COMPARISON OF A MORPHOLOGICAL LEARNER TO CHILD ACQUISITION OF THE TURKISH AORIST

KAAN BAYAR

BOGAZICI UNIVERSITY

2025

COMPARISON OF A MORPHOLOGICAL LEARNER TO CHILD ACQUISITION OF THE TURKISH AORIST

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Linguistics

by

Kaan Bayar

Boğaziçi University

2025

Comparison of a Morphological Learner

to Child Acquisition of the Turkish Aorist

The thesis of Kaan Bayar

has been approved by:

Assist. Prof. Ömer Faruk Demirok _____
(Thesis Advisor)

Assist. Prof. Ümit Atlamaz _____
(Thesis Co-Advisor)

Prof. Balkız Başaran _____

Assoc. Prof. Kadir Gökgöz _____

Assist. Prof. Junko Kanero _____
(External Member)

June 2025

DECLARATION OF ORIGINALITY

I, Kaan Bayar, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature…….....

Date …….....

ABSTRACT

Comparison of a Morphological Learner

to Child Acquisition of the Turkish Aorist

In this study, I apply a supervised morphological rule learner model, adapted from Albright and Hayes (2002), to test and improve the model's performance on the Turkish aorist by focusing on solving the problems posed by vowel harmony and irregular morphological patterns. The model generates rules based on word pairs differing in a single morphological feature. The research aims to evaluate the model's effectiveness by comparing its predictions with child language acquisition data from Nakipoğlu and Ketrez (2006). The study results show that the model captures the general patterns of vowel harmony and performs well on multisyllabic forms but struggles with monosyllabic verbs, especially those ending in sonorants. Additionally, I explore how syllable information affects the learnability of the aorist marker in Turkish and experiment with additional parameters. Overall, the study shows the limitations of rule-based approaches in modeling Turkish morphology and how syllable information helps in the learnability of the aorist marker. This study points out several avenues for future improvement, such as implementing corrective feedback, that could potentially change the results. The study hopefully has contributed to our understanding of the parallels between computational models and human language acquisition.

ÖZET

Türkçe Geniş Zaman Ekinin Öğrenimi: Çocuk Edinimi ile Morfolojik Öğrenme Modelinin Karşılaştırılması

Bu çalışma gözetimli bir biçimbilimsel kural öğrenme modeline Türkçe'deki geniş zaman ekini öğrenmesini incelemeyi ve sonuçları iyileştirecek parametreleri keşfetmeyi amaçlamaktadır. Bu çalışmada, Albright ve Hayes (2002)'den uyarlanmış gözetimli bir biçimbilimsel kural öğrenme modeli uygulanmıştır. Bu model, tek bir biçimbilimsel özelliği farklı olan sözcük çiftlerine dayanarak kurallar üretmektedir. Bu araştırmada, Nakipoğlu ve Ketrez (2006)'nın çocuk dil edinimi verilerinin, modele ait öngörülerle karşılaştırılması sonucunda modelin etkililiğini değerlendirilmesi amaçlanmaktadır. Bulgular, modelin ünlü uyumunun genel örüntülerini yakaladığını ve çok heceli biçimlerde iyi performans gösterdiğini, ancak tek heceli fiillerde -özellikle titreşimli ünsüzle bitenlerde- zorlandığını göstermiştir. Ayrıca bu araştırma, hece bilgisinin Türkçede geniş zaman ekinin öğrenilebilirliğini nasıl etkilediği incelenmiş ve modele eklenen ekstra parametrelerle de deneyler gerçekleştirilmiştir. Genel olarak, çalışma kural tabanlı yaklaşımların Türkçe biçimbilimi modellemelerindeki sınırlılığını ortaya koymakla kalmamış. Aynı zamanda hece bilgisinin, geniş zaman ekinin öğrenilebilirliğine nasıl katkı sağladığını da göstermiştir. Bu çalışma, düzeltici geri bildirim uygulanması gibi sonuçları potansiyel olarak değiştirebilecek birkaç iyileştirmeler önermektedir. Bu çalışmanın, hesaplamalı modeller ile insan dil edinimi arasındaki paralelliklerin daha iyi anlaşılmasına katkıda bulunmuştur.

ACKNOWLEDGMENTS

I want to start by expressing my most profound appreciation to Dr. Ömer Demirok and Dr. Ümit Atlamaz for their exceptional mentorship and guidance. Their insightful feedback and thoughtful critiques have been instrumental in shaping my research's direction and quality. Their high standards continually pushed me to improve. I am especially grateful for the many hours they devoted to reviewing drafts, suggesting perspectives, and helping me refine my arguments. Their support extended beyond my thesis; they have been advocates for my academic development. Above all, they have shown me what it means to be both rigorous scholars and attentive advisors. Their dedication to excellence and the success of their students has left a lasting impact on my academic journey. I feel fortunate to have benefited from their expertise, kindness, and steadfast encouragement throughout my master's degree.

I would also like to extend my sincere thanks to Dr. Junko Kanero, Dr. Kadir Gökgöz, and Dr. Balkız Öztürk Başaran for serving on my thesis committee. Their valuable feedback, thoughtful questions, and constructive suggestions have significantly contributed to the clarity and strength of this research. I am grateful for the time and care they dedicated to reviewing my work and for the perspectives they brought to this research.

I would like to express my sincere gratitude to the institutions that supported me financially during my studies. I am especially thankful to Dr. Olcay Taner Yıldız for the opportunity to work as a research intern on his TÜBİTAK-funded project (Grant No. 123E027), which enriched my skills in dataset preparation. I also

gratefully thank TÜBİTAK for their support through the 2210-A National Scholarship Programme, which played an important role in my academic journey.

Thanks to the unwavering support of Dr. Atlamaz and Dr. Demirok, I had the opportunity to intern at the Max Planck Institute for Psycholinguistics, where I met incredible people who made my experience both academically enriching and personally fulfilling. I am especially grateful to Dr. Candice Francis and Dr. Izabela Jordanoska, whose support made me feel at home and helped me grow as a researcher. I also owe special thanks to Dr. Anna Mai for their generous guidance, particularly during my PhD applications, and their attention to detail. I would like to thank my other supervisor, Dr. Filiz Tezcan, for taking the time to share her expertise and helping me deepen my understanding of the field. To Noémie, Diane, Julia, and Edoardo, thank you for the laughter, memorable conversations, and warm company that made everyday life in the Netherlands so enjoyable. Lastly, I am thankful to Dr. Andrea Martin and all other LACNS members for their valuable feedback and support throughout my time at the institute.

I feel very fortunate to have called this department my home for nearly seven years. During this time, I have had the privilege to learn and grow alongside remarkable people. I am deeply thankful to the faculty and staff for their constant support and for creating a welcoming and stimulating environment. To both of my cohorts, thank you for making this journey unforgettable. Special thanks to Onur, Metehan, Nagihan, Dilara, and everyone I couldn't name for the late-night answers, pep talks, and listening to my academic rants. Your friendship has kept me grounded and (mostly) sane. Thank you to everyone who made my bachelor's and master's studies so meaningful. I truly appreciate the support, inspiration, and kindness I have received from my professors and peers.

I would like to sincerely thank several people who have helped me immensely throughout my academic journey. I am grateful to Rüveyda. Their keen eye for detail and curiosity has left a lasting impression on my academic journey. I thank Sinem for their constant emotional support during difficult times, always offering comfort, understanding, and encouragement when I needed it most. They also helped me step out of my comfort zone, which greatly contributed to my personal and academic growth. I feel lucky to call them my friend. I am thankful to Aysemin, whom I met during my master's. We've shared similar ambitions, which made our academic journey feel less isolating and more meaningful. During one of the most intense and challenging stages of my studies, she supported me by patiently listening to my vents, offering thoughtful advice, and showing kindness and clarity that had a profound impact on me. I truly appreciate all that she has done, and I feel incredibly lucky that our paths crossed. I want to thank Serkan for editing my cover letters and for his reality checks. His honest feedbacks helped me grow academically. Finally, I want to thank Deniz for always believing in me and never ceasing her invaluable support. She has been one of my fiercest critics and strongest advocates. I am incredibly grateful to have had her by my side throughout this journey, and I feel truly fortunate to call her my best friend. I am deeply thankful to all my friends, including those I couldn't name, for their support and expertise. Without them, I would not be where I am today.

Lastly, I want to express my deepest gratitude to my family for always being there when I needed them. Their support, encouragement, and unconditional love have been the foundation of my journey. They have stood by me through every challenge, celebrated my achievements, and reminded me to keep pushing forward when things seemed overwhelming. Without their belief and constant presence, I

would not have reached this point. They have been my strongest support pillars, and I am endlessly grateful for their understanding and patience. I could not have done it without them. Thank you for being my greatest source of strength and for believing in me even when I doubt myself. Your love and encouragement have made all the difference.

Dedicated to my family for forever believing in me

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

3	Third Person
AOR	Aorist Tense Marker (Habitual Aspect Marker)
PST	Past Tense Marker
ACC	Accusative Case Marker
PL	Plural Morpheme
SG	Singular
MGL	Minimal Generalization Learner
FUT	Future Tense Marker
PROG	Progressive Aspect Marker
SON	Sonorant Feature
CONT	Continuant Feature
NAS	Nasal Feature
VOI	Voiced Feature
ANT	Anterior Feature
COR	Coronal Feature
BACK	Backness Feature
ROUND	Roundness Feature
HIGH	Vowel Height Feature
LOW	Vowel Height Feature
VOWEL	Vowel Feature

CHAPTER 1

INTRODUCTION

There have been many attempts at creating models for learning morphology. Approaches to morphological learners vary broadly, including connectionist models (Rumelhart et al., 1986; Pinker & Prince, 1988; Kirov & Cotterell, 2018), decision tree models (Belth et al., 2021), single and dual route models (Nakisa et al., 2000), and non-connectionist models (Albright & Hayes, 2002). These models aim to learn rules for form-feature matching, and the past tense in English has become a canonical test case for having a clear spectrum of different morphological patterns. These patterns include rule-governed forms (walk→walk+ed), irregular morphology (dive→dove, sing→sang), and fully suppletive morphology (go→went).

Even though the past tense in English provides a robust testing ground for these learners, Turkish has some features that challenge these models on a different set of morphophonological test cases. Turkish has a vowel harmony system, resulting in morphemes splitting into two categories: type I and type A. These types govern which features are underspecified in the vowel(s) of the morpheme. In addition to the vowel harmony, lexically conditioned allomorphy further increases the complexity of the morphology of Turkish and presents difficulties for the learner. For example, when inflected for the same category, al- 'take' becomes alır 'takes', but çal- 'steal' becomes çalar 'steals', which defies any explanation in morphophonological terms. These features present additional challenges for models and provide an excellent testing ground for the interaction between rule-governed and irregular forms. The so-called aorist marker (translated as simple present into English) provides an amazing test case for these models. As will be discussed in detail, it is not at all trivial for the

learner (or for the linguist, for the matter) to posit default/elsewhere morphophonological rules that could explain the distribution of the aorist marker in Turkish. Contrary to the rest of the morphemes in Turkish, the aorist has both morpheme types, and its distribution changes based on whether the root is multisyllabic or monosyllabic. The experiments conducted in both toddlers (Nakipoğlu & Ketrez, 2006) and adults (Nakipoğlu & Michon, 2020) indicate that learners have problems in constructing clear-cut morphophonological rules for the aorist marker. This problem surfaces in word generation errors, especially by toddlers. In summary, the aorist marker shows variability that cannot be easily accounted for by regular phonological processes alone, and this variability is particularly evident in monosyllabic verbs, where the choice between the A-type (-Ar) and I-type (-Ir) aorist markers are most often exceptional.

