicon governed by co-occurrence restrictions.

3.3.2. Most frequent pairs

The most frequent pairs were analyzed, focusing on the top 5% of pair types (16 pairs) out of a total of 324 types. This cutoff was selected to highlight the most dominant patterns while ensuring the analysis remained concise and manageable.

Table (11) presents the most frequent CC pairs and Table (12) the most frequent CVC combinations.

(11) Most frequent **CC** (5% of CC types)

CC	f(CC)	%	PoA	OCP	MoA	SonD
χʁ	92	2.38%	GUT-GUT	✗	FL	+2
tʁ	65	1.68%	CORo-GUT	✓	StL	+3
χl	64	1.65%	GUT-CORs	✓	FL	+2
kʁ	61	1.58%	DOR-GUT	✓	StL	+3
fʁ	59	1.52%	LAB-GUT	✓	FL	+2
ʁk	57	1.47%	GUT-DOR	✓	LSt	-3
χt	52	1.34%	GUT-CORo	✓	FSt	-1
ʃk	51	1.32%	CORo-DOR	✓	StrSt	NA
χʃ	50	1.29%	GUT-CORo	✓	FStr	NA
ʃl	46	1.19%	CORo-CORs	✓	StrL	NA
sʁ	46	1.19%	CORo-GUT	✓	StrL	NA
ʦʁ	45	1.16%	CORo-GUT	✓	StrL	NA
ʁt	44	1.14%	GUT-CORo	✓	LSt	-3
lχ	44	1.14%	CORs-GUT	✓	LF	-2
tχ	43	1.11%	CORo-GUT	✓	StF	+1
ʁχ	42	1.08%	GUT-GUT	✗	LF	-2

CC = the consonant pair; f(CC) = the frequency of the pair in the tested corpus; PoA = place of articulation; OCP = whether the pair satisfies the OCP (✓) or violates it (✗); MoA = manner of articulation; SonD = sonority distance level. See §2.3 for details regarding the classification by place and manner of articulation.

(12) Most frequent **CVC** (5% of CVC types)

CVC	f(CVC)	%	PoA	OCP
*ʁ*V*χ*	167	3.38%	GUT-GUT	✗
*k*V*ʁ*	138	2.80%	DOR-GUT	✓
*χ*V*ʁ*	127	2.57%	GUT-GUT	✗
*χ*V*l*	109	2.21%	GUT-CORs	✓
*ʁ*V*k*	108	2.19%	GUT-DOR	✓
*t*V*ʁ*	106	2.15%	CORo-GUT	✓
*l*V*χ*	103	2.09%	CORs-GUT	✓
*p*V*ʁ*	100	2.03%	LAB-GUT	✓
*t*V*χ*	99	2.01%	CORo-GUT	✓
*ʁ*V*t*	97	1.97%	GUT-CORo	✓
*χ*V*t*	97	1.97%	GUT-CORo	✓
*k*V*χ*	96	1.94%	DOR-GUT	✓
*ʃ*V*l*	96	1.94%	CORo-CORs	✓
*ʃ*V*χ*	96	1.94%	CORo-GUT	✓
*χ*V*ʃ*	92	1.86%	GUT-CORo	✓
*k*V*l*	91	1.84%	DOR-CORs	✓

Both tables demonstrate that the most frequent consonant pairs satisfy the OCP. However, two pairs, ***χʁ*** and ***ʁχ***, violate this principle, regardless of whether the consonants are adjacent (CC) or separated by a vowel (CVC).

This distribution can be explained from a historical perspective. Some scholars suggest that the Hebrew rhotic *r*, like its counterpart in other Semitic languages, was originally an alveolar trill or tap, classifying it as a coronal consonant rather than a guttural one (Blau 2010, Meloni 2021, among others, see also §2.3). Based on this assumption, and considering that most Modern Hebrew verbs have Biblical origins, the combination of *χ* and *ʁ* was likely not originally problematic, as it did not violate the OCP.

While the situation regarding the OCP is straightforward, the status of the sonority constraints, the SCL and the SDP, is more complex, as seen in Table (13). It is important to note that the SCL and SDP apply only to adjacent consonants (CC).

(13) Sonority distance (SonD) values of the most frequent CC

Sonority Scale		SCL		SDP	
SonD	Occurrences	SonD	Total occurrences	SonD	Total occurrences
-4	0	–	5	±4	0
-3	2	+	6	±3	4
-2	2	0	0	±2	5
-1	1	NA	5	±1	2
0	0			0	0
+1	1			NA	5
+2	3				
+3	2				
+4	0				
NA	5				

The table shows that five pairs satisfy the SCL with a negative sonority distance (C_1 more sonorous than C_2), while six pairs violate it with a positive sonority distance (C_1 less sonorous than C_2). This suggests that the most frequent pairs are not sensitive to the SCL. None of the most frequent CC pairs exhibit a plateau (SonD = 0). Additionally, five pairs are marked as not applicable (NA) due to the involvement of a strident, which behaves atypically within the sonority hierarchy (see §2.2).

Regarding the SDP, two pairs have a sonority distance (SonD) value of ±1, five have a value of ±2, and four have a value of ±3, while none exhibit a plateau (SonD = 0). This distribution suggests a preference for larger sonority distances. Additionally, no pairs with a SonD value of ±4 appear among the most frequent pairs, likely because this value involves a glide (*j*), which is rare in regular Hebrew verbs. Although these results may suggest that the SDP influences the corpus, the overall analysis does not consistently support this tendency, as detailed in §3.3.5.

3.3.3. Zero occurrences

Linguistic theory is expected to account not only for the patterns present in the lexicon but also for the absence of certain combinations. Following this premise, this subsection focuses on cases that are entirely absent, i.e., instances with zero occurrences.

In the CC list, which consisted of 324 pair types, 97 (30%) pairs exhibit zero occurrences in the corpus (see Appendix B). Among these are 18 pairs with identical consonants (*mm*, *dd*, *gg*, etc.), which are prohibited in Modern Hebrew. When such sequences arise in the stem due to morphological reasons, Hebrew speakers insert an epenthetic vowel between the consonants, resulting in a surface form of CVC (e.g., *mad**e**da* 'measure PAST.3FSG' cf. *katva* 'write PAST.3FSG').

Of the remaining 79 CC pairs on the zero-occurrence list, 43 (54%) violate the OCP. This suggests that avoiding OCP violations accounts for over half of the cases on the list, reinforcing the significant influence of this constraint on the lexicon. The situation regarding SCL and SDP, however, is less clear, as illustrated in Table (14).

(14) Sonority distance (SonD) values for zero-occurrence CC pairs

Sonority Scale		SCL		SDP	
SonD	Occurrences	SonD	Total occurrences	SonD	Total occurrences
-4	2	–	13	±4	8
-3	1	+	22	±3	4
-2	3	0	28	±2	7
-1	7	NA	34	±1	16
0	28			0	28
+1	9			NA	34
+2	4				
+3	3				
+4	6				
NA	34				

The table shows that pairs with zero occurrences are distributed across the full range of sonority distance values, including both violations and satisfactions of the SCL. Specifically, 13 pairs satisfy the SCL with a negative sonority distance, while 22 pairs violate it with a positive sonority distance. Additionally, 34 pairs are marked as not applicable (NA) due to the involvement of a strident. Moreover, no clear generalization can be made regarding the SDP, as absent pairs are found at every sonority distance value, ranging from 0 to ±4. Therefore, based on this distribution, there is no evidence to suggest that either the SCL or the SDP is responsible for the absence of these pairs in the corpus.

In the CVC list, which consisted of 324 sequence types, 50 (15%) consonant pairs exhibit zero occurrences in the corpus (see Appendix C). Among these, 36 pairs (72%) violate the OCP. This finding further reinforces the influence of OCP violations on the lexicon. Notably, the list of zero-occurrence sequences includes three pairs of identical consonants in the C_2VC_3 position: *b*V*b*, *p*V*p*, and *j*V*j*. As discussed in §2.2, these do not violate the OCP but instead represent a single consonant linked to two C positions, which Hebrew permits in the C_2-C_3 position. The absence of *b*V*b* and *p*V*p* can be attributed to post-vocalic spirantization, while *j*V*j* is absent due to the low frequency of *j* in Hebrew verbs.[12]

3.3.4. OCP

After examining the most frequent sequences and those entirely absent from the lexicon, the focus shifts to the constraints that form the core of this study, starting with the OCP. The analysis focuses on more specific phonological environments, aligning with the framework of the experiments.

The analysis begins with environments where the consonants are adjacent: C_1C_2 and C_2C_3. In the C_1C_2 environment, **92.94%** of the 1956 pairs satisfy the OCP, while 7.06% (138 pairs) violate it. Of these violations, 3.73% (73 pairs) consist of guttural consonants (*χʁ*, *ʁχ*), and 3.32% (65 pairs) consist of coronal obstruents. Among the coronal obstruents, 83.08% (54 pairs) consist of a stop and a fricative, meaning the consonants differ in manner of articulation.

[12] See footnote 6 regarding spirantization in Modern Hebrew.

Similar results are found in the C_2C_3 environment, where **92.95%** of the 1,916 pairs satisfy the OCP. Of the 7.05% that violate it (135 pairs), 3.18% (61 pairs) consist of guttural consonants (*χʁ*, *ʁχ*), 2.77% (53 pairs) consist of coronal obstruents, 0.89% (17 pairs) consist of coronal sonorants, and 0.21% (4 pairs) consist of labial consonants. As in C_1C_2, almost all coronal obstruent pairs (92.45%, 49 pairs) consist of a stop and a fricative, while only four pairs (7.55%) consist of two stops. The results are summarized in Table (15).

(15) OCP in C_1C_2 and C_2C_3

	C_1C_2		C_2C_3	
OCP satisfaction	1818	**92.94%**	1781	**92.95%**
OCP violation				
LAB *p, b, f, v, m*	0	0%	4	0.21%
CORo *t, d, ʦ, s, z, ʃ*	65	3.32%	53	2.77%
CORs *n, l, j*	0	0%	17	0.89%
DOR *k, g*	0	0%	0	0%
GUT *χ, ʁ*	73	3.73%	61	3.18%
Total	138	7.06%	135	7.05%

As noted in §2.3 and §3.3.2, some scholars suggest that the Hebrew *ʁ* was historically a coronal sonorant rather than a guttural. Given the frequent occurrence of the *x-ʁ* pair in the corpus, an additional analysis was conducted to examine how the results would change if *ʁ* were classified as a coronal sonorant. Under this assumption, **96.57%** of the C_1C_2 pairs would satisfy the OCP, and in the C_2C_3 environment, **95.35%** of the pairs would satisfy the constraint, with no guttural pairs in either environment.

Next, an analysis was conducted on environments where a vowel separates the consonants: C_1VC_2 and C_2VC_3. These environments exhibit a similar pattern regarding the OCP. In the C_1VC_2 environment, **93.48%** of the 4936 pairs satisfy the OCP, while violations occur in 6.52% of cases (322 pairs). Among these violations, 3.02% involve pairs of non-identical guttural consonants, and 2.94% consist of non-identical coronal obstruents. A small number of cases include labial pairs and coronal sonorant pairs (0.26% and 0.30%, respectively). Within the coronal obstruents, 76.55% consist of a stop and a fricative, meaning the consonants differ in manner of articulation.

Similarly, in the C_2VC_3 environment, **93.48%** of the 4976 pairs satisfy the OCP, while violations occur in 6.43% of cases (320 pairs). Of these violations, 3.14% consist of guttural consonants, and 2.21% consist of coronal obstruents. A few instances of labial pairs and coronal sonorant pairs also appear (0.54% each). Among the coronal obstruents, 94.55% consist of a stop and a fricative. Additionally, 10.43% of the C_2VC_3 pairs consist of identical C_2 and C_3. As discussed in §2.2 and §3.3.2, these are not considered OCP violations but rather cases where a single consonant is copied and linked to two consonantal positions. The results are summarized in Table (16).

(16) OCP in C_1VC_2 and C_2VC_3

	C_1VC_2		C_2VC_3	
OCP satisfaction	4614	**93.48%**	4656	**93.57%**
Of which identical C_2VC_3			519	10.43%
OCP violation				
LAB *p*, *b*, *f*, *v*, *m*	13	0.26%	27	0.54%
CORo *t*, *d*, *ʦ*, *s*, *z*, *ʃ*	145	2.94%	110	2.21%
CORs *n*, *l*, *j*	15	0.30%	27	0.54%
DOR *k*, *g*	0	0%	0	0%
GUT *χ*, *ʁ*	149	3.02%	156	3.14%
Total	322	6.52%	320	6.43%

3.3.5. SCL and SDP

This subsection completes the analysis of phonotactic constraints in the corpus by examining SCL and SDP. These constraints apply only to adjacent consonants, meaning the analysis is based solely on the CC lists and not the CVC lists. Additionally, pairs involving stridents were excluded (see §2.2). As with the OCP analysis, the results are presented by word positions: C_1C_2 and C_2C_3.

The data do not show a strong tendency to satisfy or to violate the SCL in either environment. The distribution of pairs that satisfy the SCL and those that violate it is relatively balanced, with a small group displaying a sonority plateau. Specifically, in the C_1C_2 environment, out of 1160 pairs, 44.4% (515 pairs) violate the SCL, 39.05% (453 pairs) satisfy it, and 16.55% (192 pairs) form a plateau. In the C_2C_3 environment, out of 1301 pairs, 47.19% (614 pairs) violate the SCL, 41.58% (541 pairs) satisfy it, and 11.22% (146 pairs) form a plateau. These results are summarized in Table (17).

(17) SCL in C_1C_2 and C_2C_3

		C_1C_2		C_2C_3	
SCL violation	SonD > 0	515	44.4%	614	47.19%
Plateau	SonD = 0	192	16.55%	146	11.22%
SCL satisfaction	SonD < 0	453	39.05%	541	41.58%
	Total	1160		1301	

When examining sonority distances, no clear pattern emerges, as shown in Table (18) below. While an initial look at the most frequent pairs suggests that the SDP might have some influence (§3.3.2), a detailed analysis reveals no consistent preference for larger sonority distances over smaller ones. Consequently, there is no strong evidence to support the SDP as active in the corpus.

(18) SDP in C_1C_2 and C_2C_3

SonD	C_1C_2		C_2C_3	
-4[13]	0	0%	26	2.00%
-3	105	9.05%	145	11.15%
-2	103	8.88%	177	13.60%
-1	245	21.12%	193	14.83%
0	192	16.55%	146	11.22%
+1	131	11.29%	217	16.68%
+2	246	21.21%	198	15.22%
+3	138	11.90%	199	15.30%
+4	0	0%	0	0%
Total	1160		1301	

SonD	C_1C_2		C_2C_3	
±4	0	0.00%	26	2.00%
±3	243	20.95%	344	26.44%
±2	349	30.09%	375	28.82%
±1	376	32.41%	410	31.51%
0	192	16.55%	146	11.22%
Total	1160		1301	

3.4. Discussion

The corpus analysis reveals clear differences in how the examined phonotactic constraints – OCP, SCL, and SDP – are satisfied within the Hebrew lexicon.

3.4.1. OCP: strong evidence of influence

The results demonstrate that the OCP is strongly satisfied in the Hebrew lexicon. Across all environments, both adjacent and non-adjacent consonants, the majority of consonant pairs satisfy the OCP, with only a small percentage of violations. Notably, when violations occur, they consistently involve consonants that differ in manner of articulation. Most violations involve a combination of a stop and a fricative, making the consonants distinct enough to reduce their similarity, even though they violate the OCP.

[13] As noted in §3.3.2, the low number of pairs with a SonD value of ±4 is due to the rarity of the glide *j* in regular Hebrew verbs.

Interestingly, the frequent occurrence of the *χ-ʁ* pair presents a challenge to the OCP. Synchronically, this pair, consisting of two guttural consonants, violates the principle. However, a historical perspective provides an explanation for this anomaly. Some scholars propose that the Hebrew rhotic was historically a coronal consonant, likely an alveolar trill or tap, *r* (see §2.3). Under this assumption, the pair would not have violated the OCP at the time, as the consonants had different places of articulation. In Modern Hebrew, the rhotic has shifted to a guttural, aligning phonetically with other gutturals like *χ*. This shift now causes the combination to violate the OCP. Although Modern Hebrew speakers may not be aware of this historical change, their phonological system treats this combination as an exception to the OCP, reflecting the outcome of historical developments rather than sensitivity of the constraint itself to historical origin.

3.4.2. SCL and SDP: lack of influence

In contrast to the strong evidence for OCP effect, the results do not provide evidence that the SCL or the SDP are active in the Hebrew lexicon. In both C_1C_2 and C_2C_3 environments, the distribution of consonant pairs shows no clear preference for satisfying the SCL. The percentages of pairs that satisfy and violate the SCL are relatively balanced, with violations occurring frequently and no strong tendency toward either satisfaction or violation.

Similarly, there is no evidence to suggest that the SDP plays a role in the Hebrew lexicon. The data show no consistent preference for pairs with larger sonority distances over those with smaller ones. Pairs with a range of sonority distance values (from ±1 to ±3) appear throughout the corpus, with violations occurring across the spectrum and no clear patterns emerging. While an initial observation of the most frequent pairs hinted at a possible influence of the SDP, a detailed analysis revealed no definitive evidence of its activity.

3.4.3. Conclusion

The corpus analysis reveals a clear variance in the relevance of the phonotactic constraints under study. While the OCP is highly satisfied in the Hebrew lexicon, the SCL and SDP show no discernible effect and are frequently violated.

The next chapter examines whether Hebrew speakers exhibit similar sensitivity to these constraints in psycho-phonological experiments. These experiments will investigate whether speakers' behavior reflects the status of the constraints in their lexicon or whether their phonological preferences extend beyond the lexicon, influenced by internal factors such as universal constraints that are inactive within their lexicon.

Chapter 4

Psycho-Phonological Experiments

4.1. Introduction

Chapter 3, the corpus analysis, examined the status of three phonotactic constraints: OCP, SCL, and SDP, using data from the *heTenTen* corpus. The definitions of these constraints were presented in §2.2. The analysis revealed a clear distinction between the constraints: while the OCP is highly satisfied in the Hebrew lexicon, the SCL and SDP show no discernible effect and are frequently violated.

This chapter shifts focus to the mental lexicon of Hebrew speakers, specifically examining the *phonological system in their minds*. Three psycho-phonological experiments were conducted to investigate the extent to which these constraints influence speakers' phonological systems:

a. **Experiment A (§4.2):** Evaluates OCP and SCL violations based on speakers' lexical judgments of nonce verbs.

b. **Experiment B (§4.3):** Evaluates OCP and SCL violations by measuring accuracy and reaction times in identifying nonce verbs.

c. **Experiment C (§4.4):** Evaluates SCL violations across two sonority distances and examines the role of SDP in shaping participants' preferences, based on their lexical judgments of nonce verbs.

In all three experiments, nonce verbs were used as stimuli to directly test violations of phonotactic constraints. Experiments A and B focused on OCP and SCL by comparing nonce verbs that violated these constraints with well-formed (WF) nonce verbs. Although SCL is a scalar constraint reflecting varying sonority distances between consonants, it was treated as binary in Experiments A and B to provide a straightforward test of whether the constraint plays any role in Hebrew. Experiment C then examined sonority distance in greater detail, focusing on the SDP, which directly represents the degree of sonority difference between consonants.

The experiments tested constraint violations in various phonological environments. OCP violations were examined in four environments:

a. **C_1C_2:** adjacent first and second stem consonants (e.g., *hi**kg**ina*)
b. **C_2C_3:** adjacent second and third stem consonants (e.g., *ba**td**a*)
c. **C_1VC_2:** nonadjacent first and second stem consonants (e.g., ***kag**an*)
d. **C_2VC_3:** nonadjacent second and third stem consonants (e.g., *bi**ted***)

SCL and SDP, relevant only for adjacent consonants, were tested in two environments:

a. **C_1C_2:** adjacent first two stem consonants (e.g., *hi**fn**iχa*)
b. **C_2C_3:** adjacent second two stem consonants (e.g., *da**gf**a*)

This study examines each environment independently, avoiding direct comparisons between environments (e.g., OCP violations in C_1C_2 vs. C_2C_3). Because Hebrew verbs are limited to specific templates (§3.2.2.1), testing various phonological environments requires different templates and genders (see §4.2.1.2 and §4.3.1.2 for stimuli). These templates may influence the results independently of phonotactic constraints, so cross-environment comparisons may not accurately reflect phonotactic effects.

The hypotheses are as follows:

a. If speakers' phonological systems are shaped solely by lexical frequencies, they are expected to exhibit sensitivity to OCP violations (reflecting the lexicon's sensitivity to this constraint) but not to SCL or SDP violations (as the lexicon shows no evidence of sensitivity to these constraints).
b. If speakers' phonological systems are shaped by both lexical frequencies and universal constraints, they are expected to exhibit some sensitivity to violations of all three universal constraints, including those that are not active in their lexicon.

Experiments A and B demonstrate that Hebrew speakers are sensitive to OCP violations, reflecting alignment with both the lexicon and universal constraints. However, the status of the SCL proves more complex. Experiment C investigated the SCL in relation to sonority distance. Nonce verbs were presented with consonant pairs that either satisfied or violated the SCL across two sonority distances: two obstruents (±1) or an obstruent and a sonorant (±2). The findings offer insights into the influence of the SCL and SDP on phonological preferences and their distinct roles within the phonological system.