This research extends Albright and Hayes (2002)'s proposed supervised morphological rule learner model that generates rules based on pairs of words differing in one syntactico-semantic feature by including additional parameters to capture the partially irregular morphology present in the aorist marker in Turkish. While this bottom-up model has effectively captured morphophonological (e.g., V+ed forms) and analogical rules (e.g., sing~sang) for both regular and irregular forms of the English past tense, its extension to Turkish is challenged by the autosegmental nature of the vowel harmony and seemingly random irregular forms. This research aims to implement the Albright and Hayes (2002) model and check its performance on the Turkish aorist marker by exploring specific model parameters to pinpoint those that are most crucial to the model more precisely. Crucially, by comparing the model's output with child language acquisition data (Nakipoğlu & Ketrez, 2006), this study seeks to identify the behavior of the model by checking

whether its generalizations are similar to those of children and how crucial the parameters used for the morphological learner are. The findings are expected to contribute to a better understanding of how rule learning mechanisms interact with complex phonological and morphological constraints and explore the models' similarities to the acquisition data (Nakipoğlu & Ketrez, 2006).

The following section presents the vowel harmony system of Turkish, an overview of the morphological structure of Turkish and the aorist marker, followed by an overview of the child acquisition dataset. Chapter 2 introduces the dataset and provides a detailed explanation of the model's implementation, including the explanation of the word generation algorithm used to test the model's rules. Chapter 3 reports the results of the models and discusses their results in the context of morphological learning, as well as introducing additional parameters to solve the broad context problem found in the implementation of syllable information. Broad context problem occurs when the context of the learned rules is too broad for the word generation model to predict the correct form. Chapter 4 compares the models' errors on monosyllabic roots with sonorant final segments to the acquisition data (Nakipoğlu & Ketrez, 2006) and outlines directions for future research, and Chapter 5 concludes the thesis by providing a summary.

1.1 Vowel harmony in Turkish and the aorist marker in Turkish

Turkish is a Turkic language spoken in Turkey. It is an agglutinative, head-final language with eight vowels /a e i ɯ u y o ø/. It also has vowel harmony. A chart of the vowels in Turkish is shown in Table 1 below:

Table 1. Turkish Vowel Features

Vowels	Back	Round	High	Low
a	+	-	-	+
e	-	-	-	-
ɯ	+	-	+	-
i	-	-	+	-
u	+	+	+	-
y	-	+	+	-
o	+	+	-	-
ø	-	+	-	-

Note: The table is taken from (Odden, 2013, pp. 148-149)

1.1.1 Vowel harmony in Turkish

Vowel harmony is a regular process in Turkish[1] (Kabak, 2011). It affects two contrastive features: {±round, ±back}. In principle, these features are underspecified for the vowels of the morphemes[2], and the values for these features spread rightwards from the stem-final syllable to all underspecified vowels[3]. However, non-high morphemes, A-type morphemes, only accept [±back] harmony.

(1) /-lAr/ is the plural marker in Turkish with the [±back] feature unspecified

a. pil-ler

battery-PL *'batteries'*

b. kaydırak-lar

slide-PL *'slides'*

[1] It is debated as to what extent the harmony applies within roots. I refer the reader to (Clements & Sezer, 1982; Kabak, 2011; Pöchtrager, 2011) for harmony in roots.

[2] There are fully specified vowel occurrences in some morphemes. For example, the progressive aspect (-Iyor) has one fully specified vowel (i.e., its second vowel).

[3] There is also a problem with roots ending with a palatalized sound. For example, [saat] 'clock' becomes [saat-i] 'clock.ACC'. For more in-depth discussion on vowel harmony on palatalized segments, you can check (Canalis & Dikmen, 2021).

The example in (1) shows the spreading of the feature [±back] to the plural suffix. However, both [±back] and [±round] features are harmonized when conjugating with I-type morphemes.

(2) /-(y)I/ is the accusative form with both [±back] and [±round] features unspecified

a. ay-ı	b. su-yu
moon-ACC	water-ACC
c. çöl-ü	d. sirke-yi
desert-ACC	vinegar-ACC

The examples in (2) show the spreading of both back and round features. Turkish vowel harmony occurs on two features: [±back] and [±round].

1.1.2 Morphological structure of Turkish and the aorist marker

Vowel harmony has an effect on the morphology of Turkish, namely, on the realization of the morphemes as concrete pronounceable forms. The vast majority of the morphemes in Turkish can be split into two types: I-type and A-type[2]. These types differ in which vowel features they are (under)specified for in the morpheme. I-type includes both features, whereas A-type only includes the [±back] feature.

(3) /-DI / is the past tense marker in Turkish and is an I-type morpheme[4]

a. gel-di	‘came’	c. sat-tı	‘sold’
come.PST		sell.PST	
b. çürü-dü	‘decayed’	d. oku-du	‘read’
decay.PST		read.PST	

[4] The first consonant of the morpheme’s voice feature spreads from the final sound of the stem.

(4) /-lAr/ is the plural marker in Turkish and is an A-type morpheme[5]

a. pil-ler ‘batteries’ b. okul-lar ‘schools’

battery.PL school.PL

Examples (3) and (4) show I-type and A-type morphemes. Every morphological marker in Turkish belongs to either I-type or A-type, with one exception being the aorist/habitual marker in Turkish. The aorist marker has both I-type and A-type forms.

(5) Aorist marker in Turkish

	I-type realization		*A-type realization*	
a.	al-ır	‘takes’	c. çal-ar	‘steals’
	take.AOR		steal.AOR	
b.	öl-ür	‘dies’	d. böl-er	‘divides’
	die.AOR		divide.AOR	

Example (5) shows that the aorist marker has both I-type and A-type forms, whereas others only have one type. More importantly, the stems in (5) are almost identical (and fully identical in the last two VC segments). This presents a problem in both language acquisition and predicting which form of the morpheme the verb will surface with.

The distribution of the aorist marker in Turkish is governed by two main parameters: syllable count and root type. These parameters create three groups: -r selection is based on the root type (vowel ending roots), -Ir selection based on vowel harmony and syllable count (Default/elsewhere form for multisyllabic roots), -Ar

[5] The final segment in this morpheme has its back feature underspecified, and uses the value spread from the stem.

selection based on vowel harmony and syllable count (Default/elsewhere form for monosyllabic roots). Nakipoğlu and Michon (2020) state that multisyllabic verbs almost always surface with the I-type form; however, monosyllabic verbs surface with both forms. Though the majority of the monosyllabic verbs surface with the A-type affix, there are examples where the monosyllabic verb can surface with the I-type morpheme.

(6) Monosyllabic roots surfacing with the aorist marker (Data from (Nakipoğlu & Michon, 2020, p. 5))

bak-	‘look’	bak-ar	‘looks’
seç-	‘choose’	seç-er	‘chooses’
al-	‘take’	al-ır	‘takes’
bil-	‘know’	bil-ir	‘knows’

Nakipoğlu and Michon (2020) state that there are 13 monosyllabic verbs that take the I-type form. Those irregularly behaving verbs have sonorant endings, namely /l, r, n/; however, this still does not fully encapsulate this irregularity. Nakipoğlu and Michon (2020) claim there are also verbs that end in sonorants but take A-type.

(7) Monosyllabic verbs ending in sonorants (Data from (Nakipoğlu & Michon, 2020, p. 6))

al-	‘take’	al-ır	‘takes’
çal-	‘steal’	çal-ar	‘steals’
öl-	‘die’	öl-ür	‘dies’
böl-	‘divide’	böl-er	‘divides’

This poses a challenge for the acquisition of these forms, particularly in rule formation for this category. Table 2 shows the distribution of monosyllabic roots with the aorist marker (Nakipoğlu & Michon, 2020, p. 6).

Table 2. Distribution of Monosyllabic Roots

	-Ar		-Ir		Total	
Monosyllabic roots	Type	Token	Type	Token	Type	Token
Sonorant	47	131806	13	791717	60	923523
Non-sonorant	169	786621	0	0	169	786621
Total	216	918427	13	791717	229	1710144

Note: Table is taken from (Nakipoğlu & Michon, 2020, p. 6)

It can be seen from the corpus analysis that the majority of the monosyllabic verbs surface with -Ar. It seems that -Ar could be considered the default/elsewhere exponent of the morpheme, and -Ir as its irregular realization. As shown in example (7), there seems to be a lexically conditioned allomorphy between these two forms, as there is no concrete difference between these forms. Table 3 below shows the distributions of the multisyllabic verbs:

Table 3. Distributions of Multisyllabic Roots

	-Ar		-Ir		Total	
Multisyllabic roots	Type	Token	Type	Token	Type	Token
Sonorant	1	4	7976	2121451	7977	2121455
Non-sonorant	88	42953	1393	427607	1481	470560
Total	89	42957	9369	2549058	9458	2592015

Note: Table is taken from (Nakipoğlu & Michon, 2020, p. 8)

Nakipoğlu and Michon (2020) report that occurrences of -Ar stem from the light verb *et-* (which is, of course, monosyllabic) that surfaces with -Ar form. For example, *aff+et-* 'forgive' becomes *aff+eder* 'forgives'[6]. As seen in the table above, the majority of the multisyllabic verbs surface with the -Ir form. In contrast to the monosyllabic verbs, which predominantly take -Ar (169 -Ar forms out of 229 occurrences), multisyllabic verbs seem to be almost exclusively using -Ir.

Table 4 below shows the overall distributions of -Ar and -Ir forms. The lack of vowel ending forms (-r ending forms) is because, contrary to the general convention, Nakipoğlu and Michon (2020) consider these vowel ending verbs to be combining with a vowel-initial -Ir or -Ar form, depending on the last vowel of the root. Meaning, if the last vowel is -u, as in oku-r 'reads', they treat this as an occurrence of -Ir. They claim the reason for this distinction is due to empirical results. They argue that erroneous constructions such as *ok-ar (oku-r) 'reads' and *uy-ar (uyu-r) 'sleeps' suggest the children might be treating the last occurrence of the verb as part of the morpheme as well.

Table 4. Total Counts of Both Types of the Aorist Marker

	-Ar		-Ir		Total	
Mono/multi/vowel-ending roots[7]	Type	Token	Type	Token	Type	Token
Total	998	1537983	9433	3399037	10530	4937020

Note: Table is taken from (Nakipoğlu & Michon, 2020, p. 7)

[6] Voiceless consonant becomes voiced if it is stem final and the morpheme attached starts with a voiced sound.

[7] As explained above, the reason vowel ending roots are also considered -Ar or -Ir is that they treat vowel ending as either -Ar or -Ir based on what type of vowel it ends with. For example oku-r 'read' is counted towards -Ir type.

As can be seen from Table 4, the majority of the roots surface with the -Ir form; however, as seen in Table 2 and partly in Table 3, where the distributions of the monosyllabic and multisyllabic verbs for the aorist marker are shown, there is a root-conditioned allomorphy in the distribution of these forms. This presents a problem for language acquisition with regard to the production of types with the root-conditioned allomorphy.

Overall, -Ar could be considered as the default form for monosyllabic verbs, and -Ir can be considered as the default form for multisyllabic verbs. Additionally, there is allomorphy that is lexically governed (i.e., unpredictable root-dependent choice of exponent). Regulars residing along irregulars, and in particular, the difficulty in choosing what is regular and what is irregular, adds to the complexity arising in inferring rules for this marker, which is also shown in Nakipoğlu and Ketrez (2006). The effect of this distribution problem on language acquisition will be explored in the next section.

1.2 Overview of the child acquisition dataset

As explained in the previous sub-section, the aorist marker has two forms, which results in children having to make predictions based on generalizations from their past experiences during early language acquisition. Where form selection is predictably conditioned, contextual rules are able to successfully generate surface forms. However, these rules fail when forms are irregular, resulting in overregularization when the regular rule is extended to these irregular verbs. Nakipoğlu and Ketrez (2006) consider a form to be an overregularization error when a regular rule is extended to an irregular form. For example, using ‘go+ed’ instead of ‘went’ when trying to form the past tense for ‘go’. Nakipoğlu and Ketrez (2006)

consider a form to be an irregularization error when a regular form is treated as an exception and used with the irregular rule. For example, using -Ar morpheme with a multisyllabic root is considered to an irregularization error.