4.2. Experiment A: Lexical Judgment

Experiment A examined the role of the OCP and SCL in the phonological system of Hebrew speakers by asking native speakers to provide lexical judgments for nonce verbs in a rating task. The nonce verbs that violated either OCP or SCL were compared to well-formed (WF) nonce verbs that satisfy phonotactic constraints. Violations were tested across four distinct phonological environments.

4.2.1. Method

4.2.1.1. Participants

The participants included 60 native Hebrew speakers (42 women and 18 men, aged 20–44, M = 30, SD = 5.8). None of the participants had any native language besides Hebrew, nor did they report any hearing problems. Additionally, none of them had studied Linguistics or Hebrew Language at a university level. Participants were entered into a prize draw for a breakfast voucher as compensation.

4.2.1.2. Stimuli

The stimuli consisted of nonce verbs containing tested consonant pairs that either violated the OCP or the SCL (see §2.2) or did not violate any known phonotactic constraints:

- **OCP** violation (Obligatory Contour Principle): The consonants in each tested pair shared place and manner of articulation (e.g., ***hitdi**χa*, *mi**keg***). As discussed in §3.4.1, the Hebrew lexicon largely satisfies the OCP.
- **SCL** violation (Syllable Contact Law): In each tested pair, the first consonant was less sonorous than the second. The Hebrew lexicon shows no discernible effect of the SCL (see §3.4.2). In this experiment, the sonority distance between consonants was set to ±1, resulting in pairs such as fricative-nasal (e.g., *hi**fn**iχa*) or stop-fricative (e.g., *da**gf**a*). Stridents were excluded due to their distinct behavior regarding sonority constraints (see §2.2). Furthermore, the two consonants did not share a place of articulation eliminating potential OCP violations.
- **WF** (Well-Formed): The consonants in each tested pair did not violate any known phonotactic constraints, classifying them as well-formed (e.g., *hi**ʁb**ila*, *ma**ʁg**a*). Specifically, the pairs neither shared a place of articulation (thus avoiding OCP violations) nor involved adjacent consonants where the first was less sonorous than the second (thus avoiding SCL violations).

To control for frequency effects, all tested pairs had low frequency in the Hebrew lexicon, as indicated by the corpus analysis. Pairs that violated the OCP were entirely absent from the lexicon, while all other pairs had low frequencies, with TPM ≤ 0.06.[14]

[14] The *Transitional Probability Model* (TPM; Poletiek and Wolters 2009) measures the conditional probability of one element following another in a sequence. Mathematically, it is formally expressed as ***P(x|y) = occurrences of x,y / occurrences of x***, where ***x*** is the first element and ***y*** is the second element. This metric reflects the likelihood of y occurring immediately after x within a given dataset.

The stimuli consisted of nonce verbs with three stem consonants. The tested consonant pairs appeared in either the C_1-C_2 position (e.g., ***tad**aχ*; tested pair in bold) or the C_2-C_3 position (e.g., *bi**ted***), with an untested third consonant included for each tested pair. To minimize its influence, the untested consonant did not share a place of articulation with the tested consonants, thereby avoiding OCP violations. Exceptions were made for the forms *hi**ʁz̧**iza* and ***ʁaz̧**az*, where C_3 was identical to C_2 (a permissible pattern in Hebrew; see §2.2). The untested consonant also did not result in SCL violations, as the SCL applies only to adjacent consonants, and a vowel always separated the untested consonant from the nearest tested one (e.g., C_2VC_3 when C_3 was untested). Furthermore, the sequence of the untested consonant and the nearest tested consonant had a medium-high frequency in the Hebrew lexicon (TPM ≥ 0.08).

All in all, 30 distinct consonant triplets were generated based on three types of tested pairs (OCP violations, SCL violations, and WF) × 2 C-positions (C_1-C_2 and C_2-C_3) × 5 items per group. All consonants adhered to Hebrew's spirantization rules, and none of the triplets corresponded to any existing Hebrew verbs. In C_2-C_3 position, C_3 was not *t*, as *t* in the CaCCa template (used for the stimuli) could be interpreted as a feminine marker rather than a stem consonant (e.g., *ʁat͡sta* 'want PAST.3FSG', *banta* 'build PAST.3FSG'). The complete list of consonant triplets appears in (19).

(19) Consonant triplets (tested pairs in bold)[15]

C_1-C_2	OCP	**td**χ, **td**k, **kg**d, **kg**n, **kg**m
	SCL	**fn**χ, **vn**k, **vn**n, **vn**χ, **χn**v
	WF	**ʁb**l, **ʁb**ʃ, **lg**ʁ, **ʁs**f, **ʁz**z
C_2-C_3	OCP	b**td**, m**td**, m**kg**, ʁ**kg**, t**kg**
	SCL	d**gf**, d**gv**, g**dv**, d**gχ**, p**χn**
	WF	m**ʁg**, χ**lg**, t**mg**, ʁ**mg**, s**mg**

The stimuli were conjugated into existing Hebrew verb templates. OCP violations were tested in four phonological environments, differentiated by the position of the tested pairs (C_1-C_2 vs. C_2-C_3) and the distance between the consonants (adjacent or separated by a vowel). The equivalent WF groups were created using the same templates. SCL violations, applicable only to adjacent consonants, were tested in two phonological environments. The verb templates used in the experiments are shown in (20).

(20) Nonce verb templates by phonological environments

C_1C_2 (OCP, SCL, WF)	hi**CC**iCa hif'il, PAST.3FSG	e.g., *hi**kg**ina*
C_2C_3 (OCP, SCL, WF)	Ca**CC**a qal, PAST.3FSG	e.g., *da**gf**a*
C_1VC_2 (OCP, WF)	**C**a**C**aC qal, PAST.3MSG	e.g., ***l**a**g**aʁ*
C_2VC_3 (OCP, WF)	Ci**C**e**C** pi'el, PAST.3MSG	e.g., *bi**t**e**d***

All templates were in the past tense, third-person singular: two masculine forms without a suffix and two feminine forms with the suffix *-a*. To avoid potential violations across morphemes, all consonants in the nonce verbs functioned as stem consonants. Note that

[15] These are not Hebrew roots in the traditional sense (Goshen-Gottstein 1964, Schwarzwald 2002, among others), as they represent the phonetic surface form of the consonants rather than an abstract morphological root. This interpretation aligns with Bat-El's (1994) argument for derivation directly from the base through Stem Modification (Steriade 1988), a process that eliminates the need for a consonantal root in Hebrew grammar. The broader debate regarding the concept of consonantal roots in Hebrew lies beyond the scope of this study.

the hiCCiCa template begins with the consonant *h* (glottal fricative), which is typically omitted in Modern Hebrew and is assumed to lack a place of articulation (see §2.3). Consequently, it could not cause an OCP violation.

All in all, the experiment consisted of 50 nonce verbs:

- **OCP (20):** 4 phonological environments × 5 items in each environment
- **SCL (10):** 2 phonological environments × 5 items in each environment
- **WF (20):** 4 phonological environments × 5 items in each environment

The OCP and WF groups in C_1C_2 and C_2C_3 shared stem consonants as their counterparts in C_1VC_2 and C_2VC_3. The complete list of nonce verbs is presented in (21).

(21) Nonce verbs list (tested consonant pairs in bold)

C_1C_2	hi**CC**iCa hif'il, PAST.3FSG	OCP	hi**td**iχa, hi**td**ika, hi**kg**ida, hi**kg**ina, hi**kg**ima
		SCL	hi**fn**iχa, hi**vn**ika, hi**vn**ina, hi**vn**iχa, hi**χn**iva
		WF	hi**ʁb**ila, hi**ʁb**iʃa, hi**lg**iʁa, hi**ʁs**ifa, hi**ʁz**iza
C_2C_3	Ca**CC**a qal, PAST.3FSG	OCP	ba**td**a, ma**td**a, ma**kg**a, ʁa**kg**a, ta**kg**a
		SCL	da**gf**a, da**gv**a, ga**dv**a, da**gχ**a, pa**χn**a
		WF	ma**ʁg**a, χa**lg**a, ta**mg**a, ʁa**mg**a, sa**mg**a
C_1VC_2	**C**a**C**aC qal, PAST.3MSG	OCP	**t**a**d**aχ, **t**a**d**ak, **k**a**g**ad, **k**a**g**an, **k**a**g**am
		WF	**ʁ**a**v**al, **ʁ**a**v**aʃ, **l**a**g**aʁ, **ʁ**a**s**af, **ʁ**a**z**az
C_2VC_3	Ci**C**e**C** pi'el, PAST.3MSG	OCP	bi**t**e**d**, mi**t**e**d**, mi**k**e**g**, ʁi**k**e**g**, ti**k**e**g**
		WF	mi**ʁ**e**g**, χi**l**e**g**, ti**m**e**g**, ʁi**m**e**g**, si**m**e**g**

The nonce verbs were embedded in carrier sentences, each containing a single nonce verb, creating 50 stimuli. The sentences followed the structure: **proper name + nonce verb + *et ze* 'ACC this'**, e.g., *zohaʁ makga et ze*. This format provided a natural context for the nonce verbs, while the consistent syntactic structure reduced potential semantic interference. All templates could accommodate a direct object marked with *et*. This was confirmed through an analysis of the 30 most frequent verbs in each template in the corpus (*qal* CaCaC, *pi'el* CiCeC, and *hif'il* hiCCiC), revealing that at least 50% of the verbs in each template could take a direct object with *et*. Proper names were selected to ensure that their final consonant did not share place of articulation with the initial consonant of the nonce verb.

The stimuli were recorded by a 33-year-old female native speaker of Hebrew and presented auditorily to avoid orthographic influence. The full list of stimuli appears in Appendix D, and frequency details appear in Appendix E.

4.2.1.3. Procedure

The experiment was conducted online (due to COVID-19) using the Alchemer platform. Participants were instructed to complete the experiment in a quiet room without distractions. Before starting, they were asked to provide background information: age, gender, native language(s), academic background in Linguistics or Hebrew language, and any hearing problems. Bilingual or non-native speakers and participants with hearing problems were disqualified before the experiment began and received a disqualification message immediately after filling out the background information form.

Participants were presented with 50 sentences in random order. They were asked to rate the acceptability of the nonce verb in each sentence as a potential Hebrew verb on a five-point scale, ranging from 1 (very unlikely) to 5 (very likely). They were instructed to respond quickly and intuitively without considering the meaning of the nonce verb. A pause screen (featuring a cute hedgehog image) appeared every 17 sentences. At the end of the experiment, participants could provide their email address to enter a raffle for the breakfast voucher. The experiment took approximately 10 minutes to complete.

4.2.2. Results

Ratings of the nonce verbs were collected and analyzed. The distance between each item's mean rating and its group's mean did not exceed 0.7 SD, indicating no outlier items. Consequently, all items were included in the analyses. The mean ratings are presented in (22).

(22) Mean ratings (on a 1–5 scale)

	OCP		SCL		WF	
	M	SD	M	SD	M	SD
C_1C_2	2.57	0.91	3.32	0.68	3.26	0.61
C_2C_3	2.27	0.83	3.24	0.62	3.24	0.75
C_1VC_2	2.40	0.70	NA		3.38	0.81
C_2VC_3	3.09	0.75	NA		3.47	0.70

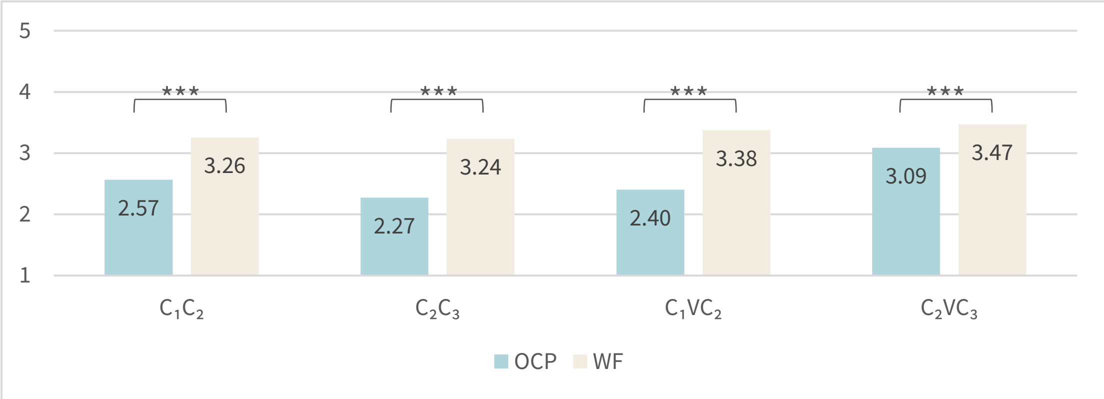
As shown in the table above, the mean ratings for the SCL and WF groups were consistently above 3 (the midpoint of the scale), while the mean ratings for the OCP groups were below 3 (in C_1C_2, C_2C_3, and C_1VC_2) or very close to 3 (in C_2VC_3, 3.09). This indicates that items violating the SCL and well-formed (WF) items were generally judged as possible Hebrew verbs ($M > 3$), whereas items violating the OCP were not ($M < 3$).

The following subsections provide statistical analyses for the OCP and SCL constraints across different phonological environments. As noted in §4.1, this study avoids comparing violations across environments (e.g., OCP violations in C_1C_2 to C_2C_3) due to the potential influence of verb templates on the results, independent of constraint violations.

4.2.2.1. OCP

Planned comparisons using t-tests revealed that the OCP groups were rated significantly lower than the WF groups across all phonological environments (C_1C_2: $t(59) = -4.88$, $p < 0.001$, Cohen's $d = -0.63$; C_2C_3: $t(59) = -7.75$, $p < 0.001$, Cohen's $d = -1$; C_1VC_2: $t(59) = -8.76$, $p < 0.001$, Cohen's $d = -1.13$; C_2VC_3: $t(59) = -4.24$, $p < 0.001$, Cohen's $d = -0.55$). These results are shown in Graph (23).

(23) OCP vs. WF across phonological environments (on a 1–5 scale)



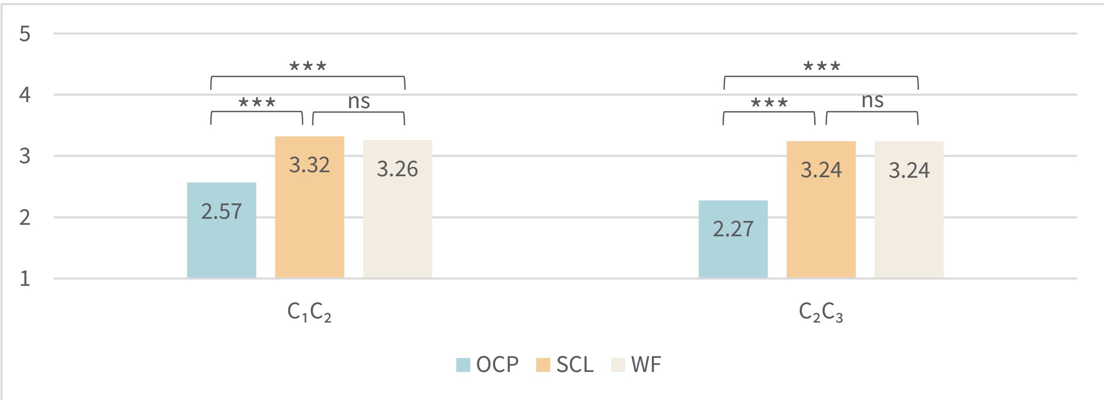
*** = p < 0.001

The graph above shows that participants judged OCP-violating nonce verbs as significantly less likely to be valid Hebrew versbs compared to WF nonce verbs.

4.2.2.2. SCL

Planned comparisons using t-tests showed that the differences between the SCL and WF groups were not statistically significant in either phonological environment (C_1C_2: $t(59) = 0.83$, $p = 0.41$; C_2C_3: $t(59) = 0.07$, $p = 0.94$). However, items in the SCL groups were rated significantly higher than those in the OCP groups (C_1C_2: $t(59) = 5.36$, $p < 0.001$, Cohen's $d = 0.69$; C_2C_3: $t(59) = 10.88$, $p < 0.001$, Cohen's $d = 1.4$). These results are shown in Graph (24).

(24) SCL vs. OCP and WF across phonological environments (on a 1–5 scale)



*** = p < 0.001; ns = p ≥ 0.05

The graph above shows that participants rated SCL-violating nonce verbs similarly to WF nonce verbs and significantly higher than OCP-violating nonce verbs.

4.2.3. Discussion

In Experiment A, native Hebrew speakers evaluated the acceptability of three groups of nonce verbs: OCP-violating, SCL-violating, and WF. All tested consonant pairs had low frequencies in the Hebrew lexicon.

The results from the WF groups highlight that frequency alone cannot account for phonotactic restrictions. Despite the low frequencies of the tested consonant pairs, WF nonce verbs consistently received ratings above the midpoint of the scale ($M > 3$). Had participants rejected nonce verbs purely based on frequency, the WF groups would likely have received lower ratings, falling below the midpoint.

When examining the phonotactic constraints, the results suggest that the OCP is active in the speakers' phonological system. In contrast, the results for the SCL provide a more ambiguous picture. The following sections provide a detailed analysis of the OCP and SCL constraints across different phonological environments.

4.2.3.1. OCP

OCP is an active constraint in the Hebrew lexicon, restricting consonant pairs that violate it (§3.4.1). Comparisons between the OCP and WF groups showed significantly lower mean ratings for OCP groups across all phonological environments. In general, OCP-violating nonce verbs were consistently rated as less acceptable items.

These findings cannot be attributed to the frequency of the consonant pairs, as no statistical differences in frequency were found between the OCP and WF groups. The observed differences are thus attributed to OCP violations.

However, these results do not clarify whether speakers' sensitivity to the OCP originates from their lexicon or represents independent phonological knowledge. The alignment of the OCP results with predictions from both Universal Grammar (UG) and the lexicon prevents any clear distinction between the two influences.

4.2.3.2. SCL

Unlike the OCP, the SCL does not appear to be active in the Hebrew lexicon, as SCL-violating consonant pairs are not consistently restricted (§3.4.2). Comparisons between SCL and WF groups revealed no statistically significant differences in ratings. Moreover, the SCL groups were rated significantly higher than the OCP groups across both tested phonological environments.

At first sight, these findings indicate that speakers satisfy phonotactic constraints active in their lexicon (OCP) while tolerating violations of constraints that are frequently violated in it (SCL). This pattern suggests that the lexicon alone may account for the results, with no definitive evidence supporting the influence of UG.

A closer examination of the sonority distance between consonant pairs reveals a more nuanced pattern. In the C_1C_2 environment, all SCL pairs consisted of an obstruent and a sonorant (*fn*, *vn*, *χn*) and showed no significant difference from the WF group. In the C_2C_3 environment, most SCL pairs included two obstruents (*gf*, *gv*, *dv*, *gχ*), except for one item with an obstruent-sonorant pair (*χn*; *pa**χn**a*). The mean rating for *paχna* was unusually high (M = 4.02) compared to the group mean (M = 3.24). However, as it was within one standard deviation, it was included in the analysis. Excluding *paχna*, the group's mean dropped to 3.05 (SD = 0.69), resulting in a statistically significant difference between the SCL and WF groups in a one-tailed test ($t(59) = -1.88$, $p = 0.03$, Cohen's $d = 0.26$).

These results suggest a possible distinction between obstruent-obstruent and obstruent-sonorant pairs in relation to SCL violations. This distinction will be further explored in Experiment B and examined in greater detail in Experiment C.

4.3. Experiment B: Lexical Decision Task

Experiment B replicated Experiment A using a different method: a lexical decision task instead of lexical judgments. Participants were presented with a set of nonce verbs based on those from Experiment A, including violations of the OCP or SCL, as well as well-formed (WF) nonce verbs without phonotactic violations. Hebrew real verbs were included as filler items. Both accuracy and reaction times (RTs) were collected and analyzed.

The methodologies of Experiment A and Experiment B complement each other: the lexical judgment task provided a direct, metalinguistic reflection of speakers' preferences, whereas the lexical decision task tapped into real-time processing. Together, they provide converging perspectives on the role of universal constraints in speakers' phonological systems.

4.3.1. Method

4.3.1.1. Participants

The participants included 70 native Hebrew speakers (49 women and 21 men, aged 20–45, M = 31.2, SD = 6.2). None of the participants had any native language besides Hebrew, nor did they report ADHD or hearing problems. Additionally, none of them had studied Linguistics or Hebrew Language at a university level.[16] Participants received a voucher for coffee and a pastry as compensation.

The results of 16 additional participants were excluded from the analysis: six participants exhibited a total accuracy below 90%, five had an accuracy below 90% on critical items, and five had mean reaction times exceeding 2.5 SD above the overall participant mean.

[16] Participant overlap between the two experiments was not checked, yet any such overlap was unlikely to create repetition effects given the several-month gap and the different methodologies.

4.3.1.2. Stimuli

The stimuli for Experiment B were based on those used in Experiment A (see §4.2.1.2). While Experiment A included five items per group, Experiment B included 10, with each consonant triplet generated in two templates rather than one, in order to provide a sufficient number of items for reliable reaction-time analyses. The verb templates used in Experiment B are listed in (25).