Nakipoğlu and Ketrez (2006) conducted a picture description experiment with 135 participants divided across five groups. The ages of the participants ranged from 2;9 to 8;2 (years;months) old, and the control group consisted of five adult native Turkish speakers who were undergraduate students at Boğaziçi University in Istanbul. Table 5 below shows the participants of the experiment:

Table 5. Experiment Groups

Group Number	Age Range (year;months)	Mean Age
G1 (n =24)	2;9-3;11	3;5
G2 (n =36)	4;1-5;1	4;5
G3 (n =32)	5;1-6;1	5;5
G4 (n =30)	6;6-7;4	7;0
G5 (n =13)	7;6-8;2	7;9
Adults (n =5)	~20	

Note: Table is taken from (Nakipoğlu & Ketrez, 2006, p. 403)

During the task, participants were asked to describe a picture that depicted a human or an animal engaging in an action. In order to describe the picture, the participants were asked to use a verb from a list of given verbs. There were 44 verbs across five categories. The verbs are shown below:

(8) Verb categories (Nakipoğlu & Ketrez, 2006, p. 403)

a. -Ir taking verbs ending in sonorants (n=10)

- al '*take'*, bul '*find'*, dur '*stop'*, gel '*come'*, gör '*see'*, kal '*stay'*, ol '*be'*, öl '*die'*, ver '*give'*, vur '*hit'*

b. -Ar taking verbs ending in sonorants (n=10)

- bin '*get on'*, çal '*play an instrument'*, dal '*dive'*, don '*freeze'*, gir '*enter'*, gül '*laugh'*, kır '*break'*, sol '*fade'*, sür '*drive'*, yan '*burn'*

c. -Ar taking verbs ending in non-sonorants (n=9)

- at '*throw'*, bak '*look',* git *'go'*, kalk '*stand up'*, tut '*hold'*, çık '*exit'*, yap '*do'*, yat '*lie down'*, yüz '*swim'*

d. Multisyllabic verbs (n=7)

- getir '*bring',* göster '*show'*, kaçır '*miss'*, otur '*sit'*, sallan '*swing'*, süpür '*sweep'*, tırman '*climb'*

e. Vowel ending verbs

i. High vowel ending verbs (n=3)

oku '*read'*, uyu '*sleep'*, yürü '*walk'*

ii. Non-high vowel ending verbs (n=5)

atla '*jump'*, de '*say'*, izle '*watch'*, oyna '*play'*, ye '*eat'*

Nakipoğlu and Ketrez (2006) argue that their findings suggest children exhibit non-adult-like performance in verb morphology until approximately the age of 7 years and 9 months. A closer examination of the data reveals that the majority of the children's errors occur in monosyllabic verbs, suggesting that syllable count plays a role in the difficulty of morphological processing. Furthermore, among these monosyllabic verbs, those ending in sonorant consonants (such as [l], [r], [m], or [n]) account for the largest proportion of the observed errors. This pattern suggests that both phonological and morphological complexity contribute to children's difficulties. Table 6 below illustrates the distribution of error rates across different verb types.

Table 6. Combined Error Rates

Error rate (%)	G1	G2	G3	G4	G5	Total
Multisyllabic	1/154 (%0.6)	21/225 (9%)	14/172 (8%)	1/208 (0.4)	0/104 (0%)	37/863 (4%)
Monosyllabic	131/494 (26%)	185/762 (24%)	128/823 (15%)	100/785 (8%)	21/354 (6%)	565/3218 (18%)
Non-sonorant ending monosyllabic	6/129 (5%)	21/168 (12%)	7/243 (3%)	2/228 (1%)	2/106 (2%)	38/874 (4%)
Sonorant ending monosyllabic	124/364 (34%)	164/594 (28%)	121/580 (21%)	98/557 (18%)	19/248 (8%)	536/2344 (26%)

Note: The table is taken from (Nakipoğlu & Ketrez, 2006, p. 405).

This problem is exemplified in (7), where nearly identical verb roots take different aorist suffixes despite their phonological similarity. For instance, the verb öl- (‘to die’) takes the aorist suffix -ür, yielding öl-ür, while böl- (‘to divide’) takes the aorist suffix -er, resulting in böl-er. The absence of clear phonological or morphological rules makes it difficult for children to form reliable generalizations. Consequently, they are left to rely on rote memorization or make incorrect analogies, which contributes to the high error rates observed in this verb class. Table 7 provides a breakdown of participants’ regularization error rates specifically for monosyllabic verbs with sonorant endings.

Table 7. Overregularization and Irregularization Error Rates for Monosyllabic Verbs with Sonorant Endings

Error rate (%)	G1	G2	G3	G4	G5
Ir > *Ar (Overreg)	75/194 (39%)	58/312 (18%)	42/296 (13%)	22/283 (7%)	5/126 (3%)
Ar > *Ir (Irreg)	50/171 (29%)	106/292 (38%)	79/284 (28%)	76/274 (27%)	14/122 (11%)

Note: Table is taken from (Nakipoğlu & Ketrez, 2006, p. 406).

As shown in Table 6, children have problems determining the correct form of the aorist for monosyllabic verbs with sonorant endings. As shown in Table 7, these errors tend to be errors of overregularization when children are younger and shift towards irregularization errors at older ages. Additionally, Nakipoğlu and Ketrez (2006) argue that the switch from preferring to overregularize to preferring to irregularize is due to children belonging to G1 mostly hearing -Ar ending monosyllabic verbs, and as more -Ir ending verbs enter children's lexicon, they start preferring irregularization. Additionally, Nakipoğlu and Ketrez (2006) suggest it is plausible that there could be a competition between -Ir and -Ar forms in G1, as there is an abundance of errors not only in the -Ar form but also in the -Ir form.

In summary, children's errors mostly occur when the verb is a monosyllabic verb that ends with a sonorant segment, and their error tendencies switch to -Ir irregularizations as they encounter more -Ir ending verbs. As they start to encounter more verbs ending with -Ir, their error rates on -Ar ending verbs decrease, which can be explained by their irregularization patterns switching from preferring the -Ar form to the -Ir form.

The following chapter will explain the dataset used to model the morphological learner and the supervised morphological learner model used in this study and provide the implementation of syllable information for the model.

CHAPTER 2

DATASET AND MODEL INFORMATION

This chapter provides an overview of the dataset, the model used in this study, and the changes made to the model, as well as outlining how the word generation algorithm picks a rule for target roots. The model is a Python implementation of the Minimal Generalization Learner from Albright and Hayes (2002). Albright and Hayes (2002) is a bottom-up model, where it builds generalized rules from the smallest and the narrowest rules possible. The generalization happens by comparing the rules and creating a more general rule based on the features shared between the rules.

2.1 Dataset

The dataset is Zargan Ltd.'s morphologically parsed Turkish dataset (Bilgin, 2016). This dataset is based on Sak et al. (2008) and the BOUN Corpus (Sak et al., 2010), a corpus that contains 405,794 entries, each of which contains word form, root, and morphological parse information of the word. Since Turkish marks subject agreement on the verb, and the 3SG is morphologically null, the pairs are generated only on the 3SG form to avoid problematic rules due to differences in person marker forms. In order to use it to train the model, verbs with 3SG in their features have been selected and then grouped in pairs such that the only difference is that one always contains an aorist marker and the other does not. Since the model extracts the designated morpheme by considering both left and right context, erroneous extractions arise in certain pairs. The example below shows one erroneous root pairing.

(9) Erroneous root pairing (sök ‘remove’)

([sökecek]$_{FUT}$, [söker]$_{AOR}$)

The left context for this pair would be ‘söke’, and because of that extra [e] in the root, the morpheme for aorist would be extracted as ‘r’; when it should be ‘er’. To combat this problem, only one other designated marker was selected to compare the aorist marker. This marker was the past tense marker (-DI). This has resulted in having 234 shared unique root pairs. This dataset had 130 -Ir, 61 -Ar, and 43 -r occurrences.

2.2 Overview of the model and changes made

Albright and Hayes (2002)’s Minimal Generalization Learner (MGL) implementations have been extended to numerous languages beyond English. For example, Albright (2002) applied the MGL to Italian verb conjugations, using it to capture phonologically conditioned “islands of reliability” in Italian inflectional classes. Albright and Hayes (2006) combined the MGL with a Gradual Learning Algorithm to model Navajo sibilant harmony, addressing patterns that are exceptionless in the input yet not extended productively by speakers. Jun and Albright (2017) used an MGL-based model to predict speakers’ innovative irregular alternations in verb paradigms in Korean. Kapatsinski (2010) employed the MGL to analyze Russian velar palatalization, testing how well its generalizations align with speakers’ intuitions. Likewise, Veríssimo and Clahsen (2014) reimplemented the MGL for Portuguese inflectional classes.

Moreover, Strik (2014) applied the MGL to Swedish. Oseki et al. (2019) incorporated MGL principles to help model Japanese inflectional morphology.

Ahyad & Becker (2020) used the MGL framework to examine vowel distribution patterns in Hijazi Arabic roots. Similarly, Kuo (2020) applied an MGL-based analysis to Tgdaya Seediq. All of these efforts showcase how the MGL framework has been variably implemented to learn morphological rules in a wide range of non-English languages. The following section outlines how Albright and Hayes (2002)'s MGL model is incorporated to capture the interaction between analogical forms and rule-governed forms in the Turkish aorist.

2.2.1 Overview of the model

As explained in the previous subsection, the model is based on Albright and Hayes (2002). It is a supervised morphological rule learner model that aims to generate morphological rules given pairs of surface forms that share a root and differ on one morphological marker. For example:

(10) Example Input for the model

([geldi]$_{PST}$, [gelir]$_{AOR}$)

When a pair is fed into the model, the model finds the maximal left shared substring and then finds the maximal right shared substring. These shared substrings are then removed from the words to get the non-shared part, which equates to B in Figure 1. Figure 1 below shows what the values 'A, B, C, D' are.

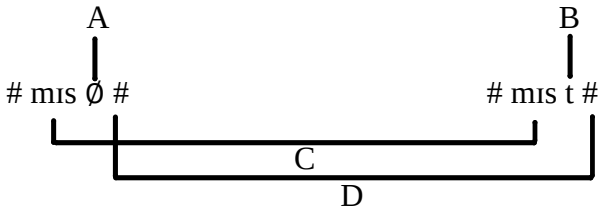

Figure 1. A diagram showing what constitutes the values: A, B, C, D
(Albright & Hayes, 2002, p. 3)

As explained above briefly, the baseline model works on the basis of shared features. The model first generates a rule for a root, then starts to generalize. Suppose the pair is <geldi, gelir>. The model first checks each sound segment starting from left to right, and when the strings don't match then the model saves matching strings up until that point and it is designated as the left side context (C). For this pair, the model would save 'gel' as the left side context. After obtaining the left-side context, the model now compares these forms starting from right to left, and, similarly, stops when the segments don't match. This segment is stored as the right-side context (D), and for this pair, it would be null as there is no shared segment from right to left. Finally, the model removes those matching left-side and right-side contexts from each segment in the pair. These are the morphemes that differ (B).