(25) Nonce verb templates by phonological environments

C_1C_2 (OCP, SCL, WF)	hi**CC**iC hif'il, PAST.3MSG	e.g., *hi**kg**in*
	hi**CC**iCa hif'il, PAST.3FSG	e.g., *hi**kg**ina*
C_2C_3 (OCP, SCL, WF)	Ca**CC**a qal, PAST.3FSG	e.g., *da**gf**a*
	Ci**CC**a pi'el, PAST.3FSG	e.g., *di**gf**a*
C_1VC_2 (OCP, WF)	**C**a**C**aC qal, PAST.3MSG	e.g., ***lag**aʁ*
	Ci**C**eC pi'el, PAST.3MSG	e.g., ***lig**eʁ*
C_2VC_3 (OCP, WF)	Ca**C**a**C** qal, PAST.3MSG	e.g., *ba**tad***
	Ci**C**e**C** pi'el, PAST.3MSG	e.g., *bi**ted***

As in Experiment A, all templates were in the past tense, third-person singular (five masculine and three feminine). All consonants in the nonce verbs were stem consonants to avoid potential morpheme violations, except for the initial *h* in the hiCCiC and hiCCiCa templates (see §4.2.1.2).

Thus, the experiment included 100 nonce verbs as critical items:

- **OCP (40):** 4 phonological environments × 10 items in each environment
- **SCL (20):** 2 phonological environments × 10 items in each environment
- **WF (40):** 4 phonological environments × 10 items in each environment

The OCP and WF groups in C_1C_2 and C_2C_3 shared stem consonants with their CVC counterparts in C_1VC_2 and C_2VC_3. These 100 nonce verbs were combined with 100 familiar Hebrew verbs, which served as fillers, maintaining a 1:1 ratio between critical items and fillers. The fillers were conjugated in the same templates as the critical items, all from the *shlemim* group (see §3.2.2.2), and had mid-range frequencies of 20–60 occurrences per million words (based on the *heTenTen* corpus). Fillers and targets showed minimal overlap in consonant pairs (four in C_1VC_2, one in C_2VC_3, and none in C_1C_2 or C_2C_3). All in all, the stimuli consisted of 200 items.

The stimuli were recorded by a 34-year-old female native Hebrew speaker and presented auditorily to avoid orthographic influence. The recordings were controlled for length, and no significant differences were found between groups in the C_1C_2, C_2C_3, and C_2VC_3 environments (C_1C_2: $F(2,27) = 2.43$, $p = 0.11$; C_2C_3: $F(2,27) = 3.06$, $p = 0.06$; C_2VC_3: $t(18) = 0.92$, $p = 0.37$). However, in the C_1VC_2 environment, the OCP group's items were significantly shorter than those in the WF group ($t(18) = 4.12$, $p < 0.001$), a difference that will be considered in interpreting the results. No significant length differences were observed between critical items and fillers ($t(193) = 1.11$, $p = 0.27$). A full list of stimuli appears in Appendix F, with frequency details in Appendix G.

4.3.1.3. Procedure

The experiment was designed using the *PsychoPy* software (Peirce et al., 2019) and conducted online (due to COVID-19). Participants were instructed to complete the experiment in a quiet room, free from distractions.

At the beginning of the experiment, participants completed a *Google* form providing the following details: age, gender, native language(s), academic background in Linguistics or Hebrew Language, dominant hand, and whether they had ADHD or hearing problems. They were also asked to provide their email address (for a voucher for coffee and a pastry) and the last three digits of their phone number, which served as their participant number. Bilingual or non-native Hebrew speakers, as well as participants reporting hearing problems or ADHD, were disqualified at this stage. Disqualified participants received a polite disqualification message and did not proceed to the experiment.

Upon completing the form, participants were provided with a link to the experiment hosted on the *Pavlovia* platform. Participant numbers were used to link form responses with experiment results, and time stamps were employed to resolve any instances of duplicate participant numbers.

Participants were instructed to listen carefully to each trial and determine whether the stimulus was a real Hebrew verb as quickly and accurately as possible. They pressed the F key for real Hebrew verbs and the J key for nonce verbs. Each trial began with a fixation point displayed at the center of the screen, followed by the auditory presentation of the stimulus. An intertrial interval of 500 ms separated responses from the next trial. Stimuli were presented in random order, with a pause screen appearing every 50 trials.

The experiment began with a short practice session consisting of 10 trials (five real verbs and five nonce verbs, all using templates from the main experiment) with feedback provided after each trial. Following the practice session, participants proceeded to the main experiment, which was conducted without feedback. Accuracy and reaction times (RTs) from the onset of the stimuli were recorded. The entire experiment took approximately 10 minutes to complete.

4.3.2. Results

Accuracy and reaction times (RTs) for critical items (i.e., nonce verbs) were analyzed. All participants, including eight left-handed individuals, were included in the analysis, as their mean RTs did not differ significantly from those of right-handed participants ($p = 0.33$). Two items with low accuracy rates were excluded from the analysis: *miʁga* (C_2C_3 WF, 74.3%) and *χileg* (C_2VC_3 WF, 71.4%). All remaining items had accuracy rates exceeding 85%.

RT analysis was conducted on correct responses only. Responses exceeding 5 seconds were excluded, and RTs exceeding 2.5 SD from the participant's mean (for correct responses to critical items) were truncated to the 2.5 SD value. This adjustment affected less than 2% of the data. Afterward, all RTs fell within the expected range, with the maximum deviation between an item's mean and its group's mean being 0.9 SD. Accuracy rates and mean RTs are presented in (26).

(26) Accuracy rates (%) and mean RTs (ms)

			RTs	
		Accuracy	M	SD
C_1C_2	OCP	96.7%	1117	110
	SCL	95.6%	1171	119
	WF	97.4%	1155	101
C_2C_3	OCP	98.1%	1029	111
	SCL	96%	1003	112
	WF	99%	1027	100
C_1VC_2	OCP	97.3%	918	109
	WF	98%	1110	112
C_2VC_3	OCP	94.9%	1112	114
	WF	98.3%	1070	104

The C_2C_3, C_1VC_2, and C_2VC_3 groups included items from two templates: *qal* (CaCCa, CaCaC) and *pi'el* (CiCCa, CiCeC) (see §4.3.1.2). Template-based differences were observed across environments. In the C_1VC_2 OCP and C_2VC_3 OCP groups, RTs were significantly shorter for the CaCaC template than for the CiCeC template ($p < 0.05$). Conversely, in the C_1VC_2 WF and C_2VC_3 WF groups, RTs were significantly shorter for the CiCeC template than for the CaCaC template ($p < 0.05$). No significant template differences were found in the C_2C_3 OCP, C_2C_3 SCL, or C_2C_3 WF groups.

These findings raise questions about the interaction between verb templates and phonotactic constraints. For example, how do verb templates interact with phonotactic constraints across different environments? Are these effects driven by phonological or by lexical-semantic factors? Addressing these questions in future research could provide valuable insights into the role of morphological structures in phonotactic processing.

The following subsections provide statistical analyses for the OCP and SCL constraints across different phonological environments. As noted in §4.1, this study avoids comparing violations across environments (e.g., OCP violations in C_1C_2 to C_2C_3) due to the potential influence of verb templates on the results, independent of constraint violations.

4.3.2.1. OCP

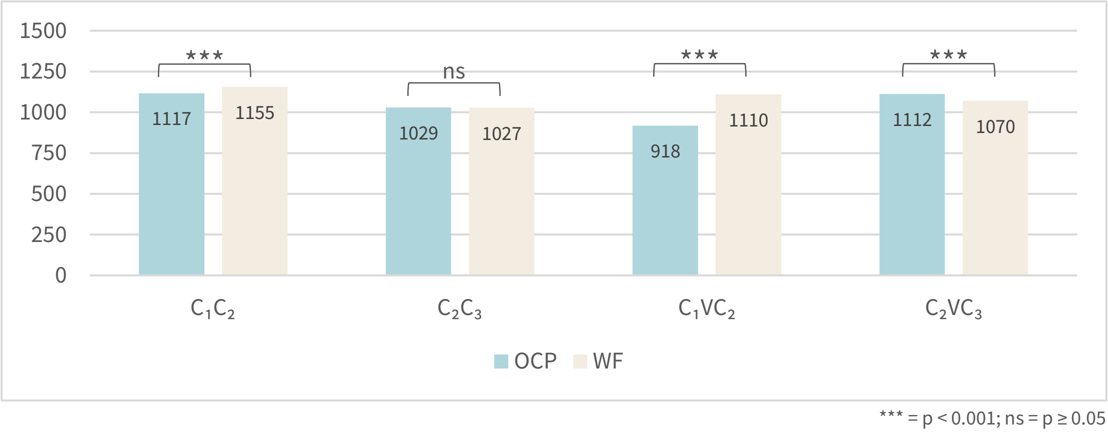
Planned comparisons using t-tests revealed that participants detected OCP-violating items significantly faster than WF items in C_1C_2 and C_1VC_2 (C_1C_2: $t(69) = -3.46$, $p < 0.001$, Cohen's $d = -0.41$; C_1VC_2: $t(69) = -18.24$, $p < 0.001$, Cohen's $d = -2.18$). In C_2C_3, no significant difference in RTs was observed ($t(69) = 0.15$, $p = 0.88$), whereas in C_2VC_3, participants detected WF items significantly faster than OCP-violating items ($t(69) = 5.48$, $p < 0.001$, Cohen's $d = 0.66$).[17] Furthermore, in C_2VC_3, participants demonstrated lower accuracy for OCP-violating items compared to WF items ($t(69) = -3.1$, $p = 0.003$, Cohen's $d = -0.37$), with no significant accuracy differences in the other environments.[18]

Comparisons of RTs are presented in Graph (27), and comparisons of accuracy are presented in Graph (28).

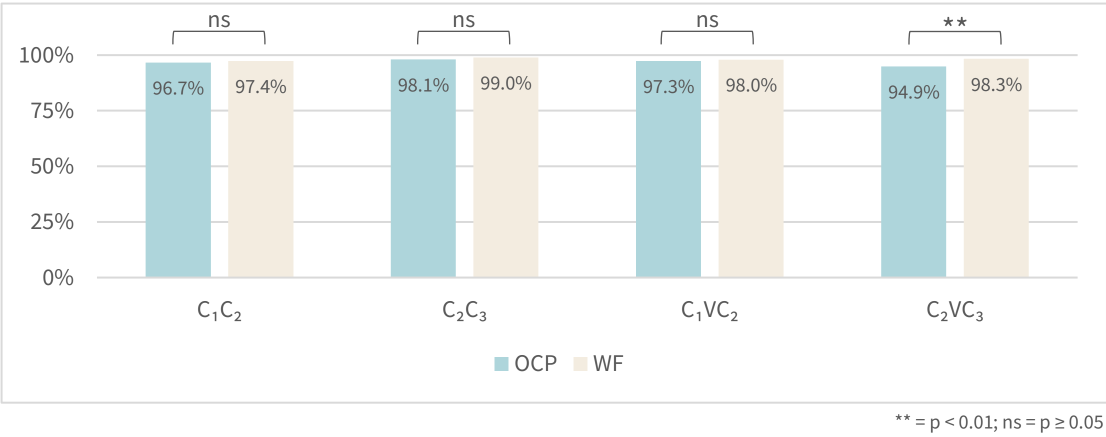
[17] The C_1VC_2 and C_2VC_3 groups showed RT differences between templates, prompting comparisons of OCP and WF within each phonological environment by template. In C_1VC_2, both template comparisons (OCP-CaCaC vs. WF-CaCaC and OCP-CiCeC vs. WF-CiCeC) were significant ($p < 0.001$). In C_2VC_3, the comparison for CiCeC was significant ($p < 0.001$), but the comparison for CaCaC was not ($p = 0.38$).

[18] The significant accuracy difference in C_2VC_3 persisted when analyzed by template (CaCaC: OCP = 96.6%, WF = 99.1%, $p = 0.005$; CiCeC: OCP = 93.1%, WF = 97.1%, $p = 0.02$).

(27) RTs of OCP vs. WF across phonological environments (in milliseconds)



(28) Accuracy of OCP vs. WF across phonological environments (% of correct responses)

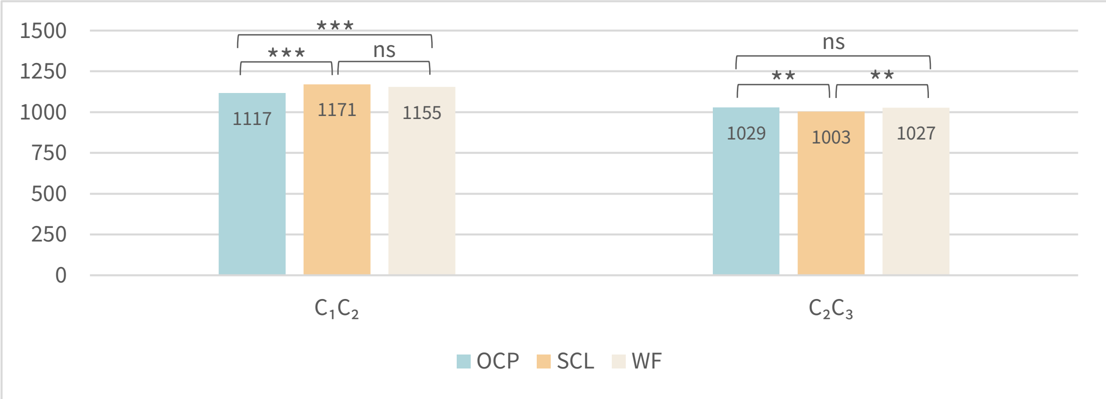


The graphs above show that OCP-violating items were detected faster in C_1C_2 and C_1VC_2. In contrast, WF items were detected faster in C_2VC_3, where OCP violations also resulted in more errors. No significant differences were observed between OCP and WF groups in C_2C_3.

4.3.2.2. SCL

Planned comparisons using t-tests revealed differences across environments. In C_1C_2, participants detected SCL-violating items significantly slower than OCP-violating items ($t(69) = 6.41$, $p < 0.001$, Cohen's $d = 0.77$), with no significant difference between SCL-violating and WF items ($t(69) = 1.4$, $p = 0.17$). Conversely, in C_2C_3, participants detected SCL-violating items significantly faster than both OCP-violating items ($t(69) = -2.72$, $p = 0.008$, Cohen's $d = -0.33$) and WF items ($t(69) = -2.87$, $p = 0.005$, Cohen's d = -0.34). Notably, the difference between OCP-violating and WF items was significant in C_1C_2 but not in C_2C_3 (see §4.3.2.1). Comparisons of RTs are shown in Graph (29).

(29) RTs of SCL vs. OCP and WF across phonological environments (in milliseconds)

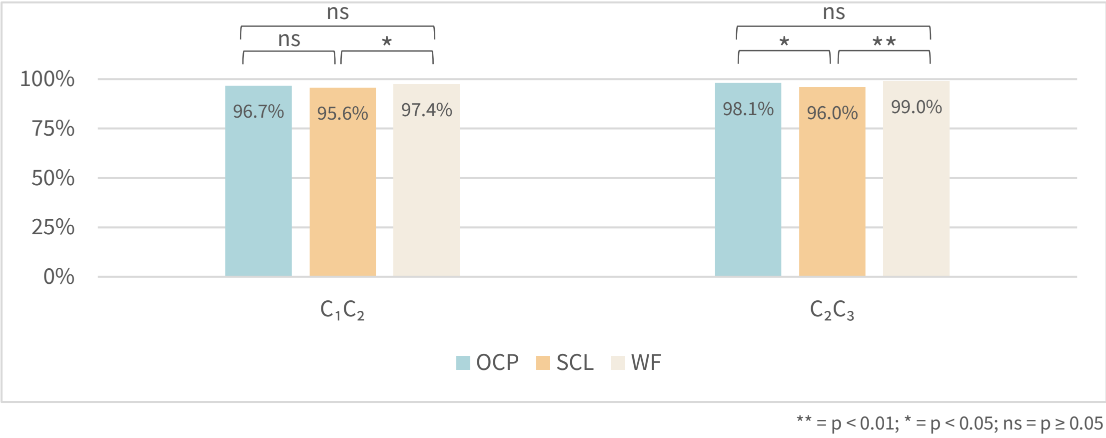


*** = p < 0.001; ** = p < 0.01; ns = p ≥ 0.05

As shown in the graph above, participants detected SCL-violating items more slowly than OCP-violating items in C_1C_2, although the difference from WF items was not significant. In contrast, in C_2C_3, participants detected SCL-violating items faster than both OCP-violating and WF items.

The SCL groups also showed lower accuracy rates compared to the OCP and WF groups. In C_1C_2, accuracy for SCL-violating items was significantly lower than for WF items ($t(69) = -2.26$, $p = 0.03$, Cohen's $d = -0.27$), with no significant difference from OCP items ($t(69) = -1.3$, $p = 0.2$). In C_2C_3, SCL-violating items had lower accuracy rates than both OCP ($t(69) = -2.3$, $p = 0.02$, Cohen's $d = -0.28$) and WF items ($t(69) = -3.37$, $p = 0.001$, Cohen's $d = -0.4$). As noted, the difference between OCP and WF groups was not significant in either environment (see §4.3.2.1). Comparisons of accuracy are shown in Graph (30).

(30) Accuracy of SCL vs. OCP and WF across phonological environments (% of correct responses)



As shown in the graph above, participants made more errors detecting SCL-violating items across all comparisons, except when compared to OCP-violating items in C_1C_2.

4.3.3. Discussion

Like Experiment A, Experiment B examined the role of OCP and SCL in speakers' phonological system. While Experiment A relied on meta-linguistic judgments, Experiment B used a psycho-phonological lexical decision task. Participants were presented with a set of auditory stimuli, including nonce verbs from OCP, SCL, and WF groups (based on Experiment A stimuli), along with real Hebrew verbs as filler items. Accuracy and reaction times were collected and analyzed. The results largely replicated those of Experiment A, suggesting that both frequency and phonological constraints (OCP in this case) influence the phonological system.

The experiment confirmed that co-occurrence restrictions are not solely frequency based. Despite all tested consonant pairs having low lexical frequency, three of the four WF groups showed significantly different RTs compared to the OCP- and SCL-violating groups. Such differences would not arise if participants rejected nonce verbs based only on consonant pair frequency, highlighting the additional influence of phonotactic constraints. The following sections examine the roles of OCP and SCL across different phonological environments.

4.3.3.1. OCP

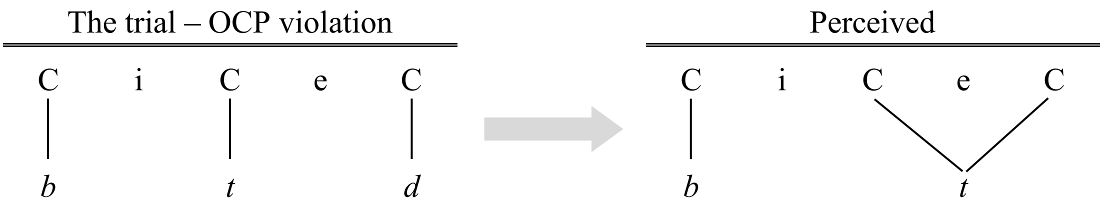
In the C_1C_2 and C_1VC_2 environments, RTs for OCP groups were significantly shorter than for the WF groups, reflecting sensitivity to OCP violations. As in Experiment A, this sensitivity could originate from the lexicon (see §3.4.1) or independent phonological knowledge; the results do not distinguish between these influences. In the C_2C_3 environment, OCP-violating items were recognized faster than WF items, though the difference was not statistically significant.

In the C_1VC_2 environment, the OCP group's items were shorter than those in the WF group, potentially increasing the effect size in this environment. However, the effect was also found in C_1C_2 and across all environments in Experiment A, indicating that item length alone does not account for the effect in C_1VC_2.

In contrast, the C_2VC_3 environment showed the opposite tendency: RTs for the OCP group were significantly longer than those for the WF group, and participants made more errors with OCP-violating items, classifying more of them as real Hebrew words. This indicates that participants hesitated longer and made more mistakes when deciding whether OCP-violating items in this environment were nonce verbs.

This finding can be attributed to phonological perception. Participants may have "corrected" the nonce verbs to align with the phonological constraints of Hebrew. In the OCP group, all items ended with a voiceless-voiced consonant pair (e.g., *t*V*d*, *k*V*g*), which is not permitted in Hebrew. However, Hebrew permits identical consonants in this position (e.g., *t*V*t*, *k*V*k*), where the consonant maps to a single C position instead of two, meaning no OCP violation occurs (see Figure 31 and §2.2). This likely led participants to perceive the consonant pair as identical, creating uncertainty about whether they heard two voiceless consonants. Consequently, this uncertainty may have caused hesitation, longer response times, and more errors.

(31) Phonological perceptual correction



This perceptual correction was not observed in Experiment A, where nonce verbs were presented within full sentences (e.g., *omeʁ mited et ze*). In that context, the final consonant of the nonce verb was followed by the accusative marker *et*, which begins with a vowel, creating a clear transition. This transition facilitated the perception of the final consonant's voicing compared to its realization in a word-final position.

4.3.3.2. SCL

Unlike the OCP, the SCL does not appear to be active in the Hebrew lexicon, as SCL-violating consonant pairs are not consistently restricted (§3.4.2). Consequently, if speakers' phonological systems were shaped solely by the lexicon, SCL violations would not be expected to influence RTs. However, if the SCL, as a universal constraint, is integrated into speakers' phonological systems, nonce verbs with SCL violations should elicit shorter RTs, indicating their lower likelihood of being perceived as real words.

The results revealed differences across phonological environments. In the C_1C_2 environment, RTs for the SCL group did not differ significantly from those for the WF group but were longer than those for the OCP group. Conversely, in the C_2C_3 environment, RTs for the SCL group were significantly shorter than those for both the WF and OCP groups.

However, based on the findings of Experiment A, this difference is likely attributable to the **sonority** of the tested pairs rather than their position. In the C_1C_2 environment, all tested pairs consisted of obstruent-sonorant sequences (e.g., *fn*, *vn*, *χn*), whereas in the C_2C_3 environment, 80% of the nonce verbs featured two obstruent-obstruent pairs (e.g., *gf*, *gv*, *dv*, *gχ*). The results suggest that participants were particularly sensitive to SCL violations involving two obstruents.

Violations of the SCL in obstruent-obstruent pairs (the C_2C_3 group) marked these items as non-Hebrew, leading participants to classify them more quickly as nonce verbs. Additionally, the SCL group exhibited lower accuracy rates compared to the equivalent WF group, suggesting that SCL violations increased the likelihood of errors.

As discussed in §3.4.2, the Hebrew lexicon generally permits SCL violations. However, the findings from Experiments A and B highlight the complexity of the SCL's role in the phonological system. While the lexicon tolerates SCL violations, participants demonstrated distinct responses to obstruent-sonorant pairs versus obstruent-obstruent pairs, a distinction further explored in Experiment C.

4.4. Experiment C: Sonority Distance

Experiment C was designed as a follow-up to Experiments A and B, where preliminary findings suggested that sonority distance (SonD) could influence speakers' tolerance to SCL violation. Specifically, Experiment C investigated how the interaction between sonority distance and its profile (rising or falling) affects acceptability, using a 2×2 design to compare these factors.

When the two consonants occur in separate syllables (…C.C…), a rise in sonority violates the SCL, while a fall satisfies it. The sonority distance (±1 or ±2) is relevant to the Sonority Dispersion Principle (SDP), which favors larger sonority distances over smaller ones.

Two sonority distance intervals were examined: a small distance between two obstruents (SonD = ±1) and a larger distance between an obstruent and a sonorant (SonD = ±2). Each interval was designated as positive (rising) or negative (falling). The procedure mirrored that of Experiment A, with participants providing lexical judgments for nonce verbs in a rating task.

4.4.1. Method

4.4.1.1. Participants

The participants included 60 native Hebrew speakers (43 women and 17 men, aged 21–44, M = 30.3, SD = 6.5). None of the participants had any native language besides Hebrew, nor did they report any hearing problems. Additionally, none of them had studied Linguistics or Hebrew Language at a university level. Participants were entered into a prize draw for a breakfast voucher as compensation.

4.4.1.2. Stimuli

The stimuli for Experiment C included nonce verbs with consonant pairs specifically designed to assess acceptability across different sonority distances (SonD). SonD represents both the absolute distance (e.g., 1 or 2) and the direction, indicating whether sonority rises (+) or falls (−) between the consonants. When mapped onto separate syllables, a rise in sonority violates the SCL, whereas a fall satisfies it.

The tested consonant pairs were organized into four groups:

- **Stp-Fri (stop-fricative):** Obstruent pairs (stop and fricative) with SonD = +1, indicating a distance of 1 and a sonority rise (SCL violation).
- **Fri-Stp (fricative-stop):** Obstruent pairs (fricative and stop) with SonD = -1, indicating a distance of 1 and a sonority fall (SCL satisfaction).
- **Stp-Nas (stop-nasal):** Obstruent-sonorant pairs (stop and nasal) with SonD = +2, indicating a distance of 2 and a sonority rise (SCL violation).
- **Nas-Stp (nasal-stop):** Sonorant-obstruent pairs (nasal and stop) with SonD = -2, indicating a distance of 2 and a sonority fall (SCL satisfaction).

Figure (32) illustrates the scale of sonority values across these four groups.

(32) The scale of sonority values

	Sonority fall (SCL satisfaction)		Sonority rise (SCL violation)	
SonD:	-2	-1	+1	+2
	←			→
	Nas-Stp nasal-stop e.g., *mg*	**Fri-Stp** fricative-stop e.g., *vg*	**Stp-Fri** stop-fricative e.g., *gv*	**Stp-Nas** stop-nasal e.g., *gm*

Additionally, the experiment included a fifth group testing OCP violations with the aim of replicating the OCP-related findings from Experiments A and B. Altogether, the experiment consisted of five groups, as outlined in (33):

(33) Experimental groups

	+ sonority rise SCL violation	– sonority fall SCL satisfaction
1 obs, obs	**Stp-Fri** stop-fricative e.g., *ʦa**gv**a*	**Fri-Stp** fricative-stop e.g., *ʦa**vg**a*
2 obs, son	**Stp-Nas** stop-nasal e.g., *sa**gm**a*	**Nas-Stp** nasal-stop e.g., *sa**mg**a*

OCP violation
OCP e.g., *ma**td**a*

As in the previous experiments, the tested pairs were frequency-controlled based on the corpus analysis. For both SonD values (±1 and ±2), pairs with a sonority rise (Stp-Fri, Stp-Nas) appeared significantly more frequently in the lexicon than their counterparts with a sonority fall (Fri-Stp, Nas-Stp) (±1: $t(7) = 7.94$, $p < 0.001$; ±2: $t(7) = 3.71$, $p = 0.008$). Thus, SCL-violating pairs were more frequent in the lexicon than the SCL-satisfying pairs. Therefore, if SCL-satisfying groups receive higher ratings than their SCL-violating counterparts, this preference cannot be explained by lexical frequency.[19]

[19] No significant frequency difference was observed between the two groups with a fall in sonority (Fri-Stp and Nas-Stp; $t(7) = 0.88$, $p = 0.4$). Among the groups with a rise in sonority, Stp-Nas pairs were significantly more frequent than Stp-Fri pairs ($t(7) = 6.88$, $p < 0.001$). The OCP pairs had a frequency of zero.

As in Experiments A and B, the stimuli were nonce verbs containing three stem consonants. The tested consonant pairs consistently appeared as C_2C_3 to allow focused comparison across groups. To ensure straightforward comparisons, the stop-fricative (Stp-Fri) and fricative-stop (Fri-Stp) groups used the same tested pairs in reversed order (e.g., *ʦa**gv**a-ʦa**vg**a*), as did the stop-nasal (Stp-Nas) and nasal-stop (Nas-Stp) groups (e.g., *sa**gm**a-sa**mg**a*). Additionally, C_3 was not *t*, since *t* in the CaCCa template (used for the stimuli) can serve as a feminine marker rather than a stem consonant (e.g., *ʁaʦta* 'want PAST.3FSG', *banta* 'build PAST.3FSG').

Each tested C_2C_3 pair was combined with an untested C_1 consonant, ensuring that no consonant triplets existed in Hebrew verbs. To maintain consistency, the same C_1 consonant was used with each matched pair (e.g., ***ʦ**agva-**ʦ**avga*, ***s**agma-**s**amga*), and C_1 did not share a place of articulation with either C_2 or C_3, avoiding OCP violations. In a few comparisons of sonority values, C_1 was not identical but a similar consonant (two stridents, as in ***s**amga-**ʦ**avga*).

In terms of C_1VC_2 frequencies, three out of the four planned comparisons (Stp-Nas vs. Nas-Stp; Stp-Fri vs. Fri-Stp; Nas-Stp vs. Fri-Stp) showed no significant differences ($p > 0.05$). The fourth comparison (Stp-Nas vs. Stp-Fri) revealed a significant difference, favoring Stp-Nas ($p = 0.03$), which was accounted for in the analysis. For the OCP group, The C_1 consonants were chosen as in Experiments A and B, ensuring that the C_1VC_2 combination had a medium-high frequency in the corpus (TPM $\geq$ 0.08).

All in all, 20 different nonce consonant triplets were generated: 5 conditions (Stp-Fri, Fri-Stp, Stp-Nas, Nas-Stp, and OCP) × 4 triplets in each condition. All consonants obey the Hebrew spirantization rules based on their position in the nonce verb. The stimuli were created by conjugating the nonce consonant triplets into two existing Hebrew verb templates: Ca**CC**a (*qal*, PAST.3FSG, e.g., *samga*) and Ca**CC**u (*qal*, PAST.3PL, e.g., *samgu*). This process generated the same verb in two forms (singular feminine and plural), resulting in a total of 40 nonce verbs. The complete list of nonce verbs is provided in (34).

(34) Nonce verbs list (tested consonant pairs in bold)

	+ sonority rise SCL violation		– sonority fall SCL satisfaction	
	Stp-Fri stop-fricative		**Fri-Stp** fricative-stop	
1 obs, obs	Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL	Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL
	ʦa**gv**a	ʦa**gv**u	ʦa**vg**a	ʦa**vg**u
	ja**gv**a	ja**gv**u	ja**vg**a	ja**vg**u
	ʦa**kf**a	ʦa**kf**u	ʦa**fk**a	ʦa**fk**u
	ja**kf**a	ja**kf**u	ja**fk**a	ja**fk**u
	Stp-Nas stop-nasal		**Nas-Stp** nasal-stop	
2 obs, son	Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL	Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL
	sa**gm**a	sa**gm**u	sa**mg**a	sa**mg**u
	ja**gm**a	ja**gm**u	ja**mg**a	ja**mg**u
	za**km**a	za**km**u	za**mk**a	za**mk**u
	ja**km**a	ja**km**u	ja**mk**a	ja**mk**u

OCP violation	
OCP	
Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL
ba**td**a	ba**td**u
ma**td**a	ma**td**u
ta**kg**a	ta**kg**u
ma**kg**a	ma**kg**u

As in experiment A, the nonce verbs were embedded in carrier sentences, each containing a single nonce verb, creating 40 stimuli. The sentences followed the structure: **proper name + nonce verb + *et ze* 'ACC this'**, e.g., *ʃiri sagma et ze.* This format provided a natural context for the nonce verbs, while the consistent syntactic structure reduced potential semantic interference. The two selected templates allowed for a direct object marked with *et*, and proper names were selected to ensure that their final consonant did not share place of articulation with the initial consonant of the nonce verb.

The stimuli were recorded by a 33-year-old female native speaker of Hebrew and presented auditorily to avoid orthographic influence. The full list of stimuli appears in Appendix H, and frequency details appear in Appendix I.

4.4.1.3. Procedure

The procedure mirrored that of Experiment A, as it successfully identified sensitivity to phonotactic violations (see §4.2.3).

The experiment was conducted online (due to COVID-19) using the Alchemer platform. Participants were instructed to complete the experiment in a quiet room without distractions. Before starting, they were asked to provide background information: age, gender, native language(s), academic background in Linguistics or Hebrew Language, and any hearing problems. Bilingual or non-native speakers and participants with hearing problems were disqualified before the experiment began and received a disqualification message immediately after filling out the background information form.

Participants were presented with 40 sentences in random order. They were asked to rate the acceptability of the nonce verb in each sentence as a potential Hebrew verb on a five-point scale, ranging from 1 (very unlikely) to 5 (very likely). They were instructed to respond quickly and intuitively without considering the meaning of the nonce verb. A pause screen (featuring a cute hedgehog image) appeared midway through the experiment after 20 sentences. At the end of the experiment, participants could provide their email address to enter a raffle for the breakfast voucher. The experiment took approximately 10 minutes to complete.

4.4.2. Results

Ratings for the nonce verbs were collected and analyzed. The deviation of each item's mean rating from its group's mean was within 0.4 SD, indicating the absence of outliers. Therefore, all items were retained for analysis. The mean ratings are presented in (35).[20]

(35) Mean ratings (on a 1–5 scale)

	+ sonority rise SCL violation	– sonority fall SCL satisfaction		OCP violation
1 obs, obs	**Stp-Fri** stop-fricative M = 3.09 SD = 0.72	**Fri-Stp** fricative-stop M = 3.23 SD = 0.67		**OCP** M = 2.24 SD = 0.78
2 obs, son	**Stp-Nas** stop-nasal M = 3.5 SD = 0.61	**Nas-Stp** nasal-stop M = 3.34 SD = 0.65		

As shown in the table above, the mean ratings for the Stp-Fri, Fri-Stp, Stp-Nas, and Nas-Stp groups consistently exceeded 3 (the midpoint of the scale). Conversely, in line with findings from Experiment A, the mean ratings for the OCP groups were below 3. This suggests that items in the Stp-Fri, Fri-Stp, Stp-Nas, and Nas-Stp groups were generally judged as possible Hebrew verbs (M > 3), while those violating the OCP were not (M < 3).

The following subsections provide a statistical analysis of the results. First, a comparison between the OCP group and all other groups is presented. Second, an analysis of the remaining four groups examines the effect of sonority distance, its direction (rising or falling), and the interaction between these two factors.

---

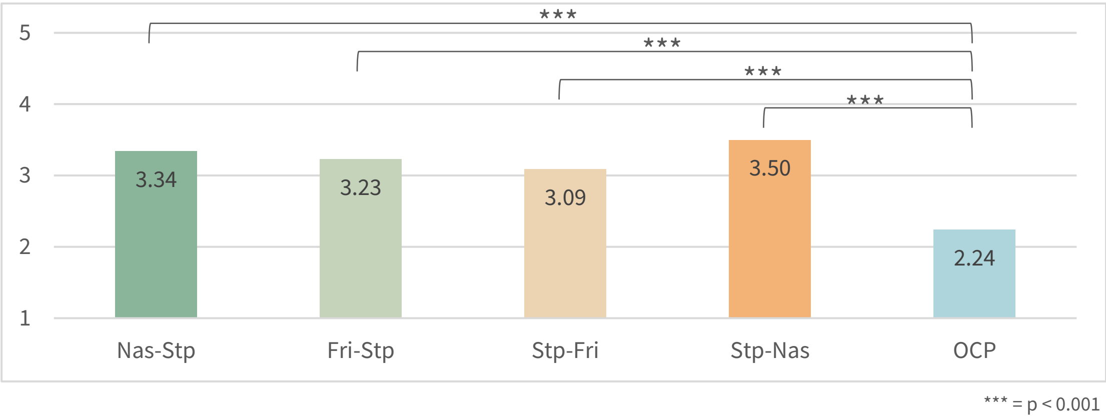
[20] In the Nas-Stp group, a significant rating difference emerged between singular feminine CaCCa items and their plural CaCCu counterparts ($p < 0.05$). This difference was not observed in other groups within the experiment. Comparisons within the Nas-Stp group, conducted separately for CaCCa and CaCCu forms, revealed rating patterns consistent with the main analysis. Further research is required to elucidate the specific impact of templates on rating patterns.

4.4.2.1. OCP violation

A one-way repeated measures ANOVA revealed a significant difference between the tested groups ($F(4,236) = 83.03$, $p < 0.001$, $\eta^2_p = 0.59$). Planned comparisons using t-tests showed that ratings for the OCP group were significantly lower than those for all other groups (OCP vs. Nas-Stp: $t(59) = 10.71$, $p < 0.001$, Cohen's $d = 1.38$; OCP vs. Stp-Nas: $t(59) = 11.77$, $p < 0.001$, Cohen's $d = 1.52$; OCP vs. Fri-Stp: $t(59) = 10.76$, $p < 0.001$, Cohen's $d = 1.39$; OCP vs. Stp-Fri: $t(59) = 10.71$, $p < 0.001$, Cohen's $d = 1.38$).

The results are presented in the Graph in (36).

(36) Comparison of OCP with other groups (on a 1–5 scale)



As shown in the graph, participants judged OCP-violating items as significantly less likely to be Hebrew verbs than items in all other groups.

4.4.2.2. Sonority distance and direction

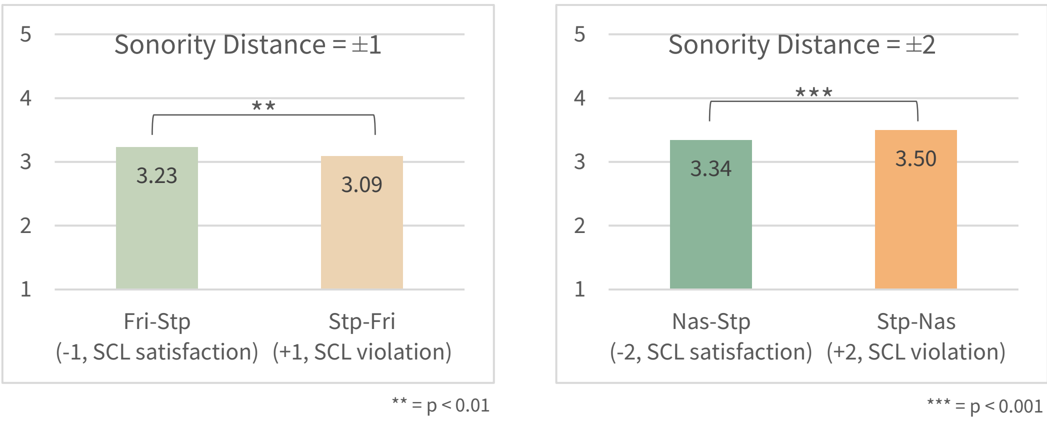
The two-way repeated measures ANOVA revealed a significant interaction between sonority distance and direction ($F(1,59) = 23.43$, $p < 0.001$, $\eta^2_p = 0.28$) , indicating that participants' ratings varied based on the combination of these factors. A significant main effect of distance was found ($F(1,59) = 21.15$, $p < 0.001$, $\eta^2_p = 0.26$), but no main effect of direction was observed ($F(1,59) = 0.05$, $p = 0.82$).

Planned t-tests examined the effect of sonority direction, which reflects participants' sensitivity to the SCL. Comparisons were made at two sonority distances: ±1, between

Stp-Fri (rising sonority, SCL violation) and Fri-Stp (falling sonority, SCL satisfaction), and ±2, between Stp-Nas (rising sonority, SCL violation) and Nas-Stp (falling sonority, SCL satisfaction).

The results showed that Fri-Stp items, which satisfy the SCL, were rated significantly higher as potential Hebrew verbs compared to Stp-Fri items, which violate the SCL ($t(59) = 3.141$, $p = 0.003$, Cohen's $d = 0.41$). In contrast, Nas-Stp items, which satisfy the SCL, were rated significantly lower than Stp-Nas items, which violate the SCL ($t(59) = 3.72$, $p < 0.001$, Cohen's $d = 0.48$). These results are presented in Graph (37).

(37) Effects of sonority direction on ratings (on a 1–5 scale)

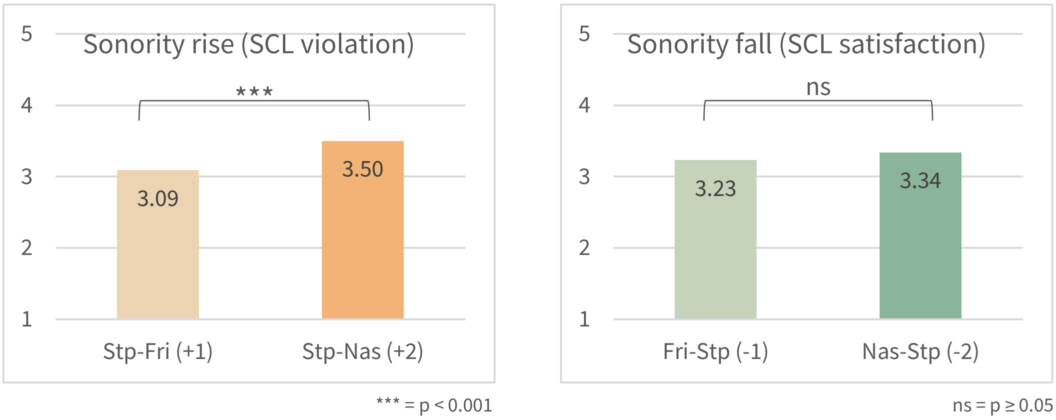


As shown in the graphs above, when the sonority distance was ±1, participants rated items with a sonority fall (Fri-Stp, SCL satisfaction) as significantly more likely to be Hebrew verbs than those with a sonority rise (Stp-Fri, SCL violation). In contrast, when the sonority distance was ±2, items with a sonority rise (Stp-Nas, SCL violation) were rated as more likely to be Hebrew verbs than those with a sonority fall (Nas-Stp, SCL satisfaction).

Next, planned comparisons were conducted to examine the effect of sonority distance, relevant to the SDP. For rising sonority, Stp-Fri (distance = +1) was compared with Stp-Nas (distance = +2), while for falling sonority, Fri-Stp (distance = -1) was compared with Nas-Stp (distance = -2).

In both comparisons, pairs with greater sonority distances received higher ratings. Stp-Nas items were rated significantly higher than Stp-Fri items as possible Hebrew verbs ($t(59) = 6.43$, $p < 0.001$, Cohen's $d = 0.83$), while Nas-Stp items received higher ratings than Fri-Stp items, though this difference was not statistically significant ($t(59) = 1.65$, $p = 0.1$). The results are shown in Graphs (38).

(38) Effects of sonority distance on ratings (on a 1–5 scale)



As shown in the graph above, participants rated items with a greater sonority distance higher than those with smaller distances. This difference was statistically significant for rising sonority but not for falling sonority.

4.4.3. Discussion

Experiment C was designed to further examine the role of the SCL in the phonological system of Hebrew speakers, with additional focus on the effect of the SDP. Following the same procedure as Experiment A, participants provided lexical judgments on nonce verbs with varying phonological characteristics to examine preference patterns.

The results revealed a significant interaction between sonority distance (±1 or ±2) and direction (rise or fall). A small sonority distance (±1) corresponded to a preference for sonority fall (SCL satisfaction), whereas a large distance (±2) aligned with a preference for sonority rise (SCL violation). These findings challenge the SCL, suggesting its context-dependent interaction with other constraints and lexical form distribution.

In addition to the SCL, the experiment highlighted the active role of the SDP, which reflects a preference for greater sonority distances between adjacent consonants. This tendency emerged despite the lack of a discernible effect of the SDP in the Hebrew lexicon, pointing to the influence of universal constraints in shaping phonological preferences. Furthermore, the results reinforced the active status of the OCP within the phonological system of Hebrew speakers.