After obtaining all of these values, the model creates a rule in the template of A->B/C_D (See Figure 1 for the values of A, B, C and D). However, since Turkish does not have right-to-left spread of segmental properties, right-side context is not included in the rule template. Furthermore, Albright and Hayes (2002) argue that A will always be empty when dealing with prefixes or suffixes. It comes into play when there is a case of ablaut. For this pair, the generated rules would be:

(11) Generated rules for the <geldi, gelir>

0->ir/gel_#

0->di/gel_#

After generating a rule for each pair in the training dataset, the model starts to create generalizations based on the rules by comparing the rules with each other and against already generalized rules. Suppose 0->ir/gel_# is the first rule to be checked and there are no generalized rules for 'ir' yet. The model then looks at the next rule containing 'ir' and takes both rules as the first pair to start generalization. Suppose

0->ir/içil_# is the next rule containing ‘ir’. The model first ensures the B values (i.e., -ir in this case) match. After that, the model starts checking whether there is a shared left-side context starting from right to left. The moment the model finds a non-matching segment between the rules, it featurizes them and stores the matching features in the generalized rule. For these rules, the model first would start from right to left to check the matching segments and would store ‘l’ as the matching segment. Then the model would featurize ‘e’ and ‘i’. After this, the model takes the shared features of these vowels. The generalized rule would be like this:

(12) Minimally generalized rule for ‘ir’ (A->B/shared string {shared feature set}

0->ir/l {vowel: True, round: False, low: False, back: False}

Suppose the next rule to be checked is 0->ir/ver_#. Since there is an already generalized rule for ‘ir’, the model compares this ungeneralized rule with an already generalized rule. First it compares the left-side context of the ungeneralized rule to the generalized rule. Since ‘r’ and ‘l’ are not matching the model stops comparing and starts to featurize both ‘r’ and ‘l’ and stores the matching features. Now the generalized rule would be:

(13) Updated minimally generalized rule for ‘ir’

0->ir/{son: True, nas: False, voi: True, ant: True, cor: True, back: False}

Since there is no shared string now, the model does not check the shared strings; rather, it only checks matching features of the last segment in the ungeneralized rule to generalized rule.

Suppose now 0->ir/erit_ is the next rule. Now the model compares the 't' with the feature dictionary of the generalized rule. The resulting rule would be:

(14) Updated minimally generalized rule for 'ir'

0->ir/{nas: False, ant: True, cor: True, back: False}

This continues until there is no rule left to be generalized. After minimal generalization is done, the model calculates the reliability and confidence scores of the rules. Reliability is the ratio of a rule's hits to its scope, where hits are the forms the model got right, and the scope is the number of forms the rule can apply to. Confidence is the lower limit statistics based on Mikheev (1997). Reliability and confidence scores are calculated as follows:

(15) Calculation of reliability and confidence scores (taken from Albright and Hayes (2002))

a. Calculation of reliability scores

$$\text{reliability} = \frac{\text{hits}}{\text{scope}}$$

b. Calculation of confidence

$$\hat{p} = \frac{\text{hits}+0.5}{\text{scope}+1.0} \qquad \text{variance} = \hat{p} \times \frac{1-\hat{p}}{\text{scope}}$$

$$z = \Phi^{-1}\left(1 - \frac{1-\alpha}{2}\right)$$

$$\text{lower confidence limit (Confidence score)} = \hat{p} - z \times \sqrt{\text{variance}(\hat{p})}$$

Judging from this rule, it is impossible for this model to capture vowel harmony information as there is no shared feature set for vowels. As noted in the previous chapter, vowel harmony in Turkish spreads from the stem to the morpheme. Since there is no vowel information on the context of the rule in examples (12), (13)

and (14), the model would not be able to discern the harmonic difference between -ir and -ır. The following example shows the pseudo-code of how the model works.

(16) **Input:** Word pairs that differ in one morphological feature, namely, the tense/aspect marker.

- **For** each input pair **DO**:
    - Get the left maximal shared substring
    - Get the right maximal shared substring
    - Remove shared substrings from the words to get B
    - Constitute a rule in the template of A -> B / C_D[8]
    - Add the rule to the ruleset
    - **For** each pair of rules in the ruleset **DO**:
        - **If** rules share the same B value **DO**
            - Get the shared part starting from the rightmost part of both strings in C, until there is a mismatching sound pair.
            - Replace mismatching sound occurrences with the smallest natural class containing features
            - Compare the features of these mismatching sound occurrences and put them in a dictionary[9]
            - Put non-shared features in different dictionaries
            - Calculate reliability and confidence values

[8] As explained before, 'A' will always be empty when dealing with suffix markers (Albright & Hayes, 2002). 'D' will always empty too since Turkish does not have harmonic features spread from rightwards

[9] The model does not distinguish the sound type, i.e., a vowel and a consonant can be compared against each other to form the generalized rule.

2.2.2 Changes made to the Albright and Hayes (2002) model

Some changes have been made to test certain language acquisition theories. Alpha value (α) is changed to be a static value rather than a value based on a lookup table. The static alpha value is set to 0.75, but further experimentation with this value is needed to see whether this has an effect on the quality of the generated rules.

The implemented model was purposefully given limited phonological information to check whether parallel acquisition of both phonology and morphology was possible. To test this, the word generation results will be checked with the results from Nakipoğlu and Ketrez (2006). The reason for limiting phonological information is the literature on acquisition. McLeod and Crowe (2018) claim acquisition of phones starts as early as 1 year 0 months old and continues until 5 years old. Ravid (2019) argues that morphological structure acquisitions occur between the ages of 2 and 6. Considering these, it is expected that both morphological and phonological acquisition will happen simultaneously, which should limit the involvement of pre-conceived phonological rules as much as possible. For this reason, the implemented model has very limited access to phonology information during rule learning. The baseline model only has access to feature theory by Odden (2013), meaning the model only knows the features associated with sound segments.

2.2.3 Implementation of the syllable information to the baseline model

In order to test the impact of phonology, syllable information is added to the model. As shown in Tables 2, 3, and 4, the aorist marker is fully rule-governed in multisyllabic and vowel-ending roots. Syllable information was implemented using two parameters. The first parameter is a binary feature that checks whether the root is

monosyllabic. This parameter is calculated by looking at the last five sound segments of the root and counting how many vowels are present. The second parameter brings a change to the way the model creates and stores the left-side context. The left side context is changed from storing both the first non-matching sound segment's shared features and a shared string, to only storing the nucleus and the coda of the last syllable. The nucleus helps in picking out the correct form of the selected type[10]. Coda information helps in distinguishing vowel-ending stems since the lack of coda information is also stored in the coda. The aim of this implementation is to find out whether this implementation could learn all rule-governed forms.

The syllable information model differs from the baseline in that only the nucleus and coda[11] of the final syllable are checked. Suppose the pair is again <geldi, gelir>. Just like in the baseline model, the model first extracts shared segments (C & D) and non-shared segments (B). After obtaining these segments, the model extracts the nucleus and the coda from the last syllable. The extraction first targets the last segment of the stem, then, if it is a consonant, it extracts the first vowel preceding the extracted consonant. If the last segment is a vowel, then the model regards the coda as non-existent. For example, the first extracted sound for a stem like 'oku' would be 'u' as it extracts the word-final segment. Since it was a vowel, the model takes the coda as non-existent as there is no coda for the final syllable. For the <geldi, gelir> pair, the model would save 'gel' as the shared segment and extract 'e' as the nucleus and 'l' as the coda. After this, the model proceeds to create a rule, as explained in the baseline version. The example (17) below shows the constructed rules based on the pair <geldi, gelir>.

[10] Since vowel harmony draws feature values from the stem's final vowel, the nucleus of the last syllable is required to successfully choose the correct morpheme form.
[11] This model receives only the last segment of the coda. For example, in a verb like kork- 'be scared', only [o] and [k] will be present in the rule.

(17) Generated rules for <geldi, gelir>

0->ir/<e,l>[12]

0->di/<e, l>[12]

Similarly to the baseline version, after constructing a rule for each pair, the model starts to generalize the rules in a similar fashion to the baseline model. The difference from the baseline model is that now the nucleus and the coda are checked between each pair, which constructs the context for the information. In order to do this the model first featurizes both the nucleus and the coda in each rule and compares the feature sets against each other to obtain more generalized rules. Suppose the rule pair for the first generalization is 0->ir/e, l and 0->ir/i, l. The generalized rule is shown in example (18).

(18) Example of a minimally generalized rule

0->ir/ {back: False, round: False, vowel: True} {son: True, cont: True, nas: False, voi: True, ant: True, cor: True, back: False}

Similarly to the baseline version, the model now compares the generalized feature set like the one in example (18). Again, the non-generalized rule's context information (the last segment of the coda and the nucleus) is featurized and then compared against the generalized rule. Assume the next rule with the morpheme 'ir' is 0->ir/ e, r. Now the generalized rule would be:

(19) Updated minimally generalized rule

0->ir/ {back: False, round: False, vowel: True} {son: True, nas: False, voi: True, ant: True, cor: True, back: False}

[12] The angle bracketed notation is only used here to show the ordering of the pair. For brevity, the used notation will be simplified as A->B/ Nucleus, Coda henceforth.

Assume now 0->ir/ i, t is the next rule. Again, the model compares the shared nucleus and coda feature sets with this rule and creates a new shared feature set for both the nucleus and the coda. The resulting rule is:

(20) Updated minimally generalized rule

0->ir/ {back: False, round: False, vowel: True} {nas: False, ant: True, cor: True, back: False}

This continues until there is no rule left to be generalized. After minimal generalizing each rule, reliability and confidence scores are calculated in a similar fashion to the baseline model (Shown in example (15)). Based on the minimally generalized rules shown in examples (18), (19) and (20), this model would be able to capture vowel harmony as there is a specified shared feature set for each morpheme. This feature set would help the model discern harmonic information and would make the model be able to discern the difference between ‘ır’ and ‘ur’. Example below is a pseudo-code of the new model with syllable information model:

(21) **Input:** Word pairs that differ in one morphological feature, namely, the tense/aspect marker.

- **For** each input pair **DO**:
  - Get the monosyllabic information of the root
  - Get the left maximal shared substring
  - Get the last nucleus and coda of the left maximal substring
  - Get the right maximal shared substring
  - Remove shared substrings from the words to get B

— Constitute a rule in the template of A -> B / C_D[13]

— Add the rule to the ruleset along with the monosyllabic information

— **For** each pair of rules in the ruleset **DO**:

  — **If** rules share the same B value **DO:**

    - Get the intersection set of the nucleus and coda features of both rules
    - Check the monosyllabic information feature
      - **If** both are True **DO**:
        - Keep the feature value the same
      - **Else DO**:
        - Change the feature value to False
    - Calculate reliability and confidence values

2.3 Word generation algorithm

The performance of the model was tested using a word generation procedure based on the learned ruleset. The model works on the basis of comparing the stored left context features and the string to the root.

The core idea for this task is to understand the performance of each model. The word generation model works on the basis of matching the context of the root with the context of the provided rule. The model first ranks the rules in order of descending confidence values. After that, the model aims to match the C value of the

[13] Instead of using what the baseline model did for left-side context (C), only last nucleus and coda are stored.

root with the C value of the rule. If there is a match, the model uses that rule to generate the desired form for the desired root; however, if there is no match, the model checks the next rule until a match is made. Suppose the syllable information model has a rule list (in order of descending confidence scores) like the one shown in example (22):

(22) Minimally generalized rule list

‘0->ır/{round: False, vowel: True } {ant: True}’

‘0->ir/{back: False, round: False, vowel: True } {ant: True}’

‘0->r/{vowel: True} {non-existent}’

‘0->ar/{back: True, vowel: True } {nas: False}’

‘0->er/{back: False, vowel: True } {ant: True}’

‘0->ür/{back: False, round: True, vowel: True } {ant: True}’

‘0->ur/{ round: True, vowel: True } {ant: True}’

Also, suppose ‘izle’ is the desired root. The model first extracts the context information of the root. The context information of the root is ‘e’ and *null*[14]. The model first featurizes ‘e’ and *null*, which are represented as:

(23) Featurized context information of the root ‘izle’

e = {high: False, back: False, round: False, vowel: True}

null = {non-existent}[15]

The model then iterates over the rules to see whether any rule’s context is the subset of ‘e’ and null. Even though {round: False, vowel: True} is a subset of ‘e’, the

[14] Note that the syllable information model is able to differentiate non-existent coda. For example, the model would extract the root ‘oku’ as ‘u’ and *null*. Since the model is only able to look into the last syllable information, coda is not existent.

[15] Non-existent is a marker for the lack of any segment.

coda does not match, and because of this, the rule is skipped. The next rule shows the same problem and is also skipped. However, the third rule 0->r/ {vowel: True} {non-existent} is a match, so the model picks that rule. The B value of the rule is then used to generate the form for the desired root, and the result is ‘izle+r’.