The discussion is organized as follows. First, it reexamines the role of the OCP within the phonological system of Hebrew speakers. Next, it explores the interaction between the SCL and the distribution in the lexicon, followed by an alternative analysis based on syllabification. Finally, it addresses the contribution of the SDP to the observed preferences, emphasizing its role in the phonological system.

4.4.3.1. OCP

The current experiment replicates the OCP findings from Experiments A and B, highlighting Hebrew speakers' sensitivity to OCP violations. The mean rating for the OCP group fell below chance levels ($M < 3$), with ratings significantly lower than those of all other groups in this experiment. These findings reinforce the status of the OCP as an active constraint within the phonological system of Hebrew speakers.

4.4.3.2. SCL

The results of Experiment C present a complex picture of the SCL within the phonological system of Hebrew speakers, revealing varying sensitivity across sonority distances. Assuming the common syllabification of Hebrew, where medial consonant sequences are typically divided across syllables (e.g., CVC.CV), participants tolerated SCL violations more when the sonority distance was small (+1). This contradicts the SCL's prediction that violation severity increases with greater sonority distance.

The following subsection explains these results by examining the alignment (or lack thereof) between participants' judgments and the lexicon. Subsequently, an alternative explanation will be considered, based on a different syllabification pattern.

4.4.3.2.1. *The Emergence of the Unmarked: beyond lexical influence*

This subsection explores the relations between lexical patterns and participants' sensitivity to the SCL, focusing on the alignment between corpus data and experimental findings.

Table (39) below presents data from the *heTenTen* corpus (see §3.2.1), showing the frequency of consonant pair groups tested in the experiment and participants' judgment ratings from Experiment C, enabling a direct comparison between corpus patterns and experimental findings on SCL satisfaction. The analysis focuses on the C_2C_3 phonological environment to ensure consistency between the datasets.[21]

(39) Corpus and experimental data by sonority distance and direction

	+ sonority rise SCL violation	– sonority fall SCL satisfaction
1 obs, obs	**Stp-Fri** stop-fricative	**Fri-Stp** fricative-stop
	Lexicon: 141 ExpC: 3.09 (0.72)	Lexicon: 98 ExpC: 3.23 (0.67)
2 obs, son	**Stp-Nas** stop-nasal	**Nas-Stp** nasal-stop
	Lexicon: 95 ExpC: 3.5 (0.61)	Lexicon: 39 ExpC: 3.34 (0.65)

Lexicon = frequency of consonant pairs in the *heTenTen* corpus; ExpC = means and standard deviations of lexicality judgment ratings on a 1–5 scale

The corpus data revealed a statistically significant interaction between sonority distance and direction ($\chi^2(1) = 5.23$, $p = 0.02$, $\varphi = 0.12$), indicating that the distribution in the

[21] The analysis focuses on the C_2C_3 environment to align with the experimental conditions and ensure comparability between the corpus and experimental data. This choice accounts for the effects of spirantization in Hebrew, which influence the distribution of stops and fricatives in adjacent consonants. In the C_3 position, *b* and *p* are absent, replaced by *v* and *f,* as seen in forms like *ganva* 'stole PAST.3FSG' and *katfa* 'pick PAST.3FSG'. By contrast, stops such as *g*, *t*, and *k* (derived from historical *q*) do not undergo spirantization and can occur in this position. Additionally, stridents were excluded from the corpus analysis due to their distinct behavior with respect to sonority-related constraints (see §2.2).

corpus differs from what would be expected without an association between these variables. Unlike the experimental results, the corpus contained more rising sonority pairs than falling ones at both distances, with a stronger preference for rising sonority at the greater sonority distance. At a sonority distance of ±1, the ratio of rising to falling sonority was 141:98 (approximately 1.44 rises per fall), while at ±2, this ratio increased to 95:39 (about 2.44 rises per fall). These findings suggest that the preference for rising sonority becomes more pronounced at the larger distance, with implications for the role of the SCL in shaping phonological preferences.

In the common syllabification of Hebrew, where medial consonant sequences are typically divided across syllables (e.g., CVC.CV), a rise in sonority violates the SCL, as the coda of the first syllable is less sonorant than the onset of the second syllable. In contrast, a fall in sonority satisfies the SCL, as the coda of the first syllable is more sonorant than the onset of the second syllable. Therefore, the corpus results above suggest insensitivity to the SCL, as both sonority intervals (±1 and ±2) show more violations (rising sonority) than satisfactions (falling sonority). This aligns with the full corpus analysis in Chapter 3, which indicates that Hebrew generally lacks sensitivity to the SCL.

Participants in Experiment C showed different preferences across the two sonority distances. At a sonority distance of ±2, their behavior aligned with the lexicon, showing no sensitivity to SCL violations. In contrast, at a sonority distance of ±1, where the lexicon showed a smaller ratio between satisfaction of and violation of the SCL, the influence of the universal constraint became evident. In other words, when the lexical pattern did not demonstrate a strong tendency, sensitivity to the constraint likely originated from an internal source, the universal constraint.

The participants' preferences can be interpreted as an instance of *The Emergence of the Unmarked* (TETU), a concept introduced by McCarthy and Prince (1994). TETU refers to situations where marked structures, though typically permitted in a language, are restricted in specific contexts, leading to the emergence of their less marked counterparts.

In Hebrew, TETU is evident in several linguistic phenomena. The trochaic foot is generally regarded as the less marked metrical structure, and in quantity-insensitive languages, stress systems are expected to consist of syllabic trochaic feet (Hayes 1995). In line with this tendency, children acquiring Hebrew show a preference for trochaic stress patterns, that is, the less marked option, even though in Hebrew this pattern is less frequent than ultimate stress (Adam and Bat-El 2009, Yariv 2021). Stress in acronyms is also often penultimate, reflecting the same preference for unmarked trochaic stress (Bat-El 1994, Zadok 2002). For example, *táχbaʦ* ('public transport') is derived from *taχbuʁá* 'transport' + *ʦibuʁít* 'public'. However, many acronyms still exhibit final stress, such as *natáʦ* 'public transport lane', derived from *natív* 'lane' + *taχburá* 'transport' + *ʦiburít* 'public'.

TETU is also evident in loanwords, where some undergo vowel harmony despite Hebrew lacking synchronic vowel harmony in its lexicon. For instance, the word ***k*o*l*o*ʁabi*** 'kohlrabi' (from English *koʊlraːbi*, originally German *koːlʁaːbi*) demonstrates vowel harmony, with the second *o* serving as an epenthetic vowel that assimilates to the first vowel's quality (Cohen 2013).

In the same vein, the participants in the experiment favored the less marked structure, which satisfies the SCL, despite the general lack of sensitivity to this constraint in the Hebrew lexicon. Unlike typical TETU phenomena, which are usually independent of lexical patterns, sensitivity to SCL violations among participants emerged only at a sonority distance of ±1, where the lexicon shows a smaller ratio of SCL violations to satisfactions.

Table (40) highlights the interplay between lexical patterns and universal constraints (UG) in shaping sensitivity to the SCL, as observed in Experiment C.

(40) SCL sensitivity: Hebrew lexicon vs. Experiment C

<table>
<tr><th></th><th>Hebrew Lexicon</th><th>Experiment C</th></tr>
<tr><td rowspan="2">1
obs, obs</td><td>SCL violations occur at a lower ratio compared to sonority distance 2.</td><td>Nonce verbs that satisfy the SCL were rated higher as possible Hebrew verbs compared to those that violate the SCL.</td></tr>
<tr><td colspan="2">When the ratio of SCL violations in the lexicon is relatively low, participants demonstrate sensitivity to the SCL, despite the presence of violations. This sensitivity likely originates from an internal source, such as universal constraints (UG), aligning with TETU.</td></tr>
<tr><td rowspan="2">2
obs, son</td><td>SCL violations occur at a higher ratio compared to sonority distance 1.</td><td>Nonce verbs that violate the SCL were rated higher as possible Hebrew verbs compared to those that satisfy the SCL.</td></tr>
<tr><td colspan="2">Participants preferred SCL-violating forms, consistent with the high violation rate in the lexicon, indicating the lexicon as an external source.</td></tr>
</table>

4.4.3.2.2. Alternative Analysis: Syllabification

The analysis above assumes the traditional syllabification pattern CVC.CV (e.g., *sag.ma*, *sag.mu*), which reflects the tendency to avoid consonant clusters and allow codas. However, the alternative syllabification pattern proposed in Albert (2014, 2019), CV.CCV (e.g., *sa.gma*, *sa.gmu*), should also be considered. This sub-section discusses why the CV.CCV syllabification can be considered a plausible option in Modern Hebrew and how it could account for the experimental results.

Traditionally, Hebrew has been thought to divide medial consonant sequences into CVC.CV. However, unlike Biblical Hebrew, Modern Hebrew permits consonant clusters, reflecting greater flexibility in syllable structure and allowing for CV.CCV syllabification. This flexibility is evident in the acceptance of clusters across various word positions. Clusters occur at the beginning of words (***kt**i.va* 'writing', ***pʁ**a.χim* 'flowers') and at the end of loanwords (*pʁo.je**kt*** 'project', *sla**ng*** 'slang'), albeit less frequently.[22] Medial clusters also occur in loan nouns, with variation in syllabification, as seen in *am.ba.**tj**a* / *am.ba**t**.**j**a* ('bathtub') and *de.mo.**kʁ**a.**tj**a* / *de.mo**k**.**ʁ**a**t**.**j**a* / *de.mo.**kʁ**a**t**.**j**a* / *de.mo**k**.**ʁ**a.**tj**a* ('democracy'). Asherov and Bat-El (2019) and Albert (2014, 2019) note that there are no phonological phenomena favor a particular type of syllabification in Hebrew, leaving syllable boundaries ambiguous in some cases. These findings suggest that dividing medial CC sequences into a single syllable under certain conditions, thereby creating a medial complex onset, is a plausible option.

Following the plausibility of CV.CCV syllabification in Hebrew, the constraints violated by each syllabification pattern can be examined. In the CV.CCV pattern, the SCL is irrelevant, as there is no consonant contact across the syllable boundary. However, this pattern violates *COMPLEX, as the second syllable contains a complex onset. In contrast, the CVC.CV pattern violates the SCL in rising sonority cases (e.g., *sag.ma*), where the coda of the first syllable is less sonorous than the onset of the following syllable. Prosodically, this pattern violates *CODA, as the first syllable includes a coda; however, this constraint is freely violated in Hebrew. Table (41) summarizes the constraints violated by each syllabification pattern.

---

[22] In addition to word-final clusters in PAST.2FSG verbs (e.g., *kata**vt*** 'write PAST.2FSG'), which are also permitted in Biblical Hebrew (Gesenius 1910, Blau 2010, among others).

(41) Summary of constraints violated by CVC.CV and CV.CCV syllabification

	CVC.CV	**CV.CCV**
SCL	Violated in rising sonority cases, where the coda of the first syllable is less sonorous than the onset of the following syllable	Not relevant: no consonant contact across the syllable boundary
Prosodic Constraint	*CODA	*COMPLEX

Given the two possible syllabification patterns, the results of Experiment C can be interpreted through an alternative analysis. When the sonority distance was 2, the Stp-Nas group (e.g., *sagma*) received higher ratings than the Nas-Stp group (e.g., s*amga*). This difference can be attributed to distinct syllabification patterns between the two groups. In the Stp-Nas group, CV.CCV syllabification (*sa.gma*) is likely preferred, creating a complex onset that is permissible in Hebrew and avoids the SCL violation associated with CVC.CV (*sam.ga*). In contrast, in the Nas-Stp group, CVC.CV syllabification (*sam.ga*) is preferred. While it violates *CODA, a frequently violated constraint in Hebrew, it satisfies the SCL. The alternative CV.CCV syllabification (*sa.mga*), which creates a complex onset, violates the Sonority Sequencing Generalization (SSG), which requires sonority to rise, or at least not fall, toward the nucleus (Clements 1990; Selkirk 1984). This principle is widely satisfied in Hebrew.

Consequently, the syllabification pattern in the Stp-Nas group is proposed as CV.CCV (*sa.gma*), while in the Nas-Stp group, it is CVC.CV (*sam.ga*). Both patterns align with Hebrew phonotactics, violating only structural constraints, and not phonotactic ones. The participants' preference for Stp-Nas sequences (*sagma*) may reflect the lexical distribution, as Stp-Nas pairs are more frequent than Nas-Stp pairs in the Hebrew lexicon (see Table 39).

When the sonority distance was 1, an opposite pattern emerged: the Fri-Stp group (*ʦavga*) received higher ratings than the Stp-Fri group (*ʦagva*). In this case, the syllabification is proposed to be CVC.CV for both groups. A sonority distance of 1 between two obstruents may be too small to favor a cluster-forming syllabification pattern. This interpretation aligns with the SDP, which states that a greater sonority distance between consonants is generally more favorable. Assuming CVC.CV syllabification, the difference in ratings is explained by the SCL: in the Fri-Stp group (*ʦav.ga*), the SCL is satisfied, whereas in the Stp-Fri group (*ʦag.va*), it is violated.

This suggests that participants assigned higher ratings to forms that satisfy the SCL, even though the language does not consistently satisfy this constraint. Such behavior provides evidence of an internal source influencing the phonological system of the speakers and reflects *The Emergence of the Unmarked* (§4.4.3.2.1).

To conclude, the results can be explained by the following constraint ranking:

***COMPLEX OBS-OBS >> SCL >> *COMPLEX OBS-SON**

When choosing between violating the SCL and violating *COMPLEX OBS-SON (i.e., a cluster consisting of an obstruent and a sonorant), the higher-ranked SCL dictates the syllabification pattern as CV.CCV. In this case, The SCL is not violated, while the lower-ranked *COMPLEX OBS-SON is. Conversely, when the choice is between violating the SCL and violating *COMPLEX OBS-OBS (i.e., a cluster of two obstruents), the higher-ranked *COMPLEX OBS-OBS prevents the complex onset, resulting in CVC.CV syllabification. Since *CODA is ranked low in Hebrew, this syllabification is acceptable. In such cases, the violation of the SCL in the Stp-Fri group (*ʦag.va*) accounts for its lower ratings compared to the Fri-Stp group (*ʦav.ga*).

Thus, differences in syllabification patterns and the interaction between phonotactic and structural constraints account for the experimental results. However, given the variability in syllabification of medial clusters among Hebrew speakers, it cannot be assumed that all participants divided the syllables in the same way. Therefore, this analysis is proposed as an alternative explanation.

4.4.3.3. SDP

The results of Experiment C indicate a clear tendency toward higher sonority distances, as participants consistently assigned higher ratings to groups with greater sonority distances between consonants (Nas-Stp > Fri-Stp, Stp-Nas > Stp-Fri). While the difference between the Nas-Stp and Fri-Stp groups did not reach statistical significance, the Stp-Nas group received significantly higher ratings than the Stp-Fri group. Notably, this pattern contrasts with the lexicon, which does not exhibit a preference for larger sonority distances (see §3.4.2).

While this preference is evident, it is not without nuance. First, the tendency does not consistently reach statistical significance across all comparisons. Second, some consonant pairs in the Stp-Nas group were more frequent in the lexicon than those in the Fri-Stp group (see Appendix I). However, the lack of a corresponding preference in the lexicon itself strengthens the argument that this tendency originates from an internal source, such as universal constraints, rather than being lexicon-driven. These findings highlight the potential role of universal constraints in shaping phonological preferences, demonstrating their ability to guide preferences independently of lexical distributions.

Chapter 5

Concluding Remarks

Universal phonotactic constraints play a fundamental role in defining permissible consonant sequences in a language, shaping speakers' segmental preferences. This study examined the interaction between universal constraints and lexical frequencies in shaping restrictions within native Hebrew speakers' phonological systems. While lexical frequencies play a significant role in shaping these systems, they may not fully account for all patterns that speakers' phonological systems permit. This study explores whether universal constraints can account for patterns in phonological systems that operate beyond the influence of the lexicon.

To address this matter, three phonotactic constraints were examined:

a. **The Obligatory Contour Principle** (OCP; Leben 1973; Goldsmith 1976; McCarthy 1979, 1981, 1986), which restricts the occurrence of identical elements within a melodic level;
b. **The Syllable Contact Law** (SCL; Murray and Vennemann 1983, Vennemann 1988, Clements 1990), which governs the contact between adjacent syllables, favoring a coda that is more sonorous than the onset of the following syllable;
c. **The Sonority Dispersion Principle** (SDP; Clements 1990), which states that greater sonority distances between adjacent segments are preferred over smaller ones.

The influence of these constraints was evaluated with reference to two types of data. First, an extensive corpus analysis examined their effect in the Modern Hebrew lexicon (Chapter 3). Second, three psycholinguistic experiments tested their role in the phonological system of native Hebrew speakers (Chapter 4).

The corpus analysis was based on the *heTenTen* corpus, a large and diverse collection of Modern Hebrew texts, providing a comprehensive representation of the language's lexicon. The analysis focused on tri-consonantal verb stems, chosen to ensure systematic examination of consonant co-occurrence patterns in adjacent (C_1C_2, C_2C_3) and non-adjacent (C_1VC_2, C_2VC_3) positions. This approach allowed for a detailed examination of co-occurrence restrictions across different phonological environments. The results revealed that the OCP is strongly satisfied in the Hebrew lexicon, with clear restrictions on consonants that share a place of articulation. In contrast, the SCL and the SDP are heavily violated in the Modern Hebrew lexicon, with no clear evidence of their activity in shaping the distribution of consonant pairs.

The three psycholinguistic experiments explored the role of these phonotactic constraints in the phonological system of native Hebrew speakers. The first experiment, a lexical judgment task, measured sensitivity to OCP and SCL violations by asking participants to rate the acceptability of constraint-violating nonce verbs. The second experiment, a lexical decision task, assessed reaction time and accuracy in identifying well-formed versus constraint-violating nonce verbs, providing further insights into the impact of these constraints. The third experiment focused on the SCL, examining preferences for nonce verbs with varying sonority distances (±1 or ±2) and sonority profiles (rising or falling). This also allowed an evaluation of the SDP, which favors greater sonority distances. Together, these experiments investigated the interplay between universal constraints and Hebrew speakers' phonological preferences.

The experimental results led to the following conclusions:

a. **OCP:** The results demonstrate that the OCP is active in the phonological system of native Hebrew speakers. Participants consistently showed sensitivity to violations of the OCP, as reflected in their lower acceptability ratings (Experiments A and C) and reaction time (Experiment B) when identifying them as nonce verbs. This aligns with the findings from the corpus analysis, where the OCP was shown to strongly influence the phonological structure of words, disfavoring consonant pairs that share the same place of articulation. Therefore, these results cannot tease apart the source of OCP effects in speakers' mental system, whether it is from the lexicon, UG, or both.

b. **SCL:** The results revealed a nuanced sensitivity to the SCL, varying by the sonority relation between adjacent consonants. In pairs consisting of an obstruent and a sonorant, participants significantly preferred SCL-violating forms, favoring obstruent-sonorant over sonorant-obstruent pairs. This preference aligns with the lexicon, which also favors obstruent-sonorant (SCL-violating) pairs. In contrast, for obstruent-obstruent pairs, participants significantly preferred forms that satisfied the SCL, a preference that contradicts the lexicon. Notably, the preference for SCL-violating pairs in the lexicon was stronger in obstruent-sonorant pairs than in obstruent-obstruent pairs.

   This divergence indicates that the SCL exerts a stronger influence in obstruent-obstruent pairs, overriding lexical tendencies. This tendency reflects the phenomenon of *The Emergence of the Unmarked* (McCarthy and Prince 1994), wherein universally less marked forms are preferred, even when lexical frequencies suggest otherwise (see §4.4.3.2.1). An alternative analysis (§4.4.3.2.2), based on a different syllabification pattern, further highlights the role of the SCL in shaping phonological judgments, even in the absence of strong lexical evidence for its activity.

c. **SDP:** Although the lexicon shows no discernible evidence of the SDP, the constraint appears to influence the phonological systems of speakers. Participants consistently preferred nonce verbs with larger sonority distances between adjacent consonants, aligning with the predictions of the SDP. This suggests that the SDP shapes phonological judgments independently of lexical patterns, highlighting its role as a universal constraint that operates beyond the lexicon.

The findings from the experiments highlight the activity of each of the three constraints, as summarized in Table (42).

(42) Summary of phonotactic constraint activity (OCP, SCL, SDP) in the lexicon and experimental data

	OCP	**SCL**		**SDP**
Lexicon	✓	✗		✗
Experimental Data	✓	✓ obs-obs	✗ obs-son	✓

This table highlights the distinct patterns of activity for each constraint in both the lexicon and experimental data, illustrating the nuanced role of universal constraints in shaping phonological systems. ✓ indicates cases where the constraint influences the data, while ✗ indicates no discernible effect. The distinction between *obs-obs* (obstruent-obstruent pairs) and *obs-son* (obstruent-sonorant pairs) reflects differences in how the SCL operates across varying sonority distances.

To conclude, the study explored how universal constraints shape phonotactic restrictions in the phonological system of speakers in light of lexical frequencies. By focusing on three phonotactic constraints – OCP, SDP, and SCL – and analyzing data from corpus studies and psycholinguistic experiments, the research revealed a complex interplay between universal constraints and speakers' phonological systems. The findings demonstrate that while some constraints, such as the OCP, align closely with patterns observed in the lexicon, others, including the SDP and aspects of the SCL, influence phonological preferences in ways that cannot be solely attributed to lexical frequencies. This highlights the unique contributions of universal constraints in shaping phonological systems, both through their lexical patterns and beyond.