Overall, the word generation model for syllable information works by comparing the features of the vowels and consonants. The word generation model aims to compare the left side context (C) with the root to decide which vowel harmony features to use. Since the empty set is also a subset of the feature dictionary of the first non-matching string of the root, an empty set is also considered to be matching.

Since syllable information changes how the learned rules are stored, the word generation code is changed slightly. Below is the pseudo-code of word generation for the model with syllable information implementation.

As for the word generation task for the baseline model, the only difference is how the context is constructed. For the baseline model, this task first compares the shared string. If it is a match, then the model removes the shared string from the desired root, and featurizes the last segment. After featurizing, the model compares the shared feature set in the minimally generalized rule to the feature set of the desired root. If the shared feature set for minimally generalized rule is the subset of the feature set for the desired root, the model uses this rule to generate the form. The following examples show the pseudo-code for the word generation model for syllable information (Example (24)) and the baseline model (Example (25)) below:

(24) **Input:** roots from which the forms will be generated, the desired marker to generate the form for, and the rule list

- **For** target root in input word list **Do**:
    - **For** roots in rule list, **Do**:
        - **If** target root is present in roots, **Do**:
            - Produce the form
    - **For** rules in rule list, **Do**:
        - **If** the monosyllabic information feature of the rule is True, and the target root's feature is False **DO**:
            - Check the next rule
        - **Else Do:**
            - Get the features of the nucleus and coda of the target root
            - Check whether the shared coda and nucleus features on the rule are a subset of the coda and nucleus features of the target root
                - **If** it is a match **DO:**
                    - Produce the form
                - **Else Do**:
                    - Check the next rule

(25) **Input:** roots from which the forms will be generated, the desired marker to generate the form for, and the rule list

- **For** target root in input word list **Do**:
    - **For** roots in rule list, **Do**:

— **If** target root is present in roots, **Do**:

- Produce the form

— **For** rules in rule list, **Do**:

— Check whether the shared string is present in the target root

— **If** the string is present DO:

- Remove the string from the target root
- Check whether the shared features set in C is a subset of the first mismatching sound segment in the target root.
- **If** it is a match **DO:**
  - Produce the form
- **Else Do**:
  - Check the next rule

Chapter 3 reports on the results of both the baseline rule learner, model with the implementation of the syllable information, and introduces three additional parameters implemented to the syllable information model.

CHAPTER 3

RESULTS AND ADDED PARAMETERS

This chapter provides an overview of the results of Minimal Generalization experiments and the results of the word generation model. This chapter also discusses the added parameters and their results. The results discussed in this chapter are both the training results of the Minimal Generalization Learner and the word generation task used to assess the models' performance.

As explained in Chapter 2, Minimal Generalization happens by first constructing non-generalized rules by comparing two forms that share a root but differ in one syntactico-semantic feature. Afterwards, these non-generalized rules are then compared against each other and against already generalized rules in order to get more generalized rules. After exhausting all available forms in the training dataset, reliability and confidence ratings are calculated. Reliability is the ratio of the number of forms in which the morpheme extraction[16] is correct to the total number of forms present in the training dataset. Whereas the word generation task compares the context of the generalized rules to the target root. If the features in the context of the generalized rule are a subset of the context of the target root, then the rules are used to generate the form. The same 43 roots used in the Nakipoğlu and Ketrez (2006) experiment are used in this word generation task as well.

[16] Since the model compares to forms that differ in one syntactico-semantic feature, there is a possibility that the rules could be extracted wrongly when dealing with certain forms. For example, when comparing pişir.ir 'cook.AOR' to pişir.iyor 'cook.PROG', the algorithm will extract 'r' as the aorist form. This is an instance of a 'hit' which hurts the reliability.

3.1 Baseline model results

The model produced eight rules for aorist forms, and out of the eight rules, the model successfully realized seven correct forms, and the wrong form ‘der’. The baseline model has an average confidence rate of ~76% and an average reliability rate of ~84%. The trained rules are below: (See Table 8)

Table 8. Baseline Model’s Minimally Generalized Outputs

Featurized Rule (A->B/{shared features}{shared segment}[17])	Reliability	Confidence
0->ir->{}*	1	0.97
0->ır->{}*	0.984	0.95
0->r->{}*	0.977	0.93
0->ar->{}*	0.968	0.91
0->er->{}*	0.965	0.90
0->ür->{'son': True, 'voi': True}*	0.916	0.77
0->ur->{'son': True, 'voi': True}*	0.916	0.77
0->der->{}*	0	-0.09

The erroneous aorist form in Table 8 most likely results from how the MGL builds rules. As shown in Figure 1, each rule isolates the maximally shared left and right substrings (C, D) and treats the remaining segment as the non-shared B value. For some lexically conditioned stems ending in the alveolar stop [t], there is an adjustment rule that voices [t] to [d] when it is followed by a vowel. Since the model posits the B value (i.e., the aorist form) as the non-shared maximal substring, and this is a regular phonological process, the model picked this to be a rule as well.

[17] * sign denotes there are not any shared string

For the word generation task using these minimally generalized rules, the model only produced 4 correct forms out of 43 forms used in the Nakipoğlu and Ketrez (2006) experiment. Since there are no distinguishing features in all but two rules, the model picked the rule with the highest confidence and produced the form.

3.2 Syllable information model results

Since the baseline model was not able to capture morphophonological information, syllable information is implemented into the model, as explained in Chapter 2, to explore whether the inclusion of syllable information could help the model capture these morphophonological rules. Briefly, the syllable information model stores the last coda and nucleus as two different segments, and minimal generalization occurs by comparing these two segments to each other to generate generalized rules.

The minimal generalization results produced the same forms; however, the reliability ratings and the average confidence scores are higher than the baseline results. This model averaged a reliability rate of ~87% and a confidence rate of ~81%. Word generation task shows the model is able to capture vowel harmony information and is able to produce correct forms for fully rule-governed roots; however, it failed to capture monosyllabic rules correctly and picked the rule with the highest confidence valued -Ir form. Out of 43 roots used in the Nakipoğlu and Ketrez (2006) experiment, the model is able to predict 25 correct forms. All errors occurred on the monosyllabic roots, most specifically sonorant ending monosyllabic roots. Since the codas were mostly empty on the generalized rules, the model mostly chose -Ir as it has the highest confidence values. Table 9 below shows the learned rules of the syllable information model and Table 10 shows some errors from the word generation task.

Table 9. Syllable Information Model's Generalized Outputs

Featurized Rule (A->B/{nucleus}{coda})	Reliability	Confidence
0->ır->{'back': True, 'round': False, 'vowel': True}{}	1	0.97
0->r->{'vowel': True}{non-existant}	1	0.97
0->ir->{'low': False, 'back': False, 'round': False, 'vowel': True}{}	1	0.97
0->ar->{'back': True, 'vowel': True}{}	1	0.96
0->er->{'low': False, 'back': False, 'vowel': True}{}	1	0.95
0->ur->{'low': False, 'back': True, 'round': True, 'vowel': True}{'cor': True, 'back': False}	1	0.89
0->ür->{'high': True, 'low': False, 'back': False, 'round': True, 'vowel': True}{'cor': True, 'back': False}	1	0.89
0->der->{'high': False, 'low': False, 'back': False, 'round': False, 'vowel': True}{'son': False, 'cont': False, 'nas': False, 'voi': False, 'ant': True, 'cor': True, 'back': False }	0	-0.09

Table 10. Word Generation Results of the Syllable Information Model

Produced Form	Correct Form	Root
atır	atar	at 'to throw'
bakır	bakar	bak 'to look'
görer	görür	gör 'to see'
dalır	dalar	dal 'to dive'
yanır	yanar	yan 'to burn'

3.3 Added parameters and their results

The errors occurred due to the model having an underspecified context, meaning the coda context of the rules is mostly empty, which results in the model preferring the highest-ranked compatible rule. The empty context likely occurred due to the

irregular forms. For example, the -ar form contains çal ‘steal’, tut ‘hold’, çak ‘hit’, don ‘freeze’. Despite the small sample size of 4, there is no shared feature to include in the context. Due to the underspecification of the contexts, the highest-ranked rule is almost always a perfect match for the target root during the word generation task. To combat this situation, three new parameters are independently checked to see whether they could address this issue.

The first parameter ensures that the model identifies at least one shared feature in both the nucleus and the coda, each within their respective feature dictionaries (This parameter will be referred to as “no empty context” henceforth). The “no empty context” model differs from the syllable information model in that this model now aims to have at least one shared feature in each generalized rule and does not include the exceptions to the generalization. Rather it treats those cases as exceptions. Recall that syllable information model stores the last occurrence of the nucleus and the coda of the stem[18]. The parameter in question (i.e., ‘no empty context’) blocks the model from positing an empty set as a shared feature set for the generalized rules. Assume example (26) is the generalized rule.

(26) Minimally generalized rule used to show the interaction of the parameter

0->ir/ {back: False, round: False, vowel: True} {ant: True, back: False}

Further assume that the next rule is 0->ir/ o, k[19]. The model compares the shared nucleus and coda feature sets of the generalized rule with this rule and creates

[18] As explained earlier, the model targets the last segment of the coda. So, in a root like kork- ‘be scared’, the model first checks the final segment [k], and since it is a consonant, it looks for the first vowel preceding the final consonant, [o].

[19] This rule is generated to show how this parameter would interact with the model. There is no form that in Turkish that would generate this morphophonological rule.

a new shared feature set for both the nucleus and the coda. The resulting rule is shown in example (27):

(27) Updated minimally generalized rule

0->ir/ {vowel: True} {}

Since the resulting rule would include an empty set in the coda feature set, the model posits the rule 0->ir/ o, k^{18} as an exception and continues with the last generalized rule which is 0>ir/ {back: False, round: False, vowel: True} {ant: True, back: False'}, and moves on to the next ungeneralized rule. The forms that are counted as exceptions are stored along with the minimally generalized rule. These exceptions are accessible during the word generation task. During this task, the algorithm checks the exception list as well to see whether the desired root is counted as an exception. As with all other models, the "no empty context" model continues until there is no minimally generalized rule left. As can be seen from Tables 9 and 11, the model now provides a better context to match the roots with a rule, and as a result, the performance is markedly improved in the word generation task. Overall, this implementation marked only a limited number of forms as exceptions—a total of 9 forms across 4 rules. This shows that the model is trying to capture information rather than categorizing everything as an exception. The word generation task shows an increase in the amount of roots the model got right. Out of 43 forms, the model got 29 of them correct. This model performs better overall and is able to correct some errors of the syllable information model. Table 11 below shows the learned rules from the syllable information model with "no empty context" parameter implemented, and Table 12 shows some errors from the word generation task for this model.

Table 11. The Syllable Information Model and "No Empty Context"s Minimally Generalized Outputs

Featurized Rule (A->B/{nucleus}{coda})	Reliability	Confidence
0->ır->{'back': True, 'round': False, 'vowel': True}{'cor': True, 'back': False}	1	0.97
0->r->{'vowel': True}{non-existent}	1	0.97
0->ir->{'back': False, 'round': False, 'vowel': True}{'cor': True, 'back': False}	1	0.96
0->ar->{'back': True, 'vowel': True}{'nas': False}	1	0.95
0->er->{'back': False, 'vowel': True}{'nas': False}	1	0.95
0->ür->{'high': True, 'back': False, 'round': True, 'vowel': True}{'cor': True, 'back': False}	1	0.89
0->ur->{'back': True, 'round': True, 'vowel': True}{'cor': True, 'back': False}	1	0.89
0->der->{'high': False, 'back': False, 'round': False, 'vowel': True}{'son': False, 'cont': False, 'nas': False, 'voi': False, 'ant': True, 'cor': True, 'back': False}	0	-0.09

Table 12. Word Generation Results of the Syllable Information and "No Empty Context" Model

Produced Form	Correct Form	Root
durar	durur	dur 'to stop'
olar	olur	ol 'to become'
görer	görür	gör 'to see'
dalır	dalar	dal 'to dive'
yanır	yanar	yan 'to burn'

As can be seen from the results, not only does the model perform better than its baseline, i.e., syllable information model, but also it was able to correct 4 mistakes: bak ‘to look’, çık ‘to exit’, yap ‘to do/make’. This shows the impact of a narrower context in generalized rules.

The next parameter checks whether the context of a non-minimally generalized rule with a different morpheme[20] overlaps with an already minimally generalized rule’s context (This parameter will be referred to as the “rule exception checker” henceforth). The “rule exception checker” model differs from the “no empty context” model in that this model aims to remove clashing rules from the generalization aiming to reduce the bleeding effect. Similarly to the syllable information model, the model constructs the non-generalized rules based on shared segments. Different from the syllable information model, after each generalization in this model, all non-generalized rules’ context information (i.e., the feature sets of the nucleus and the coda) are checked to see whether there is a rule whose context information clashes with an already generalized rule. Suppose example (28) is the current generalized rule:

(28) Minimally generalized rule used to show the interaction of the parameter

0->ir/ {back: False, round: False, vowel: True} {ant: True, back: False}

The model now scans every non-generalized rule, including the ones that have different B values, and decides whether any ungeneralized rule has the same context information as this model. The model decides that a rule has the same

[20] For example, a rule like ‘0 -> ir ->{‘high’; False, ‘low’: False, 'back': False, 'vowel': True}{'son': False, 'cont': False, 'nas': False, 'voi': True, 'ant': True, 'cor': True, 'back': False}’ overlaps with the following minimally generalized rule ‘0->ar->{'back': True, 'vowel': True}{'nas': False}’

context information if the context information of the minimally generalized rule is a subset of the ungeneralized rule. If there are any such rules, the model removes them from the generalization and counts them as exceptions in a similar fashion to the "no empty context" model. Suppose '0->ar/ 'e' 'l'_' is a rule. During this process the model would flag this rule as an exception as the context is the same as the generalized one. Since the shared feature set for the generalized rule in example (28) is a subset of the feature set of the segment 'l', the model would flag the non-generalized rule as an exception and continue generalizing. The comparison happens every time a rule is generalized and each time the model compares every rule in the non-generalized rule set. Crucially, now the model will skip the rule 0->ar/e, l during minimal generalization for the 'ar' morpheme. The aim of this model is to effectively remove clashing context information of rules that contradict each other. The results of this model can be seen in Table 13 below:

Table 13. The Syllable Information and "Rule Exception Checker"s Minimally Generalized Outputs

Featurized Rule (A->B/{nucleus}{coda})	Reliability	Confidence
0->ır->{'back': True, 'round': False, 'vowel': True}{}	1	0.97
0->r->{'vowel': True}{non-existant}	1	0.97
0->ir->{'back': False, 'round': False, 'vowel': True}{}	1	0.97
0->ar->{'back': True, 'vowel': True}{}	1	0.96
0->er->{'back': False, 'vowel': True}{}	1	0.95
0->ür->{'high': True, 'back': False, 'round': True, 'vowel': True}{'cor': True, 'back': False}	1	0.89
0->ur->{'back': True, 'round': True, 'vowel': True}{'cor': True, 'back': False}	1	0.89
0->der->{'high': False, 'back': False, 'round': False, 'vowel': True}{'son': False, 'cont': False, 'nas': False, 'voi': False, 'ant': True, 'cor': True, 'back': False}	0	-0.09

As seen from Tables 9 and 13, they behave the same, and this is reflected in the word generation as well. This parameter behaved exactly the same, with 25 correct forms produced out of 43 verbs. One reason this may be the case is that empty context is not prohibited, as it was in the previous parameter. Since this model uses a context similar to that of the syllable information model, it exhibits similar behavior.

The last implemented parameter first scans all of the rules, then finds the most commonly used {feature: value} pair for each morpheme. Then the model ensures that the {feature: value} pair is always present in the morpheme's context (This parameter will be referred to as the "most common feature in context" henceforth). The "most common feature in context" model differs from the "no empty context" model in that this model aims to have the most common feature shared for all the same morphemes present in the context. Whereas the "no empty context" model only aimed to have at least one {feature: value} pair in their context.

After generating a rule for each pair in the training dataset, similar to the syllable information model, the model featurizes every rule before starting to create generalizations. Assume 0->ir/e, l and 0->ir/i, k are rules. The model would featurize these rules, respectively, as shown in example (29) below.

(29) Minimally generalized rule used to show the interaction of the parameter

0->ir/ {back: False, round: False, high: False, low: False, vowel: True} {son: True, cont: True, nas: False, voi: True, ant: True, cor: True, back: False}

0->ir/ {back: False, round: False, high: True, low: False, vowel: True} {son: False, cont: False, nas: False, voi: False, ant: False, cor: False, back: True}

After featurizing all rules, the model first groups the rules based on their morpheme (i.e., B values. For example, grouping all 'ar' rules together). Then the model starts keeping a tally for each {feature: value} pair in each group. The model then selects the winning {feature: value} pair for each morpheme (i.e. the B value). The generalization happens similarly to the syllable information model. During the generalization, the model posits a rule as an exception if the winning {feature: value} pair is not included in the context. For example, assume only 0->ir/e, l and 0->ir/i, k are rules for the 'ir' morpheme. The model would count each {feature: value} pair encountered in the featurized rules shown in the example (29). The winning {feature: value} pair for the nucleus would be either 'back: False' or 'round: False'[21], and for the coda would be 'nas: False'. After determining the {feature: value} pairs for every morpheme, the model starts to compare the rules against each other. For example, assuming 'round: False' is a winning {feature: value} pair for -ir a rule like 0->ir/o, k[22] would be counted as an exception since the winning {feature: value} pair is not included in the context information of the rule 0->ir/ o, k[21].

The aim of this model is to generate stricter context information based on the most recurrent feature. The handling of the exception is similar to the "no empty context" and "rule exception checker" parameters.

This parameter, along with the first parameter ("no empty context"), generates a narrower context than the "rule exception checker" parameter and the syllable information model. However, having a narrower context did not help in achieving better results, as this model performed worse than the syllable information model on the word generation task. Out of 43 forms, this model produced 23 of them

[21] In the case of a tie, the model picks the feature that occurs alphabetically earlier
[22] There is no form in Turkish that would generate this rule. This is used only to show how the model interacts.

correctly. Table 14 below shows the learned rules of the “most common feature in context” model and Table 15 shows some errors from the word generation task.

Table 14. The Syllable Information and “Most Common Feature in Context”s Minimally Generalized Outputs

Featurized Rule (A->B/{nucleus}{coda})	Reliability	Confidence
0->r->{'vowel': True}{non-existant}	1	0.97
0->ir->{'back': False, 'round': False, 'vowel': True}{'cor': True, 'back': False}	1	0.96
0->ır->{'back': True, 'round': False, 'vowel': True}{'cor': True, 'back': False}	1	0.90
0->ar->{'back': True, 'vowel': True}{'nas': False}	1	0.89
0->ur->{'back': True, 'round': True, 'vowel': True}{'cor': True, 'back': False}	1	0.89
0->ür->{'high': True, 'back': False, 'round': True, 'vowel': True}{'cor': True, 'back': False}	1	0.89
0->er->{'back': False, 'vowel': True}{'nas': False}	1	0.85
0->der->{'high': False, 'back': False, 'round': False, 'vowel': True}{'son': False, 'cont': False, 'nas': False, 'voi': False, 'ant': True, 'cor': True, 'back': False }	0	-0.09

Table 15. Word Generation Results of the Syllable Information and “Most Common Feature in Context” Model

Produced Form	Correct Form	Root
donur	donar	don ‘to be frozen’
yüzür	yüzer	yüz ‘to swim’
gülür	güler	gül ‘to laugh’
yerir	yerer	yer ‘to criticize’

3.4 Dataset manipulation experiments

3.4.1 Removal of sonorants from the dataset

After seeing the models having problems with mainly sonorant ending roots, sonorant ending monosyllabic roots are removed from both the training and the testing dataset to check whether the model can produce correct results on monosyllabic forms. As a result, the number of forms present in the test dataset reduced to 19 from 43. This manipulation was aimed at checking whether the model could learn to figure out the distinction that occurs because of the syllabicity.

The syllable information has better accuracy of ~63% (Corresponds to 12 correct out of 19 possible test cases), whereas in the previous experiment, the accuracy was ~58%. Table 16 shows some errors from the word generation task.

Table 16. Word Generation Results of the Syllable Information Model Trained on a Dataset with No Sonorant Ending Monosyllabic Roots

Produced Form	Correct Form	Root
atır	atar	at ‘to throw’
yatır	yat	yat ‘to lay’
gitir	gider	git ‘to go’
bakır	bakar	bak ‘to look’
kalkır	kalkar	kalk ‘to get up’

As seen from Table 16, the broad context problem, where the rules’ context does not convey enough information to pick a rule correctly, persists. Because of the broad context problem, the same experiment is also done with the additional parameters designed to solve the broad context problem. “No empty context” parameter has ~84% accuracy (16 correct forms out of 19), whereas in the previous

experiment, it had ~67% accuracy. This boost suggests that narrowing the context yields better results overall, as this model only failed to capture forms ending with [t]. Table 17 below shows all of the wrong productions of the word generation task.

Table 17. Word Generation Results of the Syllable Information and "No Empty Context" Model Trained on a Dataset with No Sonorant Ending Monosyllabic Roots

Produced Form	Correct Form	Root
atır	atar	at 'to throw'
yatır	yatar	yat 'to lay'
gitir	gider	git 'to go'

Rule exception checking parameter yielded the same results as the syllable information model, suggesting that checking whether a different rule's context encroaches on another rule might not truly be the case. This parameter had 12 correct forms generated out of 19. Admittedly, this is an increase from the base model; however, the model still failed to produce rule-governed monosyllabic forms as there were no -Ir ending monosyllabic roots present in both the training and the testing dataset. Table 18 shows some of the errors produced by this parameter.

Table 18. Word Generation Results of the Syllable Information and "Rule Exception Checker" Model Trained on a Dataset with No Sonorant Ending Monosyllabic Roots

Produced Form	Correct Form	Root
atır	atar	at 'to throw'
yatır	yatar	yat 'to lay'
gitir	gider	git 'to go'
çıkır	çıkar	çık 'to leave'

The surprising factor is that the "most common feature in the context" parameter not only performed better than both the syllable information model and the rule exception parameter but also was the best-performing model out of all four models. The model produced 17 forms correctly out of 19, with an 89% accuracy. Table 19 below shows the errors generated by the word generation task.

Table 19. Word Generation Results of the Syllable Information and "Most Common Feature in Context" Model Trained on a Dataset with No Sonorant Ending Monosyllabic Roots

Produced Form	Correct Form	Root
yüzür	yüzer	yüz 'to swim'
git	gitir	git 'to go'

The results of this experiment suggest that the "most common feature in context" parameter is the best at producing rule-governed forms if the training dataset does not contain irregularities.

Overall, these results suggest that the removal of most problematic cases shows the models can learn better. Most interestingly, the "most common feature in context" model performed the best, which suggests this parameter is best at capturing rule-governed forms.

3.4.2 Changing the skewness of the dataset

Following the interesting results in the previous section, where all models improved in performance, a third experiment is conducted to see whether artificially manipulating -Ar ending forms in the training dataset could posit -Ar as the default and pick -Ar as the prominent form. In order to test this, additional -Ar ending non-

words were added to the training dataset. The added forms always ended in non-sonorant endings. The training dataset now has 401 shared unique root pairs, and the distribution of these pairs is: 114 -ar, 114 -er, 65 -ir, 45 -ır, 11 -ur, 11 -ir, and 43 -r. Unlike the previous experiment, where the testing dataset was reduced to 19 forms, the testing dataset is not reduced; all 43 forms used in the acquisition experiment (Nakipoğlu & Ketrez, 2006) are present in the testing dataset of this experiment.