This study contributes to the broader understanding of phonological theory by shedding light on the role of universal constraints in shaping phonotactic preferences, focusing on Modern Hebrew. By highlighting the interplay between universal constraints and language-specific factors, it reveals the roles of constraints reflected in the lexicon and those operating beyond it. While centered on Hebrew, the study invites future research to explore whether these patterns extend to other lexical categories or to constraints in different languages, enriching cross-linguistic understanding. This dissertation thus adds to the broader field of phonological theory, enhancing our understanding of how universal constraints interact with the unique properties of individual languages.

References

Adam, G. (2002). *From variable to optimal grammar: Evidence from language acquisition and language change* [Doctoral dissertation, Tel Aviv University].

Adam, G., & Bat-El, O. (2009). When do universal preferences emerge in language development? The acquisition of Hebrew stress. *Brill's Annual of Afroasiatic Languages and Linguistics*, *1*, 1–28.

Adriaans, F., & Kager, R. (2010). Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language*, *62*(3), 311–331.

Albert, A. (2014). *Phonotactic universals in Modern Hebrew: Evidence for prosodic alignment of stops* [Master's thesis, Tel Aviv University].

Albert, A. (2019). The state of stop–fricative alternation in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics*, *11*(1), 135–161.

Albert, A. (2023). *A model of sonority based on pitch intelligibility*. BoD–Books on Demand.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*(2), 119–161.

Arad, M. (2005). *Roots and patterns: Hebrew morpho-syntax*. Springer Science & Business Media.

Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes*. MIT Press.

Asherov, D., & Bat-El, O. (2019). Syllable structure and complex onsets in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics*, *11*(1), 69–95.

Asherov, D., & Cohen, E. G. (2019). A phonetic description of Modern Hebrew consonants and vowels. *Brill's Journal of Afroasiatic Languages and Linguistics*, *11*(1), 3–27.

Avinery, Y. (1965). *Yad HaLashon*. Yizre'el. [in Hebrew]

Bat-El, O. (1989). *Phonology and word structure in Modern Hebrew* [Doctoral dissertation, University of California, Los Angeles].

Bat-El, O. (1994). Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory*, *12*(4), 572–596.

Bat-El, O. (1994). The optimal acronym word in Hebrew. In P. Koskinen (Ed.), *Proceedings of the 1994 annual conference of the Canadian Linguistic Association* (pp. 23–37). Toronto Working Papers in Linguistics.

Bat-El, O. (1996). Selecting the best of the worse: The grammar of Hebrew blends. *Phonology*, *13*(3), 283–328.

Bat-El, O. (2003). Semitic verb structure within a universal perspective. In J. Shimron (Ed.), *Language processing and acquisition in languages of Semitic, root-based, morphology* (pp. 29–59). John Benjamins Publishing Company.

Bat-El, O. (2008). Morphologically conditioned V–Ø alternation in Hebrew: Distinction among nouns, adjectives & participles, and verbs. In S. Armon-Lotem, G. Danon, & S. Rothstein (Eds.), *Current issues in generative Hebrew linguistics* (pp. 27–60). John Benjamins Publishing Company.

Becker, M., & Gouskova, M. (2016). Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry*, *47*(3), 391–425.

Ben-David, A., & Bat-El, O. (2016). Paths and stages in acquisition of the phonological word in Hebrew. In R. Berman (Ed.), *Acquisition and development of Hebrew: From infancy to adolescence* (pp. 39–68). John Benjamins Publishing Company.

Berent, I., Everett. D. L., & Shimron, J. (2001). Do phonological representations specify variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology*, *42*(1), 1–60.

Berent, I., Lennertz, T., & Rosselli, M. (2012). Universal phonological restrictions and language-specific repairs: Evidence from Spanish. *The Mental Lexicon*, *13*, 275–305.

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, *105*(14), 5321–5325.

Berent, I., & Shimron, J. (1997). The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition*, *64*(1), 39–72.

Berman, R. (1978). *Modern Hebrew structure*. Tel Aviv University, University Publishing Projects.

Berrebi, S. (2021). *Hearing ethnicity: Classification, stereotypization and processing of socially marked phonetic features in Modern Hebrew* [Doctoral dissertation, Tel Aviv University & École normale supérieure – PSL].

Berrebi, S., Bassel, N., & Gafter, R. J. (2022). Hearing Hebrew pharyngeals: Experimental evidence for a covert phonemic distinction. *University of Pennsylvania Working Papers in Linguistics*, *28*(1), 11–20.

Bessell, N. J. (1992). *Towards a phonetic and phonological of post-velar articulation* [Doctoral dissertation, University of British Columbia].

Blau, J. (2010). *The phonology and morphology of Biblical Hebrew*. The Academy of Hebrew Language. [in Hebrew]

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Bolozky, S. (1978). Some aspects of Modern Hebrew phonology. In R. Berman (Ed.), *Modern Hebrew structure* (pp. 11–67). Universities Publishing Projects.

Bolozky, S. (1997). Israeli Hebrew phonology. *Phonologies of Asia and Africa*, *1*, 287–311.

Bolozky, S., & Becker, M. (2006). *Living lexicon of Hebrew nouns* [Manuscript]. University of Massachusetts Amherst. *http://becker.phonologist.org/LLHN*

Bolozky, S., & Kreitman, R. (2007). Uvulars in Israeli Hebrew – their phonetic and phonological status. Paper presented at *The National Association of Professors of Hebrew's International Conference on Hebrew Language, Literature and Culture*, Sydney, Australia.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425–455.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language 82*(4), 711–733.

Chomsky, N. (2006). *Language and mind*. Cambridge University Press.

Clements, G. N. (1985). The geometry of phonological features. *Phonology*, *2*, 225–252.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 282–333). Cambridge University Press.

Cohen, E. G. (2013). The emergence of the unmarked: Vowel harmony in Hebrew loanword adaptation. *Lingua*, *131*, 66–79.

Cohen, E. G., (2019). Loanword phonology in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics*, *11*(1), 182–200.

Cohen, E. G., Laks, L., & Savu, C. (2019). The phonetics of Modern Hebrew rhotics. *Brill's Journal of Afroasiatic Languages and Linguistics*, *11*(1), 28–48.

Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., & Norrmann, I. (2011). Explaining sonority projection effects. *Phonology*, *28*(2), 197–234.

Doron, E. (2003). Agency and voice: The semantics of the Semitic templates. *Natural Language Semantics*, *11*(1), 1–67.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. P., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568–1578.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*(3), 199–210.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. MIT Press.

Enguehard, G., & Faust, N. (2018). Guttural ghosts in Modern Hebrew. *Linguistic Inquiry*, *49*(4), 685–721.

Faust, N. (2005). *The fate of Hebrew gutturals* [Master's thesis, Tel-Aviv University].

Faust, N. (2019). Gutturals in General Israeli Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics*, *11*(1), 162–181.

Foley, J. (1972). Rule precursors and phonological change by meta-rule. In R. S. Stockwell & R. K. Macaulay (Eds.), *Linguistic change and generative theory* (pp. 96–100). Indiana University Press.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, *22*(1), 179–228.

Frisch, S. A., & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place in Arabic, *Language*, *77*(1), 91–106.

Fudge, E. C. (1969). Syllables. *Journal of Linguistics*, *5*(2), 253–286.

Gafter, R. J. (2014). *"The most beautiful and correct Hebrew": Authenticity, ethnic identity and linguistic variation in the greater Tel Aviv area* [Doctoral dissertation, Stanford University].

Gesenius, W., & Kautzsch E. F. (1910). *Gesenius' Hebrew grammar*. Clarendon Press.

Goad, H. (2011). The representation of sC clusters. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The blackwell companion to phonology* (Vol. 2, pp. 898–923). Wiley-Blackwell.

Goldsmith, J. A. (1976). *Autosegmental phonology* [Doctoral dissertation, Massachusetts Institute of Technology].

Goldwater, S., Johnson, M., & Griffiths, T. L. (2006). Interpolating between types and tokens by estimating power-law generators. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 459–466). MIT Press.

Goshen-Gottstein, M. (1964). Semitic morphological structures: The basic morphological structure of Biblical Hebrew. In H. B. Rosen (Ed.), *Studies in egyptology and linguistics in honour of H. J. Polotsky* (pp. 104–116). Israel Exploration Society.

Gouskova, M. (2001). Falling sonority onsets, loanwords, and syllable contact. In M. Andronis, C. Ball, H. Elston, & S. Neuvel (Eds.), *CLS 37: The main session* (Vol. 1, pp. 175–185). Chicago Linguistic Society.

Gouskova, M. (2002). Exceptions to sonority distance generalizations. In M. Andronis, E. Debenport, A. Pycha, & K. Yoshimura (Eds.), *CLS 38: The main session* (Vol. 1, pp. 253–268). Chicago Linguistic Society.

Greenberg, J. H. (1950). The patterning of root morphemes in Semitic. *Word*, *6*(2), 162–181.

Greenberg, J. H. (Ed.). (1978). *Universals of human language*. Stanford University Press.

Handelsman, N., Haim, M., Yariv, D., Gafni, C., Asherov, D., Adam, G., & Bat-El, O. (2021). The phonology of nouns and verbs in early Hebrew speech. In G. Danon (Ed.), *Proceedings of IATL 2018–2019* (MIT Working Papers in Linguistics, 92, pp. 103–119). MIT Working Papers in Linguistics.

Hay, J., Pierrehumbert, J., & Beckman, M. (2004). Speech perception, well-formedness, and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI: Phonetic interpretation* (pp. 58–74). Cambridge University Press.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Hayes, B., Kirchner, R., & Steriade, D. (2004). Introduction. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 1–33). Cambridge University Press.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379–440.

Hayes, B., Zuraw, K., Siptár, P., & Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, *85*(4), 822–863.

Horvath, J., & Siloni, T. (2008). Active lexicon: Adjectival and verbal passives. In S. Armon-Lotem, G. Danon, & S. Rothstein (Eds.), *Current issues in generative Hebrew linguistics* (pp. 105–134). John Benjamins Publishing Company.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013, July). The TenTen corpus family. In A. Hardie & R. Love (Eds.), *Abstract book of the 7th International Corpus Linguistics conference (CL 2013)* (pp. 125–127). UCREL, Lancaster University.

Kaisse, E. (1988). *Modern Greek continuant dissimilation and the OCP* [Manuscript]. University of Washington.

Kenstowicz, M. J. (1994). *Phonology in generative grammar*. Cambridge: Blackwell Publishers.

Laks, L. (2011). *Morpho-phonological and Morpho-thematic relations in Hebrew and Arabic verb formation* [Doctoral dissertation, Tel Aviv University].

Laks, L., Cohen, E. G., & Azulay-Amar S. (2016). Paradigm uniformity and the locus of derivation: the case of vowel epenthesis in Hebrew verbs. *Lingua*, *170*, 1–22.

Laufer, A. (1990). Hebrew. *Journal of the International Phonetic Association*, *20*(2), 40–43.

Leben, W. R. (1973). *Suprasegmental phonology* [Doctoral dissertation, Massachusetts Institute of Technology].

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, *49*(1), 199–227.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge University Press.

Matras, Y., & Schiff, L. (2005). Spoken Israeli Hebrew revisited: Structures and variation. *Studia Semitica: Journal of Semitic Studies Supplement*, *16* (Jubilee Volume), 145–191.

McCarthy, J. (1979). *Formal problems in Semitic phonology and morphology* [Doctoral dissertation, Massachusetts Institute of Technology].

McCarthy, J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, *12*(3), 373–418.

McCarthy, J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, *17*(2), 207–263.

McCarthy, J. (1988). Feature geometry and dependency: A review. *Phonetica*, *45*(2-4), 84–108.

McCarthy, J., & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. *NELS*, *24*, 333–379.

Meloni, C. (2021). *The Reš riddle: Identifying the Biblaical Hebrew rhotic* [Master's thesis, Tel Aviv University].

Mintz, T. H., Newport, E. L, & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*(4), 393–424.

Morelli, F. (2003). The relative harmony of /s+stop/ onsets: Obstruent clusters and the sonority sequencing principle. In C. Féry & R. van de Vijver (Eds.), *The syllable in Optimality Theory* (pp. 356–371). Cambridge University Press.

Murray, R. W., & Vennemann, T. (1983). Sound change and syllable structure in Germanic phonology. *Language*, *59*(3), 514–528.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49*(2), 85–117.

Padgett, J. (1995). *Stricture in feature geometry.* SCLI Publications.

Parker, S. G. (2002). *Quantifying the sonority hierarchy* [Doctoral dissertation, University of Massachusetts Amherst].

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Peretz, Y. (1971). *Ivrit Kahalakhah*. Y. Sreberk. [In Hebrew]

Pierrehumbert, J. B. (1993). Dissimilarity in the Arabic verbal roots. In A. J. Schafer (Ed.), *Proceedings of the North East Linguistic Society 23* (Vol. 1, pp. 367–381). GLSA, University of Massachusetts at Amherst.

Pierrehumbert, J. B. (2001). Stochastic phonology. *Glot International, 5*(6), 195–207.

Pimentel, T., Roark, B., & Cotterell, R. (2020). Phonotactic complexity and its trade-offs. *Transactions of the Association for Computational Linguistics*, *8*, 1–18.

Poletiek, F. H., & Wolters, G. (2009). What is learned about fragments in artificial grammar learning? A transitional probabilities approach. *The Quarterly Journal of Experimental Psychology, 62*(5), 868–876.

Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. Wiley-Blackwell.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, *275*(5306), 1604–1610.

Rasin, E., & Katzir, R. (2016). On evaluation metrics in Optimality Theory. *Linguistic Inquiry*, *47*(2), 235–282.

Rebuschat, P., & Williams, J. N. (2012). *Statistical learning and language acquisition*. De Gruyter Mouton.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive science*, *22*(4), 425–469.

Reinhart, T., & Siloni, T. (2005). The Lexicon-syntax parameter: reflexivization and other arity operations. *Linguistic Inquiry*, *36*(3), 389–436.

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

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, *80*(3), 475–531.

Sagey, E. (1986). *The representation of features and relations in non-linear phonology* [Doctoral dissertation, Massachusetts Institute of Technology].

Schwarzwald (Rodrigue), O. (2002). *Studies in Hebrew morphology* (Vol. 3). Open University. [in Hebrew]

Schwarzwald (Rodrigue), O. (2008). The special status of nif'al in Hebrew. In S. Armon-Lotem, G. Danon, & S. Rothstein (Eds.), *Current issues in generative Hebrew linguistics* (pp. 61–76). John Benjamins Publishing Company.

Schwarzwald (Rodrigue), O. (2019). *Studies in contemporary Hebrew*. The Academy of Hebrew Language. [in Hebrew]

Selkirk, E. (1984). On the major class features and the syllable theory. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure* (pp. 107–144). MIT Press.

Sima'an, K., Itai, A., Winter, Y., Altman, A., & Nativ, N. (2001). Building a tree-bank of Modern Hebrew text. *Traitement Automatique des Langues*, *42*(2), 247–380.

Smolensky, P., & Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar (Cognitive architecture)*. MIT press.

Steriade, D. (1982). *Greek prosodies and the nature of syllabification* [Doctoral dissertation, Massachusetts Institute of Technology].

Steriade, D. (1988). Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology*, *5*(1), 73–155.

Steriade, D. (1999). Alternatives to syllable-based accounts of consonantal phonotactics. In O. Fujimura, B. D. Joseph, & B. Palek (Eds.), *Proceedings of the 1998 Linguistics and Phonetics conference* (pp. 205–242). Karolinum Press.

Suchomel, V., & Pomikálek, J. (2012, April). Efficient web crawling for large text corpora. In A. Kilgarriff & S. Sharoff (Eds.), *Proceedings of the seventh Web as Corpus workshop (WAC7)* (pp. 39–43). Lyon, France.

Tomasello, M. (1998). Introduction: A cognitive-functional perspective on language structure. In M. Tomasello (Ed.), *The new psychology of language: cognitive and functional approaches to language structure* (Vol. 1, pp. vii–xxiii). Lawrence Erlbaum Associates.

Tomasello, M. (2003). *Constructing a language: A usage-based account of language acquisition*. Harvard University Press.

Ussishkin, A. (1999). The inadequacy of the consonantal root: Modern Hebrew denominal verbs and output–output correspondence. *Phonology*, *16*(3), 401–442.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change*. De Gruyter Mouton.

Yaeger-Dror, M. (1988). The influence of changing group vitality on convergence toward a dominant linguistic norm: An Israeli example. *Language & Communication*, *8*(3–4), 285–305.

Yariv, D. (2021). *The markedness over faithfulness bias: Coda moraicity in early Hebrew speech* [Master's thesis, Tel Aviv University].

Yeverechyahu, H. (2014). *The role of similarity in cooccurrence restrictions: Evidence from the Hebrew verbal system* [Master's thesis, Tel Aviv University].

Yeverechyahu, H. (2019). Consonant co-occurrence restrictions in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics*, *11*(1), 57–68.

Yip, M. (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry*, *19*(1), 65–100.

Zadok, G. (2002). *Abbreviations: A unified analysis of acronym words, clippings, clipped compounds and hypocoristics* [Master's thesis, Tel Aviv University].

Zadok, G. (2012). *On the instability of Hebrew weak verbs: similarity, variation, and change* [Doctoral dissertation, Tel-Aviv University].

Zipf G. K. (1936). *The psychobiology of language*. Routledge.

Zipf G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.

Appendix A
Verb Paradigms

Complete paradigms are presented for all non-passive verb templates (*binyanim*) of regular verbs, in which all three stem consonants are phonetically realized throughout the templates of the inflectional paradigm (Zadok 2012, see §3.2.2.2). Stems appear in bold, and Cs represent the stem's consonants.[23]

binyan qal (pa'al)[24,25]							
Past		Present[26]		Future			Infinitive[27]
1SG	**CaCaC**-ti	MSG	**CoCeC**	1SG	ʔe-**CCoC**	ʔe-**CCaC**	li-**CCoC**
2MSG	**CaCaC**-ta	FSG	**CoCeC**-et	2MSG	ti-**CCoC**	ti-**CCaC**	
2FSG	**CaCaC**-t	MPL	**CoCC**-im	2FSG	ti-**CCeC**-i	ti-**CCeC**-i	
3MSG	**CaCaC**	FPL	**CoCC**-ot	3MSG	i-**CCoC**	i-**CCaC**	
3FSG	**CaCC**-a			3FSG	ti-**CCoC**	ti-**CCaC**	
1PL	**CaCaC**-nu			1PL	ni-**CCoC**	ni-**CCaC**	
2MPL	**CaCaC**-tem			2PL	ti-**CCeC**-u	ti-**CCeC**-u	
2FPL	**CaCaC**-ten			3PL	i-**CCeC**-u	i-**CCeC**-u	
3PL	**CaCC**-u						

[23] As mentioned in §3.2.2.4, imperative forms (e.g., *ktov* 'write, IMP.2MSG') are not part of the current study, as future forms often replace them in Modern Hebrew. The historical plural feminine form (e.g., *tiχtovna* 'write, FUT.2FPL / FUT.3FPL') is also not part of the current study due to its rare use, as masculine forms are commonly used for feminine.

[24] Some verbs in *binyan qal* conjugate in the present tense with the CaCeC template instead of the more common CoCeC template, primarily verbs expressing states rather than actions (e.g., *jaʃen* 'sleep, PRESENT.MSG'). However, these verbs are relatively uncommon.

[25] In the future tense, *binyan qal* has two stems: one with the vowel *o* and another with the vowel *a*. This division, which is lexical rather than phonological, determines whether a verb conjugates with the *o* base (e.g., *iχtov* 'write FUT.3MSG') or the *a* base (e.g., *ilbaʃ* 'wear FUT.3MSG').

[26] Present forms (in all verb templates) are not specified for person, only for number and gender.