For the syllable information model, there seems to be an increase in accuracy, as this one was able to get 34 forms correct out of 43 with 79% accuracy. The errors, however, were always towards overregularization, i.e., the model always picked the -Ar form over -Ir. Additionally, the erroneous forms ended with a sonorant except in one case, git 'go'. As explained before, to capture that form, the model needed to learn phonological rules of Turkish, for which the model is not trained; however, it still picked the -Ar form for the production. Table 20 below shows some errors from the results of the word generation task.

Table 20. Word Generation Results of the Syllable Information Model Trained on a Manipulated Dataset Where -Ar Ending Forms are the Majority

Produced Form	Correct Form	Root
bular	bulur	bul 'to find'
kalar	kalır	kal 'to stay'
görer	görür	gör 'to see'

The model with "no empty context" parameter performed worse than the syllable information model; however, it had both types of regularization errors. This suggests the model is trying to capture certain features regarding these forms. The model with the "no empty context" parameter had 31 correct out of 43 forms, with a

72% accuracy. Again, the increase in performance continues, but this increase in performance could be due to either having a bigger dataset[23] or having a more skewed distribution of forms in monosyllabic roots. Crucially, unlike the syllable information model, "no empty context" parameter produced both types of regularization errors. See Table 21 for errors on the word generation task.

Table 21. Word Generation Results of the Syllable Information and "No Empty Context" Model Trained on a Manipulated Dataset Where -Ar Ending Forms are the Majority

Produced Form	Correct Form	Root
durar	durur	dur 'to stop'
bular	bulur	bul 'to find'
donur	donar	don 'to freeze'
yanır	yanar	yan 'to burn'

The trend of the rule exception parameter producing the same results as the syllable information model continues in this experiment as well. The model produced 34 correct forms out of 43 with 79% accuracy. Similar to the syllable information model, "rule exception checker" model also produced only -Ar ending errors. These observations reinforce the inference that this parameter might not be working as expected. By removing rules that bleed into other forms, the model might be positing rules that could have been useful for the morpheme that the rule is part of as exceptions. Table 22 illustrates some errors on the word generation task.

[23] The implications of this will be explained more in the following chapter.

Table 22. Word Generation Results of the Syllable Information and "Rule Exception Checker" Model Trained on a Manipulated Dataset Where -Ar Ending Forms are the Majority

Produced Form	Correct Form	Root
alar	alır	al 'to take'
bular	bulur	bul 'to find'
donur	donar	don 'to freeze'
kalar	kalır	kal 'to stay'
görer	görür	gör 'to see'

The "most common feature in the context" parameter performed worse than the syllable information model. It only made overregularization errors on sonorant ending monosyllabic roots, similar to how the syllable information model performed. This parameter performed worse than the syllable information model, with the additional parameter. This parameter had 28 correct forms out of 43, with 65% accuracy. Table 23 shows some errors in the word generation task.

Table 23. Word Generation Results of the Syllable Information and "Most Common Feature in Context" Model Trained on a Manipulated Dataset Where -Ar Ending Forms are the Majority

Produced Form	Correct Form	Root
bular	bulur	bul 'to find'
donur	donar	don 'to freeze'
kalar	kalır	kal 'to stay'
görer	görür	gör 'to see'

Overall, this experiment shows the models are able to capture a dominant form. However, the behavior of these models diverges from that of the children, in

that these models, with the exception of "no empty context", always prefer to overregularize toward the skewed type in the dataset. Both regularization errors seen on the model with the "no empty context" parameter are due to the features in the context. -Ar forms have a nasal feature as a feature in the coda's context information, and coronal and back features are found as a feature in the coda's context information.

The syllable information model performed a lot better than the baseline model; however, it had errors mostly due to the underspecified context information present in the rules. It seems the no context parameter solves the underspecification issue of the context, but this solution also impacts its ability to identify a default/elsewhere form. The "most common feature in the context" parameter is able to identify the default/elsewhere form if there are no exceptions to the case, but it performs worse if there are exceptions present, even though the majority of the forms are rule-governed.

3.5 Chapter summary

In summary, this chapter has shown that having a more specific context helps these models to capture the distribution better. Additionally, it is also shown that the models capture information better if there are no exceptions, and the performance increases as the number of rule-governed forms increases. Table 24 shows the word generation task results of all the experiments conducted with these models.

Table 24. Summarized Results of Word Generation Tasks

	Default	Removal of Sonorant ending monosyllabic roots	Changing the skewness of the dataset
Baseline Model	4/43 (9%)	N/A	N/A
Syllable Information Model	25/43 (58%)	12/19 (63%)	34/43 (79%)
"No Empty Context" Model	29/43 (67%)	16/19 (84%)	31/43 (72%)
"Rule Exception Checker" Model	25/43 (58%)	12/19 (63%)	34/43 (79%)
"Most Common Feature in Context" Model	23/43 (53%)	17/19 (89%)	28/43 (65%)

The results in Table 24 suggest the models show an increase in accuracy when the dataset is fully rule-governed. This suggests that these models are best at capturing rule-governed forms. However, the performance increase on changing the skewness suggests the models are also able to posit a default/elsewhere form that is used when the context is not narrow enough to correctly predict the correct form.

The next chapter will discuss the regularization errors of these models and compare them against the child acquisition dataset.

CHAPTER 4

DISCUSSION

This chapter aims to interpret the behavior of the models and assess how closely they align with the regularization patterns observed in the child acquisition dataset. Building on the previous chapter, where syllable information and additional parameters were introduced to address the model's limitations (particularly in handling monosyllabic roots). This chapter evaluates whether these modifications lead to patterns resembling those found in child language acquisition. Specifically, model outputs are compared to the developmental trajectory observed in Nakipoğlu and Ketrez (2006). As children are increasingly exposed to -Ir ending forms, their preferred aorist form shifts away from the -Ar, which reflects the changing balance between overregularization and irregularization (See Figure 2). This chapter explores whether similar patterns emerge in the model under comparable distributional conditions.

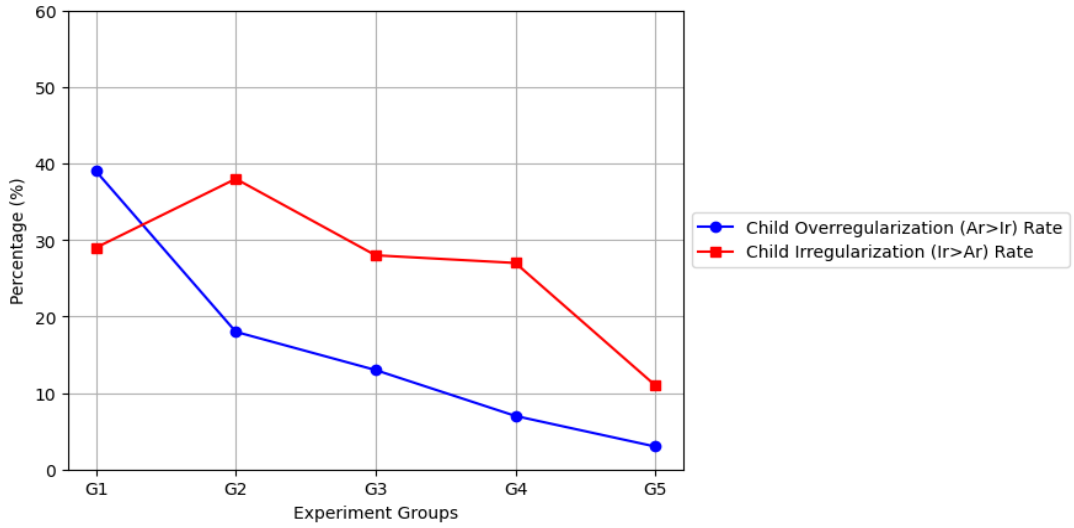

Figure 2. Regularization error rates for children
(Nakipoğlu and Ketrez (2006, p.406)

As explained in Chapter 1, the child acquisition experiment (Nakipoğlu & Ketrez, 2006) is conducted on 135 children as participants, and they showed the participants a picture and asked them to articulate what is happening in the picture. They found that their production during the early language acquisition period is closely tied to their lexicon, which is reflected in the figure above. As they encounter more -Ir using roots, the children start to prefer -Ir over -Ar, which is reflected in the noticeable increase of irregularization rate on G2 (Nakipoğlu & Ketrez, 2006).

4.1 Comparison of regularization rates of the models to the child acquisition dataset

4.1.1 Reanalysis of the regularization terminology

Nakipoğlu and Ketrez (2006) explain overregularization errors as cases where an irregular form is produced as if it were regular, thus regularizing an irregular form. Irregularization errors, on the other hand, occur when a regular form is produced as if it were irregular, thus irregularizing a regular form. The graphs used in the following subsections follow these definitions closely, and they reflect the distinction between overregularization and irregularization as described by Nakipoğlu and Ketrez (2006). As explained in Chapter 1, the dominantly encountered aorist form for monosyllabic roots is -Ar, and Nakipoğlu and Ketrez (2006) take -Ar to be the default form. They argue that children are exposed to more -Ar-taking forms than -Ir-taking ones during early acquisition, which explains the sudden spike in irregularization error rate observed in G2 in Figure 2. However, the training dataset for the models had high counts of -Ir taking forms as explained in Chapter 2. This is a problem because if -Ar is the default/elsewhere form for monosyllabic roots based on the distribution of aorist form, then -Ir should be the default form for the training dataset, as the distribution of -Ir/-Ar in the whole dataset is 130/61.

On the other hand, training data (Without any manipulations) contains only 62 monosyllabic forms. Of those 62 forms, 17 of them end with a sonorant consonant, and of those 17 forms, 14 of them take -Ar. In total, out of 62 monosyllabic forms, 59 of them end in -Ar. However, Minimal generalization links whether a generalized rule is composed of only monosyllabic roots, which results in generalized -Ir rules having roots that are both monosyllabic and multisyllabic. This creates a problem of labelling whether -Ar or -Ir should be the default form for the models.

Overall, the model does not fully replicate child behavior regardless of the error labels. However, some models do capture the irregularization pattern observed in children, where the error rate first increases and then decreases. More importantly, all models, with the exception of the baseline model, are able to produce rule-governed forms without error. This indicates that the minimal generalization learner is capable of learning productive morphological rules when the syllable information is provided. While the comparison to child data suggests potential similarities in behavior, further experimentation and more naturalistic training data are needed before claiming that the model mirrors child learning patterns. Next section compares the results to the acquisition data (Nakipoğlu & Ketrez, 2006).

4.1.2 Experiment results of the type manipulated training dataset

To simulate the acquisition experiment conducted in Nakipoğlu and Ketrez (2006), the model was trained 135 times, matching the number of children in their study. These 135 runs are divided into five groups, where each group represents a different age range and is defined by the amount of input data given to the model. To simulate lexicon expansion as children age, type frequency is manipulated by repeating the

training data with numerical prefixes added to each root and form. For example, G1 had no manipulation, while G3 had the original dataset plus additional copies of it marked with the numbers 1 and 2 at the start, which triples the amount of training input. Example (30) below shows some rows from the training dataset for G3:

(30) Some rows from the training dataset for G3

{gel, geldi, gelir}, {1gel, 1geldi, 1gelir}, {2gel, 2geldi, 2gelir}

Datasets are shuffled randomly each time to simulate variation in the order that children hear words. Since the model builds generalized rules based on shared features between forms, fixed item order would otherwise lead to the same rules being learned each time. This setup allows the model to better reflect the kind of variability that occurs in child language acquisition.

Berman (2016) explains Pearson's correlation coefficient (r) as a measure of the similarity between two data objects and describes the p-value as the likelihood that this similarity occurred by chance. Pearson's coefficient (r) value ranges from -1 to 1, and a positive value suggests a positive relationship, meaning both are increasing or decreasing at the same time, and vice versa for a negative value. A Pearson's correlation coefficient (r) greater than 0.5 (or less than -0.5 for negative relationships) suggests a high degree of similarity between the two data points. A positive relationship means both values are increasing or decreasing at the same time. A p-value lower than 0.05 indicates that the observed relationship is statistically significant, meaning it is unlikely to have occurred by random chance.