[27] Exception: *liʃkav* ('to lie down').

binyan nif'al						
Past		Present		Future		Infinitive
1SG	ni-**CCaC**-ti	MSG	ni-**CCaC**	1SG	ʔe-**CaCeC**	le-hi-**CaCeC**
2MSG	ni-**CCaC**-ta	FSG	ni-**CCeC**-et	2MSG	ti-**CaCeC**	
2FSG	ni-**CCaC**-t	MPL	ni-**CCaC**-im	2FSG	ti-**CaCC**-i	
3MSG	ni-**CCaC**	FPL	ni-**CCaC**-ot	3MSG	i-**CaCeC**	
3FSG	ni-**CCeC**-a			3FSG	ti-**CaCeC**	
1PL	ni-**CCaC**-nu			1PL	ni-**CaCeC**	
2MPL	ni-**CCaC**-tem			2PL	ti-**CaCC**-u	
2FPL	ni-**CCaC**-ten			3PL	i-**CaCC**-u	
3PL	ni-**CCeC**-u					

binyan hif'il						
Past		Present		Future		Infinitive
1SG	hi-**CCaC**-ti	MSG	ma-**CCiC**	1SG	ʔa-**CCiC**	le-ha-**CCiC**
2MSG	hi-**CCaC**-ta	FSG	ma-**CCiC**-a	2MSG	ta-**CCiC**	
2FSG	hi-**CCaC**-t	MPL	ma-**CCiC**-im	2FSG	ta-**CCiC**-i	
3MSG	hi-**CCiC**	FPL	ma-**CCiC**-ot	3MSG	ja-**CCiC**	
3FSG	hi-**CCiC**-a			3FSG	ta-**CCiC**	
1PL	hi-**CCaC**-nu			1PL	na-**CCiC**	
2MPL	hi-**CCaC**-tem			2PL	ta-**CCiC**-u	
2FPL	hi-**CCaC**-ten			3PL	ja-**CCiC**-u	
3PL	hi-**CCiC**-u					

binyan pi'el[28]						
Past		Present		Future		Infinitive
1SG	**CiCaC**-ti	MSG	me-**CaCeC**	1SG	ʔa-**CaCeC**	le-**CaCeC**
2MSG	**CiCaC**-ta	FSG	me-**CaCeC**-et	2MSG	te-**CaCeC**	
2FSG	**CiCaC**-t	MPL	me-**CaCC**-im	2FSG	te-**CaCC**-i	
3MSG	**CiCeC**	FPL	me-**CaCC**-ot	3MSG	je-**CaCeC**	
3FSG	**CiCC**-a			3FSG	te-**CaCeC**	
1PL	**CiCaC**-nu			1PL	ne-**CaCeC**	
2MPL	**CiCaC**-tem			2PL	te-**CaCC**-u	
2FPL	**CiCaC**-ten			3PL	je-**CaCC**-u	
3PL	**CiCC**-u					

binyan hitpa'el						
Past		Present		Future		Infinitive
1SG	hit-**CaCaC**-ti	MSG	mit-**CaCeC**	1SG	ʔet-**CaCeC**	le-hit-**CaCeC**
2MSG	hit-**CaCaC**-ta	FSG	mit-**CaCeC**-et	2MSG	tit-**CaCeC**	
2FSG	hit-**CaCaC**-t	MPL	mit-**CaCC**-im	2FSG	tit-**CaCC**-i	
3MSG	hit-**CaCeC**	FPL	mit-**CaCC**-ot	3MSG	it-**CaCeC**	
3FSG	hit-**CaCC**-a			3FSG	tit-**CaCeC**	
1PL	hit-**CaCaC**-nu			1PL	nit-**CaCeC**	
2MPL	hit-**CaCaC**-tem			2PL	tit-**CaCC**-u	
2FPL	hit-**CaCaC**-ten			3PL	it-**CaCC**-u	
3PL	hit-**CaCC**-u					

[28] In *binyan pi'el* and *binyan hitpa'el*, verbs in which C_2 and C_3 are identical typically conjugate with *o* as the first vowel (e.g., *p**o**tsets* 'explode PAST.3MSG', *hitp**o**tsets* 'explode oneself PAST.3MSG'). When these identical consonants are adjacent, an *e* vowel is inserted between them to separate them in pronunciation (e.g., *pots**e**tsa* 'explode PAST.3FSG'; *hitpots**e**tsa* 'explode oneself PAST.3FSG').

Appendix B

Corpus Analysis: Zero appearances, CC

CC	PoA	OCP	MoA	SonD
pp	LAB-LAB	✗	StSt	0
pb	LAB-LAB	✗	StSt	0
pm	LAB-LAB	✗	StN	2
pf	LAB-LAB	✗	StF	1
pv	LAB-LAB	✗	StF	1
pz	LAB-CORo	✓	StStr	NA
pʃ	LAB-CORo	✓	StStr	NA
pj	LAB-CORs	✓	StG	4
pg	LAB-DOR	✓	StSt	0
bp	LAB-LAB	✗	StSt	0
bb	LAB-LAB	✗	StSt	0
bf	LAB-LAB	✗	StF	1
bv	LAB-LAB	✗	StF	1
bz	LAB-CORo	✓	StStr	NA
bj	LAB-CORs	✓	StG	4
bg	LAB-DOR	✓	StSt	0
mp	LAB-LAB	✗	NSt	-2
mb	LAB-LAB	✗	NSt	-2
mm	LAB-LAB	✗	NN	0
mf	LAB-LAB	✗	NF	-1
mv	LAB-LAB	✗	NF	-1
mj	LAB-CORs	✓	NG	2
mg	LAB-DOR	✓	NSt	-2
fp	LAB-LAB	✗	FSt	-1
fb	LAB-LAB	✗	FSt	-1
ff	LAB-LAB	✗	FF	0
fv	LAB-LAB	✗	FF	0
fj	LAB-CORs	✓	FG	3
vp	LAB-LAB	✗	FSt	-1
vb	LAB-LAB	✗	FSt	-1
vm	LAB-LAB	✗	FN	1
vf	LAB-LAB	✗	FF	0
vv	LAB-LAB	✗	FF	0
vs	LAB-CORo	✓	FStr	NA
vj	LAB-CORs	✓	FG	3
tt	CORo-CORo	✗	StSt	0
td	CORo-CORo	✗	StSt	0
tz	CORo-CORo	✗	StStr	NA
tj	CORo-CORs	✓	StG	4
dt	CORo-CORo	✗	StSt	0
dd	CORo-CORo	✗	StSt	0
ds	CORo-CORo	✗	StStr	NA
dz	CORo-CORo	✗	StStr	NA
dʦ	CORo-CORo	✗	StStr	NA
dn	CORo-CORs	✓	StN	2

CC	PoA	OCP	MoA	SonD
dj	CORo-CORs	✓	StG	4
sv	CORo-LAB	✓	StrF	NA
ss	CORo-CORo	✗	StrStr	NA
sz	CORo-CORo	✗	StrStr	NA
sʦ	CORo-CORo	✗	StrStr	NA
sʃ	CORo-CORo	✗	StrStr	NA
sj	CORo-CORs	✓	StrG	NA
zp	CORo-LAB	✓	StrSt	NA
zt	CORo-CORo	✗	StrSt	NA
zd	CORo-CORo	✗	StrSt	NA
zs	CORo-CORo	✗	StrStr	NA
zz	CORo-CORo	✗	StrStr	NA
zʦ	CORo-CORo	✗	StrStr	NA
zʃ	CORo-CORo	✗	StrStr	NA
zj	CORo-CORs	✓	StrG	NA
ʦt	CORo-CORo	✗	StrSt	NA
ʦs	CORo-CORo	✗	StrStr	NA
ʦz	CORo-CORo	✗	StrStr	NA
ʦʦ	CORo-CORo	✗	StrStr	NA
ʦʃ	CORo-CORo	✗	StrStr	NA
ʦj	CORo-CORs	✓	StrG	NA
ʃs	CORo-CORo	✗	StrStr	NA
ʃʃ	CORo-CORo	✗	StrStr	NA
ʃj	CORo-CORs	✓	StrG	NA
nn	CORs-CORs	✗	NN	0
nl	CORs-CORs	✗	NL	1
nj	CORs-CORs	✗	NG	2
nʁ	CORs-GUT	✓	NL	1
lz	CORs-CORo	✓	LStr	NA
ln	CORs-CORs	✗	LN	-1
ll	CORs-CORs	✗	LL	0
lj	CORs-CORs	✗	LG	1
lʁ	CORs-GUT	✓	LL	0
jp	CORs-LAB	✓	GSt	-4
jb	CORs-LAB	✓	GSt	-4
jz	CORs-CORo	✓	GStr	NA
jj	CORs-CORs	✗	GG	0
kj	DOR-CORs	✓	StG	4
kk	DOR-DOR	✗	StSt	0
kg	DOR-DOR	✗	StSt	0
gp	DOR-LAB	✓	StSt	0
gt	DOR-CORo	✓	StSt	0
gʦ	DOR-CORo	✓	StStr	NA
gj	DOR-CORs	✓	StG	4
gk	DOR-DOR	✗	StSt	0
gg	DOR-DOR	✗	StSt	0
χj	GUT-CORs	✓	FG	3
χχ	GUT-GUT	✗	FF	0
ʁp	GUT-LAB	✓	LSt	-3
ʁl	GUT-CORs	✓	LL	0
ʁj	GUT-CORs	✓	LG	1
ʁʁ	GUT-GUT	✗	LL	0

Appendix C

Corpus Analysis: Zero appearances, CVC

CVC	PoA	OCP
pVp	LAB-LAB	✗
pVb	LAB-LAB	✗
pVm	LAB-LAB	✗
pVv	LAB-LAB	✗
bVp	LAB-LAB	✗
bVb	LAB-LAB	✗
bVf	LAB-LAB	✗
mVp	LAB-LAB	✗
mVb	LAB-LAB	✗
mVf	LAB-LAB	✗
mVv	LAB-LAB	✗
fVp	LAB-LAB	✗
fVb	LAB-LAB	✗
fVv	LAB-LAB	✗
vVp	LAB-LAB	✗
vVb	LAB-LAB	✗
vVm	LAB-LAB	✗
vVf	LAB-LAB	✗
tVd	CORo-CORo	✗
dVt	CORo-CORo	✗
dVs	CORo-CORo	✗
dVz	CORo-CORo	✗
dVʦ	CORo-CORo	✗
dVn	CORo-CORs	✓
sVz	CORo-CORo	✗
sVʦ	CORo-CORo	✗
sVʃ	CORo-CORo	✗
zVp	CORo-LAB	✓
zVt	CORo-CORo	✗
zVd	CORo-CORo	✗
zVs	CORo-CORo	✗
zVʦ	CORo-CORo	✗
zVʃ	CORo-CORo	✗
ʦVp	CORo-LAB	✓
ʦVb	CORo-LAB	✓
ʦVs	CORo-CORo	✗
ʦVz	CORo-CORo	✗
ʦVʃ	CORo-CORo	✗
nVl	CORs-CORs	✗
nVʁ	CORs-GUT	✓
lVz	CORs-CORo	✓
lVj	CORs-CORs	✗
lVʁ	CORs-GUT	✓
jVz	CORs-CORo	✓
jVj	CORs-CORs	✗

CVC	PoA	OCP
kVg	DOR-DOR	✗
gVt	DOR-CORo	✓
gVʦ	DOR-CORo	✓
gVk	DOR-DOR	✗
ʁVb	GUT-LAB	✓

Appendix D

Experiment A: Full Stimuli List

All stimuli were presented within carrier sentences of the form:

proper name + nonce verb + *et ze* 'ACC this'.

Nonce verbs are underlined, and tested consonant pairs are in bold.

C_1C_2 hi**CC**iCa hif'il, PAST.3FSG		
OCP	SCL	WF
1. nofaʁ hi**td**iχa et ze.	1. tamaʁ hi**fn**iχa et ze.	1. eden hi**ʁb**ila et ze.
2. ʃiʁa hi**td**ika et ze.	2. hagaʁ hi**vn**ika et ze.	2. jaʁden hi**ʁb**iʃa et ze.
3. miχal hi**kg**ida et ze.	3. jaaʁa hi**vn**ina et ze.	3. nomi hi**lg**iʁa et ze.
4. jael hi**kg**ina et ze.	4. ofʁa hi**vn**iχa et ze.	4. moʁan hi**ʁs**ifa et ze.
5. lital hi**kg**ima et ze.	5. liʁon hi**χn**iva et ze.	5. maajan hi**ʁz**iza et ze.

C_2C_3 Ca**CC**a qal, PAST.3FSG		
OCP	SCL	WF
1. ofʁi ba**td**a et ze.	1. ʁotem da**gf**a et ze.	1. hadaʁ ma**ʁg**a et ze.
2. ʃiʁi ma**td**a et ze.	2. naama da**gv**a et ze.	2. ʃiʁan χa**lg**a et ze.
3. zohaʁ ma**kg**a et ze.	3. keʁen ga**dv**a et ze.	3. baʁ ta**mg**a et ze.
4. niʦan ʁa**kg**a et ze.	4. miʁiam da**gχ**a et ze.	4. sivan ʁa**mg**a et ze.
5. inbaʁ ta**kg**a et ze.	5. ʃiʁ pa**χn**a et ze.	5. tamaʁa sa**mg**a et ze.

(Continued on next page)

C_1VC_2 **C**a**C**aC qal, PAST.3MSG	
OCP	WF
1. itamaʁ tadaχ et ze.	1. doʁon ʁaval et ze.
2. amiʁ tadak et ze.	2. oʁen ʁavaʃ et ze.
3. daniel kagad et ze.	3. noam lagaʁ et ze.
4. gil kagan et ze.	4. idan ʁasaf et ze.
5. ejal kagam et ze.	5. ejtan ʁazaz et ze.

C_2VC_3 Ci**C**e**C** pi'el, PAST.3MSG	
OCP	WF
1. saaʁ bited et ze.	1. tomeʁ miʁeg et ze.
2. omeʁ mited et ze.	2. ben χileg et ze.
3. ofeʁ mikeg et ze.	3. ʃaχaʁ timeg et ze.
4. alon ʁikeg et ze.	4. dan ʁimeg et ze.
5. niʁ tikeg et ze.	5. ofiʁ simeg et ze.

Appendix E

Experiment A: Stimuli Frequencies

C_1C_2 hi**CC**iCa hif'il, PAST.3FSG								
	Nonce verb	Tested pair (C_1C_2)	f	TPM	Untested pair (C_2VC_3)	f	TPM
OCP	hitdiχa	td	0	0	dVχ	21	0.09
	hitdika	td	0	0	dVk	24	0.10
	hikgida	kg	0	0	gVd	19	0.09
	hikgina	kg	0	0	gVn	19	0.09
	hikgima	kg	0	0	gVm	27	0.13
SCL	hifniχa	fn	2	0.01	nVχ	40	0.25
	hivnika	vn	0	0	nVk	22	0.14
	hivnina	vn	0	0	nVn	34	0.21
	hivniχa	vn	0	0	nVχ	40	0.25
	hiχniva	χn	18	0.05	nVv	14	0.09
WF	hiʁbila	ʁb	4	0.03	bVl	41	0.17
	hiʁbiʃa	ʁb	4	0.03	bVʃ	30	0.13
	hilgiʁa	lg	2	0.04	gVʁ	41	0.19
	hiʁsifa	ʁs	0	0	sVf	20	0.10
	hiʁziza	ʁz	0	0	zVz	24	0.18

(Continued on next page)

C2C3 **CaCCa** qal, PAST.3FSG							
	Nonce verb	Tested pair (C_2C_3)	f	TPM	Untested pair (C_1VC_2)	f	TPM
OCP	batda	td	0	0	bVt	24	0.11
	matda	td	0	0	mVt	27	0.10
	makga	kg	0	0	mVk	32	0.12
	ʁakga	kg	0	0	ʁVk	52	0.16
	takga	kg	0	0	tVk	28	0.13
SCL	dagfa	gf	4	0.04	dVg	22	0.15
	dagva	gv	3	0.03	dVg	22	0.15
	gadva	dv	3	0.04	gVd	35	0.12
	dagχa	gχ	5	0.06	dVg	22	0.15
	paχna	χn	12	0.06	pVχ	27	0.08
WF	maʁga	ʁg	9	0.04	mVʁ	32	0.12
	χalga	lg	7	0.05	χVl	66	0.11
	tamga	mg	0	0	tVm	17	0.08
	ʁamga	mg	0	0	ʁVm	25	0.08
	samga	mg	0	0	sVm	46	0.12

C1VC2 **CaCaC** qal, PAST.3MSG							
	Nonce verb	Tested pair (C_1VC_2)	f	TPM	Untested pair (C_2VC_3)	f	TPM
OCP	tadaχ	tVd	0	0	dVχ	21	0.09
	tadak	tVd	0	0	dVk	24	0.10
	kagad	kVg	0	0	gVd	19	0.09
	kagan	kVg	0	0	gVn	19	0.09
	kagam	kVg	0	0	gVm	27	0.13
WF	ʁaval	ʁVv	13	0.04	vVl	19	0.13
	ʁavaʃ	ʁVv	13	0.04	vVʃ	12	0.08
	lagaʁ	lVg	4	0.03	gVʁ	41	0.19
	ʁasaf	ʁVs	16	0.05	sVf	20	0.10
	ʁazaz	ʁVz	0	0	zVz	24	0.18

(Continued on next page)

C_2VC_3 **CiCeC** pi'el, PAST.3MSG							
	Nonce verb	Tested pair (C_2VC_3)	f	TPM	Untested pair (C_1VC_2)	f	TPM
OCP	bited	td	0	0	bVt	24	0.11
	mited	td	0	0	mVt	27	0.10
	mikeg	kg	0	0	mVk	32	0.12
	ʁikeg	kg	0	0	ʁVk	52	0.16
	tikeg	kg	0	0	tVk	28	0.13
WF	miʁeg	ʁg	18	0.03	mVʁ	32	0.12
	χileg	lg	16	0.04	χVl	66	0.11
	timeg	mg	0	0	tVm	17	0.08
	ʁimeg	mg	0	0	ʁVm	25	0.08
	simeg	mg	0	0	sVm	46	0.12

Appendix F

Experiment B: Full Stimuli List

Critical Items

tested consonant pairs are in bold.[29,30]

C_1C_2 hi**CC**iC hif'il, PAST.3MSG hi**CC**iCa hif'il, PAST.3FSG					
OCP		SCL		WF	
1. hi**td**iaχ	6. hi**td**iχa	1. hi**fn**iaχ	6. hi**fn**iχa	1. hi**ʁb**il	6. hi**ʁb**ila
2. hi**td**ik	7. hi**td**ika	2. hi**vn**ik	7. hi**vn**ika	2. hi**ʁb**iʃ	7. hi**ʁb**iʃa
3. hi**kg**id	8. hi**kg**ida	3. hi**vn**in	8. hi**vn**ina	3. hi**lg**iʁ	8. hi**lg**iʁa
4. hi**kg**in	9. hi**kg**ina	4. hi**vn**iaχ	9. hi**vn**iχa	4. hi**ʁs**if	9. hi**ʁs**ifa
5. hi**kg**im	10. hi**kg**ima	5. hi**χn**iv	10. hi**χn**iva	5. hi**ʁz**iz	10. hi**ʁz**iza

C_2C_3 Ca**CC**a qal, PAST.3FSG Ci**CC**a pi'el, PAST.3FSG					
OCP		SCL		WF	
1. ba**td**a	6. bi**td**a	1. da**gf**a	6. di**gf**a	1. ma**ʁg**a	6. mi**ʁg**a
2. ma**td**a	7. mi**td**a	2. da**gv**a	7. di**gv**a	2. χa**lg**a	7. χi**lg**a
3. ma**kg**a	8. mi**kg**a	3. ga**dv**a	8. gi**dv**a	3. ta**mg**a	8. ti**mg**a
4. ʁa**kg**a	9. ʁi**kg**a	4. da**gχ**a	9. di**gχ**a	4. ʁa**mg**a	9. ʁi**mg**a
5. ta**kg**a	10. ti**kg**a	5. pa**χn**a	10. pi**χn**a	5. sa**mg**a	10. si**mg**a

(Continued on next page)

---

[29] In Hebrew, when the final *χ* (ח) follows a vowel other than *a*, an epenthetic *a* is inserted before the final *χ* (e.g., *hitdiaχ*).

[30] In the items *ʁaval-ʁibel* and *ʁavaʃ-ʁibeʃ* (WF, C_1VC_2), the *v-b* alternation occurs due to spirantization.

C_1VC_2 **C**a**C**aC qal, PAST.3MSG **C**i**C**eC pi'el, PAST.3MSG			
OCP		WF	
1. **tad**aχ	6. **tide**aχ	1. **ʁav**al	6. **ʁib**el
2. **tad**ak	7. **tid**ek	2. **ʁav**aʃ	7. **ʁib**eʃ
3. **kag**ad	8. **kig**ed	3. **lag**aʁ	8. **lig**eʁ
4. **kag**an	9. **kig**en	4. **ʁas**af	9. **ʁis**ef
5. **kag**am	10. **kig**em	5. **ʁaz**az	10. **ʁiz**ez

C_2VC_3 **C**a**C**aC qal, PAST.3MSG **C**i**C**eC pi'el, PAST.3MSG			
OCP		WF	
1. ba**tad**	6. bi**ted**	1. ma**ʁag**	6. mi**ʁeg**
2. ma**tad**	7. mi**ted**	2. χa**lag**	7. χi**leg**
3. ma**kag**	8. mi**keg**	3. ta**mag**	8. ti**meg**
4. ʁa**kag**	9. ʁi**keg**	4. ʁa**mag**	9. ʁi**meg**
5. ta**kag**	10. ti**keg**	5. sa**mag**	10. si**meg**

(Fillers on next page)

Fillers

hiCCiC hif'il, PAST.3MSG	
hidgiʃ	'emphasized'
hifχit	'diminished'
higbiʁ	'strengthened'
higdil	'enlarged'
higdiʁ	'defined'
hikdim	'arrived early'
hikdiʃ	'dedicated'
hikpid	'insisted on'
hikʃiv	'listened'
himlits	'recommended'
himtin	'waited'
hispik	'made it on time'
hitkin	'installed'
hivdil	'distinguished'
hiχʁiz	'declared'

hiCCiCa hif'il, PAST.3FSG	
hidgiʃa	'emphasized'
hifχita	'diminished'
higbiʁa	'strengthened'
higdila	'enlarged'
higdiʁa	'defined'
hikdima	'arrived early'
hikdiʃa	'dedicated'
hikpida	'insisted on'
hikʃiva	'listened'
himlitsa	'recommended'
himtina	'waited'
hispika	'made it on time'
hitkina	'installed'
hivdila	'distinguished'
hiχʁiza	'declared'

CaCCa qal, PAST.3FSG	
tsavʁa	'gathered'
gadla	'grew up'
gazʁa	'cut'
kaʃʁa	'knotted'
paska	'decided'
patʁa	'solved'
ʁaʃma	'wrote'
safga	'absorbed'
safʁa	'counted'
taʁma	'donated'
χakʁa	'researched'
χalma	'dreamed'
χasχa	'saved'
jazma	'initiated'
zaʁka	'threw'

CiCCa pi'el, PAST.3FSG	
biʃla	'cooked'
tsilma	'filmed'
kibda	'respected'
kitsʁa	'shortened'
mitna	'moderated'
nigna	'played (music)'
pitʁa	'fired'
sidʁa	'arranged'
sikma	'summarized'
ʃilva	'combined'
simna	'marked'
ʃimʁa	'preserved'
ʃipʁa	'told'
χidʃa	'renew'
χizka	'reinforced'

CaCaC qal, PAST.3MSG	
baʁaχ	'ran away'
tsamaχ	'grew'
daʁaχ	'stepped on'
kafats	'jumped'
kalat	'perceived'
lavaʃ	'dressed'
laχats	'pressed'
maʃaχ	'pulled'
pasak	'decided'
pataʁ	'solved'
ʁaʃam	'wrote'
ʃakal	'weighed'
samaχ	'trusted'
tamaχ	'supported'
taʁam	'donated'
χakaʁ	'researched'
χalaf	'passed'
χalam	'dreamed'
χasaχ	'saved'
zaʁak	'threw'

CiCeC pi'el, PAST.3MSG	
biʃel	'cooked'
tsilem	'filmed'
tsijeʁ	'drew'
diveaχ	'reported'
kibed	'respected'
kitseʁ	'shortened'
kiven	'diverted'
miten	'moderated'
nigen	'played (music)'
pisek	'spread'
piteʁ	'fired'
sideʁ	'arranged'
sikem	'summarized'
ʃilev	'combined'
simen	'marked'
ʃimeʁ	'preserved'
ʃipeʁ	'told'
ʃivek	'marketed'
χideʃ	'renew'
χizek	'reinforced'

Practice Session

	Nonce verbs	Real verbs	
hiCCiC hif'il, PAST.3MSG	himdiaχ	hit͡sliaχ	'succeeded'
hiCCiCa hif'il, PAST.3FSG	himdiχa	hit͡sliχa	'succeeded'
CiCCa pi'el, PAST.3FSG	bisga	kibla	'got'
CaCCa qal, PAST.3FSG	taʁga	katva	'wrote'
CaCaC qal, PAST.3MSG	χagam	lakaχ	'took'

Appendix G

Experiment B: Stimuli Frequencies (critical items)

C_1C_2 hi**CC**iC hif'il, PAST.3MSG, hi**CC**iCa hif'il, PAST.3FSG							
	Nonce verb	Tested pair (C_1C_2)	f	TPM	Untested pair (C_2VC_3)	f	TPM
OCP	hitdiaχ	td	0	0	dVχ[31]	21	0.09
	hitdik	td	0	0	dVk	24	0.10
	hikgid	kg	0	0	gVd	19	0.09
	hikgin	kg	0	0	gVn	19	0.09
	hikgim	kg	0	0	gVm	27	0.13
	hitdiχa	td	0	0	dVχ	21	0.09
	hitdika	td	0	0	dVk	24	0.10
	hikgida	kg	0	0	gVd	19	0.09
	hikgina	kg	0	0	gVn	19	0.09
	hikgima	kg	0	0	gVm	27	0.13
SCL	hifniaχ	fn	2	0.01	nVχ	40	0.25
	hivnik	vn	0	0	nVk	22	0.14
	hivnin	vn	0	0	nVn	34	0.21
	hivniaχ	vn	0	0	nVχ	40	0.25
	hiχniv	χn	18	0.05	nVv	14	0.09
	hifniχa	fn	2	0.01	nVχ	40	0.25
	hivnika	vn	0	0	nVk	22	0.14
	hivnina	vn	0	0	nVn	34	0.21
	hivniχa	vn	0	0	nVχ	40	0.25
	hiχniva	χn	18	0.05	nVv	14	0.09
WF	hiʁbil	ʁb	4	0.03	bVl	41	0.17
	hiʁbiʃ	ʁb	4	0.03	bVʃ	30	0.13
	hilgiʁ	lg	2	0.04	gVʁ	41	0.19
	hiʁsif	ʁs	0	0	sVf	20	0.10
	hiʁziz	ʁz	0	0	zVz	24	0.18
	hiʁbila	ʁb	4	0.03	bVl	41	0.17
	hiʁbiʃa	ʁb	4	0.03	bVʃ	30	0.13
	hilgiʁa	lg	2	0.04	gVʁ	41	0.19
	hiʁsifa	ʁs	0	0	sVf	20	0.10
	hiʁziza	ʁz	0	0	zVz	24	0.18

[31] The frequencies of CVC are discussed despite the presence of two vowels between the consonants, as the second vowel is epenthetic and does not form part of the paradigm.

C2C3 CaCCa qal, PAST.3FSG, CiCCa pi'el, PAST.3FSG							
	Nonce verb	Tested pair (C_2C_3)	f	TPM	Untested pair (C_1VC_2)	f	TPM
OCP	batda	td	0	0	bVt	24	0.11
	matda	td	0	0	mVt	27	0.10
	makga	kg	0	0	mVk	32	0.12
	ʁakga	kg	0	0	ʁVk	52	0.16
	takga	kg	0	0	tVk	28	0.13
	bitda	td	0	0	bVt	24	0.11
	mitda	td	0	0	mVt	27	0.10
	mikga	kg	0	0	mVk	32	0.12
	ʁikga	kg	0	0	ʁVk	52	0.16
	tikga	kg	0	0	tVk	28	0.13
SCL	dagfa	gf	4	0.04	dVg	22	0.15
	dagva	gv	3	0.03	dVg	22	0.15
	gadva	dv	3	0.04	gVd	35	0.12
	dagχa	gχ	5	0.06	dVg	22	0.15
	paχna	χn	12	0.06	pVχ	27	0.08
	digfa	gf	4	0.04	dVg	22	0.15
	digva	gv	3	0.03	dVg	22	0.15
	gidva	dv	3	0.04	gVd	35	0.12
	digχa	gχ	5	0.06	dVg	22	0.15
	piχna	χn	12	0.06	pVχ	27	0.08
WF	maʁga	ʁg	9	0.04	mVʁ	32	0.12
	χalga	lg	7	0.05	χVl	66	0.11
	tamga	mg	0	0	tVm	17	0.08
	ʁamga	mg	0	0	ʁVm	25	0.08
	samga	mg	0	0	sVm	46	0.12
	miʁga	ʁg	9	0.04	mVʁ	32	0.12
	χilga	lg	7	0.05	χVl	66	0.11
	timga	mg	0	0	tVm	17	0.08
	ʁimga	mg	0	0	ʁVm	25	0.08
	simga	mg	0	0	sVm	46	0.12

(Continued on next page)

C_1VC_2 **CaCaC** qal, PAST.3MSG, **CiCeC** pi'el, PAST.3MSG							
	Nonce verb	Tested pair (C_1VC_2)	f	TPM	Untested pair (C_2VC_3)	f	TPM
OCP	tadaχ	tVd	0	0	dVχ	21	0.09
	tadak	tVd	0	0	dVk	24	0.10
	kagad	kVg	0	0	gVd	19	0.09
	kagan	kVg	0	0	gVn	19	0.09
	kagam	kVg	0	0	gVm	27	0.13
	tideaχ	tVd	0	0	dVχ	21	0.09
	tidek	tVd	0	0	dVk	24	0.10
	kiged	kVg	0	0	gVd	19	0.09
	kigen	kVg	0	0	gVn	19	0.09
	kigem	kVg	0	0	gVm	27	0.13
WF	ʁaval	ʁVv	13	0.04	vVl	19	0.13
	ʁavaʃ	ʁVv	13	0.04	vVʃ	12	0.08
	lagaʁ	lVg	4	0.03	gVʁ	41	0.19
	ʁasaf	ʁVs	16	0.05	sVf	20	0.10
	ʁazaz	ʁVz	0	0	zVz	24	0.18
	ʁibel	ʁVb	0	0	bVl	41	0.17
	ʁibeʃ	ʁVb	0	0	bVʃ	30	0.13
	ligeʁ	lVg	4	0.03	gVʁ	41	0.19
	ʁisef	ʁVs	16	0.05	sVf	20	0.10
	ʁizez	ʁVz	0	0	zVz	24	0.18

(Continued on next page)

C2VC3 **CaCaC** qal, PAST.3MSG, **CiCeC** pi'el, PAST.3MSG							
	Nonce verb	Tested pair (C_2VC_3)	f	TPM	Untested pair (C_1VC_2)	f	TPM
OCP	batad	td	0	0	bVt	24	0.11
	matad	td	0	0	mVt	27	0.10
	makag	kg	0	0	mVk	32	0.12
	ʁakag	kg	0	0	ʁVk	52	0.16
	takag	kg	0	0	tVk	28	0.13
	bited	td	0	0	bVt	24	0.11
	mited	td	0	0	mVt	27	0.10
	mikeg	kg	0	0	mVk	32	0.12
	ʁikeg	kg	0	0	ʁVk	52	0.16
	tikeg	kg	0	0	tVk	28	0.13
WF	maʁag	ʁg	18	0.03	mVʁ	32	0.12
	χalag	lg	16	0.04	χVl	66	0.11
	tamag	mg	0	0	tVm	17	0.08
	ʁamag	mg	0	0	ʁVm	25	0.08
	samag	mg	0	0	sVm	46	0.12
	miʁeg	ʁg	18	0.03	mVʁ	32	0.12
	χileg	lg	16	0.04	χVl	66	0.11
	timeg	mg	0	0	tVm	17	0.08
	ʁimeg	mg	0	0	ʁVm	25	0.08
	simeg	mg	0	0	sVm	46	0.12

Appendix H

Experiment C: Full Stimuli List

All stimuli were presented within carrier sentences of the form:

proper name(s) + nonce verb + *et ze* 'ACC this'.

Nonce verbs are underlined, and tested consonant pairs are in bold.

Stp-Fri stop-fricative, SonD = +1, SCL violation	
Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL
1. tamaʁ ʦa**gv**a et ze.	1. ohad ve nomi ʦa**gv**u et ze.[32]
2. tamaʁa ja**gv**a et ze.	2. liat ve iʦik ja**gv**u et ze.
3. agam ʦa**kf**a et ze.	3. adi ve omeʁ ʦa**kf**u et ze.
4. naama ja**kf**a et ze.	4. jael ve adam ja**kf**u et ze.

Fri-Stp fricative-stop, SonD = -1, SCL satisfaction	
Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL
1. inbaʁ ʦa**vg**a et ze.	1. miχal ve asaf ʦa**vg**u et ze.
2. alma ja**vg**a et ze.	2. ʁinat ve ofeʁ ja**vg**u et ze.
3. moʁ ʦa**fk**a et ze.	3. gali ve boaz ʦa**fk**u et ze.
4. ʃiʁa ja**fk**a et ze.	4. daniela ve baʁ ja**fk**u et ze.

(Continued on next page)

[32] *ve* 'and'

Stp-Nas stop-nasal, SonD = +2, SCL violation	
Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL
1. ʃiʁi sa**gm**a et ze.	1. ela ve amiʁ sa**gm**u et ze.
2. meʁav ja**gm**a et ze.	2. tal ve dafi ja**gm**u et ze.
3. sapiʁ za**km**a et ze.	3. ido ve miʁiam za**km**u et ze.
4. hagaʁ ja**km**a et ze.	4. anat ve baʁak ja**km**u et ze.

Nas-Stp nasal-stop, SonD = -2, SCL satisfaction	
Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL
1. libi sa**mg**a et ze.	1. ʃiʁ ve omʁi sa**mg**u et ze.
2. nofaʁ ja**mg**a et ze.	2. lital ve tomeʁ ja**mg**u et ze.
3. noga za**mk**a et ze.	3. niʦan ve jotam za**mk**u et ze.
4. iʁis ja**mk**a et ze.	4. ajala ve ʃaχaʁ ja**mk**u et ze.

OCP violation	
Ca**CC**a qal, PAST.3FSG	Ca**CC**u qal, PAST.3PL
1. jasmin ba**td**a et ze.	1. ajelet ve daniel ba**td**u et ze.
2. tali ma**td**a et ze.	2. eʁan ve hila ma**td**u et ze.
3. ʁotem ta**kg**a et ze.	3. ofʁa ve oz ta**kg**u et ze.
4. noa ma**kg**a et ze.	4. gilad ve bʁit ma**kg**u et ze.

Appendix I

Experiment C: Stimuli Frequencies

	Nonce verb	Tested pair (C_2C_3)	f	TPM	Untested pair (C_1VC_2)	f	TPM
Stp-Fri	tsagva	gv	3	0.03	tsVg	0	0
	jagva	gv	3	0.03	jVg	0	0
	tsakfa	kf	11	0.07	tsVk	0	0
	jakfa	kf	11	0.07	jVk	8	0.05
	tsagvu	gv	3	0.03	tsVg	0	0
	jagvu	gv	3	0.03	jVg	0	0
	tsakfu	kf	11	0.07	tsVk	0	0
	jakfu	kf	11	0.07	jVk	8	0.05
Fri-Stp	tsavga	vg	2	0.02	tsVv	22	0.12
	javga	vg	2	0.02	jVv	3	0.02
	tsafka	fk	5	0.07	tsVf	10	0.06
	jafka	fk	5	0.07	jVf	5	0.03
	tsavgu	vg	2	0.02	tsVv	22	0.12
	javgu	vg	2	0.02	jVv	3	0.02
	tsafku	fk	5	0.07	tsVf	10	0.06
	jafku	fk	5	0.07	jVf	5	0.03
Stp-Nas	sagma	gm	12	0.13	sVg	20	0.05
	jagma	gm	12	0.13	jVg	0	0
	zakma	km	15	0.09	zVk	21	0.18
	jakma	km	15	0.09	jVk	8	0.05
	sagmu	gm	12	0.13	sVg	20	0.05
	jagmu	gm	12	0.13	jVg	0	0
	zakmu	km	15	0.09	zVk	21	0.18
	jakmu	km	15	0.09	jVk	8	0.05
Nas-Stp	samga	mg	0	0	sVm	46	0.12
	jamga	mg	0	0	jVm	3	0.02
	zamka	mk	9	0.07	zVm	17	0.14
	jamka	mk	9	0.07	jVm	3	0.02
	samgu	mg	0	0	sVm	46	0.12
	jamgu	mg	0	0	jVm	3	0.02
	zamku	mk	9	0.07	zVm	17	0.14
	jamku	mk	9	0.07	jVm	3	0.02

	Nonce verb	Tested pair (C_2C_3)	f	TPM	Untested pair (C_1VC_2)	f	TPM
OCP	batda	td	0	0	bVt	24	0.11
	matda	td	0	0	mVt	27	0.10
	takga	kg	0	0	tVk	28	0.13
	makga	kg	0	0	mVk	32	0.12
	batdu	td	0	0	bVt	24	0.11
	matdu	td	0	0	mVt	27	0.10
	takgu	kg	0	0	tVk	28	0.13
	makgu	kg	0	0	mVk	32	0.12

תקציר

אילוצים פונוטקטיים ממלאים תפקיד מרכזי בהגדרת רצפי העיצורים האפשריים בשפה ובעיצוב העדפותיהם של הדוברים בנוגע לרצפי עיצורים שונים. אף שהשפעתם משתנה משפה לשפה, נטען כי מכלול האילוצים הוא אוניברסלי ומשותף לכל המערכות הפונולוגיות באשר הן (ראו, למשל, Prince and Smolensky 1993/2004, 1997; Chomsky 2006; Smolensky and Legendre 2006). אולם, בנוסף לאילוצים האוניברסליים, הידע הלשוני מתעצב בהתאם למידע שמקורו בשפה המדוברת בסביבה, כאשר השכיחות ממלאת תפקיד משמעותי בכך.

המחקר הנוכחי בוחן כיצד אילוצים אוניברסליים ושכיחויות בלקסיקון מעצבים הגבלות במערכת הפונולוגית של הדוברים. המחקר עוסק בעברית מודרנית, ומתמקד בשלושה אילוצים (או עקרונות) המוכרים היטב בספרות הבלשנית: עקרון ה-OCP (*Obligatory Contour Principle*; Leben 1973; Goldsmith 1976; McCarthy 1979, 1981, 1986), עקרון ה־SCL (*Syllable Contact Law*; Murray and Vennemann 1983, Vennemann 1988, Clements 1990) ועקרון ה־SDP (*Sonority Dispersion Principle*; Clements 1990).

המחקר משלב שתי גישות מתודולוגיות – ניתוח קורפוס וניסויים פסיכו־פונולוגיים. ניתוח הקורפוס בוחן את השפעת האילוצים האוניברסליים על הלקסיקון, ואילו הניסויים עוסקים בתפקידם של האילוצים בעיצוב המערכת הפונולוגית של הדוברים. שילוב השיטות מאפשר לבחון אם המערכת הפונולוגית מושפעת מהלקסיקון בלבד או גם מאילוצים אוניברסליים שאינם משתקפים בו.

עבודת הדוקטורט כוללת חמישה פרקים. פרק 1 מהווה הקדמה ומציג את שאלת המחקר ומטרותיו. פרק 2 סוקר את הרקע התיאורטי, תוך התמקדות בגישות לידע לשוני, באילוצים הפונוטקטיים הנדונים במחקר (OCP, SCL, SDP) ובמאפיינים של מצאי העיצורים בעברית מודרנית.

פרק 3 מציג את ניתוח הקורפוס. הקורפוס שנותח במחקר הוא *heTenTen*, שמבוסס על אוסף טקסטים נרחב בעברית מודרנית ומייצג את הלקסיקון של השפה (Jakubíček et al. 2013). הניתוח התמקד בפעלים שבגזע שלהם יש שלושה עיצורים ובחן את דפוסי המופעים של עיצורים סמוכים (C_1C_2, C_1VC_2) ושל עיצורים שאינם סמוכים (C_1VC_2, C_2VC_3). התוצאות חושפות הבדלים ברורים בין האילוצים: הלקסיקון מכבד את ה־OCP, כלומר נמנע באופן נרחב מהפרת האילוץ, ואילו ההשפעה של SCL ושל SDP אינה ניכרת בלקסיקון, והם מופרים לעיתים קרובות.

פרק 4 מציג את הניסויים הפסיכו־פונולוגיים. ניסויים A ו־B בחנו את האילוצים OCP ו־SCL בשתי שיטות שונות: ניסוי A בחן את ההשפעה של הפרות האילוצים האלה על שיפוטים לקסיקליים לפעלי תפל, ואילו ניסוי B בחן את אותן הפרות באמצעות מדידת דיוק וזמני תגובה בזיהוי פעלי תפל. תוצאות הניסויים מצביעות על השפעה חד־משמעית של ה־OCP על המערכת הפונולוגית של הדוברים, בעוד שה־SCL מציג דפוס מורכב, התלוי בפערי הצליליות (sonority) שבין העיצורים הסמוכים.

ניסוי C התמקד בהפרות של אילוץ ה־SCL, כהמשך לתוצאות של שני הניסויים הקודמים. באמצעות בחינה של שיפוטים לקסיקליים לפעלי תפל, הניסוי השווה בין כיבוד והפרה של ה־SCL בשתי רמות צליליות: בזוגות של שני חוסמים (obstruents), שהפער בצליליות ביניהם הוא 1, ובזוגות של חוסם וצלילי (sonorant), שהפער בצליליות ביניהם הוא 2. התוצאות מראות כי בזוגות של חוסם וצלילי, הנבדקים העדיפו באופן מובהק את הרצפים שהפרו את ה־SCL, בהתאם למגמה בלקסיקון. עם זאת, בזוגות של שני חוסמים, הנבדקים העדיפו באופן מובהק את הרצפים המכבדים את ה־SCL, אף שבלקסיקון שני חוסמים נוטים להפר את האילוץ.

ממצא זה מצביע על כך של-SCL יש השפעה רבה יותר בזוגות של שני חוסמים, תוך שהוא גובר על המגמה בלקסיקון. בכך האילוץ משקף תופעה הנקראת *The Emergence of the Unmarked* (McCarthy and Prince 1994), כלומר "הופעת הלא מסומנן", שבה ניתנת עדיפות לצורות שאינן מסומננות מבחינה אוניברסלית, על אף שכיחותן הנמוכה בלקסיקון.

בנוסף, הפערים בצליליות בין פעלי התפל בניסוי C אפשרו לבחון את אילוץ ה־SDP. התוצאות מראות כי הנבדקים העדיפו פערי צליליות גדולים יותר בין עיצורים סמוכים, וזאת בהתאם ל־SDP, אף שהלקסיקון אינו מראה העדפה ברורה שכזו.

לסיום, פרק 5 מסכם את ממצאי המחקר ומדגיש את האינטראקציה המורכבת בין אילוצים אוניברסליים, שכיחות בלקסיקון והמערכת הפונולוגית של הדוברים. המחקר מראה כי הלקסיקון והאילוצים הפועלים בו משפיעים על המערכת הפונולוגית. יחד עם זאת, גם אילוצים שהשפעתם אינה ניכרת בלקסיקון משפיעים על המערכת הפונולוגית, דבר המצביע על תפקידם הייחודי של אילוצים אוניברסליים בעיצוב מערכת זו.

TEL AVIV UNIVERSITY אוניברסיטת תל אביב

הפקולטה למדעי הרוח ע״ש לסטר וסאלי אנטין

בית הספר לפילוסופיה, בלשנות ולימודי מדע

פונוטקטיקה עברית:
שכיחות ודקדוק

חיבור לשם קבלת התואר

״דוקטור לפילוסופיה״

מאת

הדס יברכיהו

העבודה הוכנה בהנחיית

פרופ׳ אותי בת־אל פוקס

דצמבר 2024