For this experiment, the error rates of the acquisition experiment (Nakipoğlu & Ketrez, 2006) are used as a reference to the models' error rates on monosyllabic roots with sonorant endings. Table 25 below shows Pearson's coefficient score and p-values for each model.

Table 25. Pearson's Coefficient Score and P-values for All Models

Model Name	Regularization Type	Pearson's Coefficient (r)	P-Value
Syllable	Overregularization	0.54	0.34
Information	Irregularization	0.19	0.75
No empty	Overregularization	0.72	0.16
Context	Irregularization	-0.52	0.36
Rule Exception	Overregularization	-0.21	0.72
Checker	Irregularization	0.83	0.08
Most Common	Overregularization	0.85	0.06
Feature in Context	Irregularization	-0.73	0.15

Even though results suggest there are no statistically significant results in one model to claim that it behaves similarly, some models show strong correlation to the acquisition study (r-score); however, none of them show statistical significance (p-value < 0.05). Overall, these statistical results suggest that manipulation of type frequency does not seem to show statistically significant similarities between the models and the children.

The results of the syllable information model suggest the model is not improving in performance and is not behaving similarly to the child with regards to the regularization rates, which is supported by the r and p values. The model seemingly has a stable error rate regardless of the increase in the amount of data

input, which underlines the problem of broad context again. Even though the model has access to bigger data, because the context is so broad that the model mostly picks the highest ranked generalized rule. Figure 3 shows the error rates of this model and the child acquisition experiment.

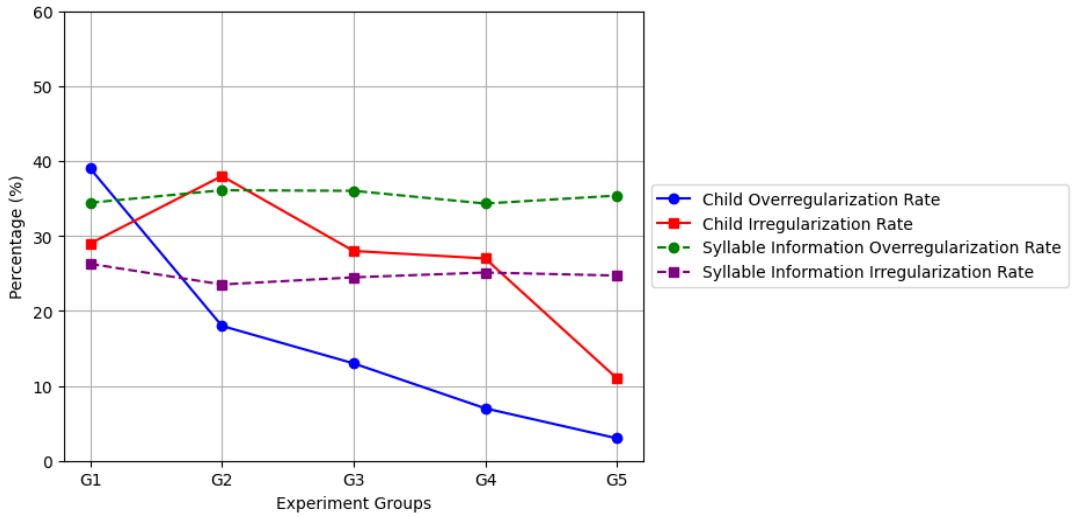

Figure 3. Child acquisition dataset and syllable information model regularization rates
(Nakipoğlu and Ketrez (2006, p.406)

The “No empty context” parameter, on the other hand, shows a similar curve to the child in terms of overregularization error rates; however, as explained previously, the regularization error types are opposites. The reversal of the labels for the model shows a correlation in the decrease in irregularization error rate throughout the groups. Figure 4 shows the error rates of this model and the child acquisition experiment.

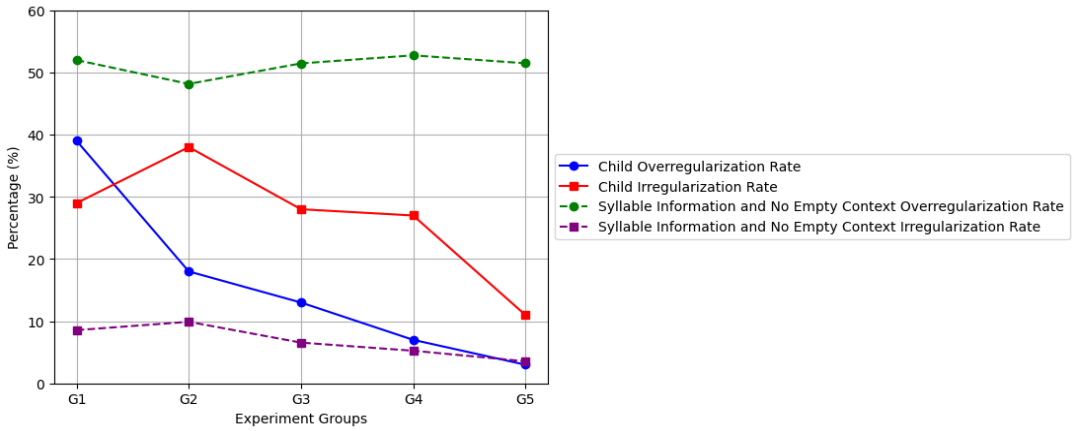

Figure 4. Child acquisition and syllable information and "no empty context" model regularization rates
(Nakipoğlu and Ketrez (2006, p.406)

As mentioned in the previous chapter, the increase in performance is due to both a larger training dataset and the skewed distribution of input forms. The error rate is reduced in both overregularization and irregularization in G5, suggesting that having a bigger dataset increases performance. However, at the same time, the error rates in G1, G2, and G3 suggest that the skewness of the training dataset also has an impact on performance. Overall, the "no empty context" parameter's behavior shows similarities to children's behavior in overregularization errors, but fails at capturing irregularization behavior, highlighting the broad context problem, which occurs when the context is not narrow enough to predict a root correctly, and the impact of skewness of the dataset.

The "rule exception checker" parameter shows results similar to those of the syllable information model. Figure 5 shows the error rates of this model and the child acquisition experiment.

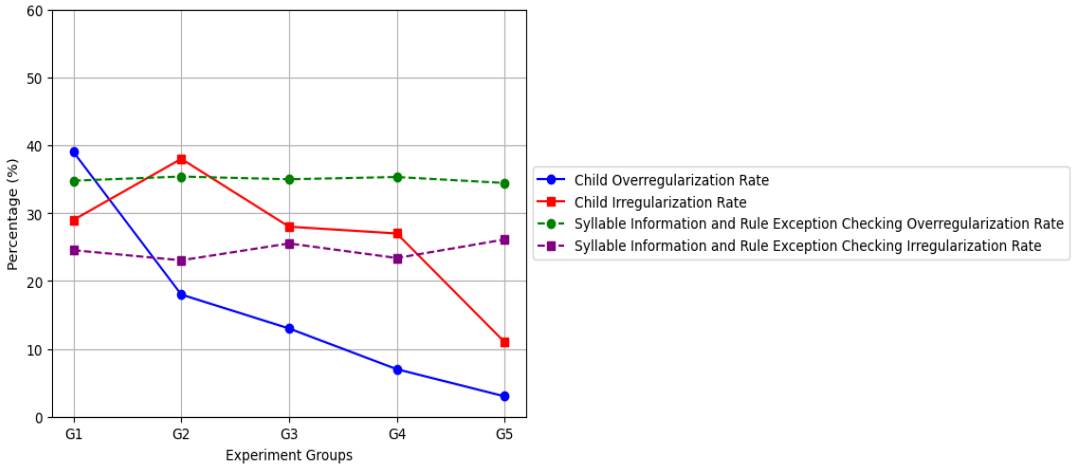

Figure 5. Child acquisition and syllable information and "rule exception checker" model regularization rates
(Nakipoğlu and Ketrez (2006, p.406)

For irregularization, there seem to be more fluctuations here than in the base—syllable information model, which could be because of the impact of the "rule exception checker" parameter, as it removes a rule from generalization if its context overlaps with that of an already generalized rule for a different morpheme. Again, similarly to the syllable information model, there is no similar behavior between this model and the child.

The "most common feature in context" parameter yielded results somewhat similar to the child acquisition experiment. Additionally, the model tended to overregularize in the simulation corresponding to G5, the oldest age group in the child dataset. Figure 6 shows the error rates of this model and the child acquisition experiment.

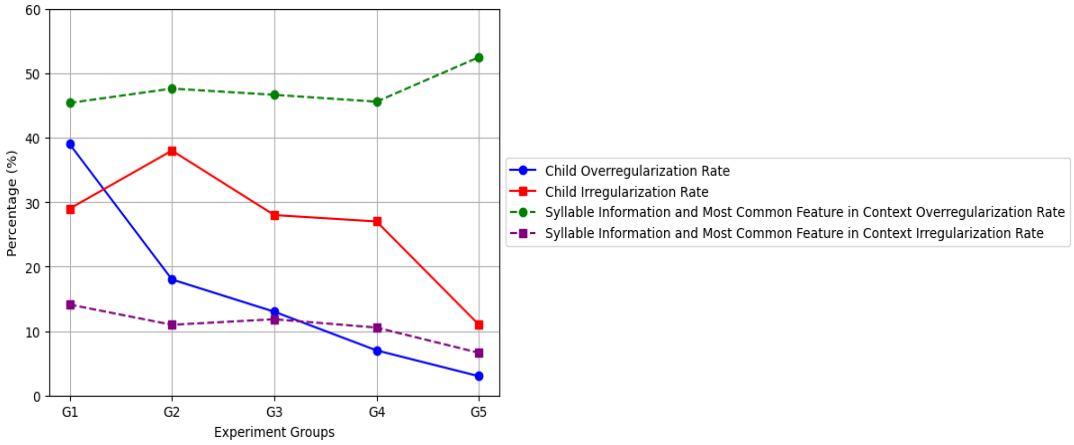

Figure 6. Child acquisition and syllable information and "most common feature in context" model regularization rates
(Nakipoğlu and Ketrez (2006, p.406)

This is interesting because not only does the model replicate the trend of decreasing overregularization rate throughout groups, but also there is a hike in the G5 which could not be explained by the skewness of the training dataset, because the dataset is skewed towards -Ir types; 650 -Ir-taking forms out of 1170 whereas only 305 -Ar-taking rots out of 1170. Because this parameter checks all the forms and constructs the context around the most common for that morpheme, it could be the case that the increase of the dataset could result in the change of the most common feature, which expands the context of the -Ar rule. This expansion could have resulted in a broader context that incorporated some irregular Ir taking forms to be overregularized towards Ar forms.

The models favored overregularization (Ar>Ir), which cannot be explained by the skewness of the training dataset. It seems that even though the training dataset is skewed towards -Ir-taking forms, the models were able to posit -Ar as the default for the monosyllabics which results in higher overregularization errors.

However, irregularization rates differ. In most models, the irregularization rate remains either too stable or overlaps with the overregularization behavior of the

child acquisition experiment. This could also be taken as a matching overregularization rate if the error labels are reversed. The "most common feature in context" and “no empty context” model best mimic the irregularization trend in the child acquisition experiment. The rest, however, exhibit either flat or inconsistent irregularization patterns across groups.

Overall, while these graphs suggest the models do not behave similarly but the "most common feature in context" and “no empty context” models were able to show similar behavior of overregularization error rates of the acquisition experiment, but this similarity is observed on the irregularization rates[24]. Though the p-values suggest this is not statistically significant, the previous chapter showed the “most common feature in context” parameter is able to perform better with bigger data and the same chapter also showed the “no empty context” parameter’s performance is closely related to the amount of data present for a generalized rule. The results of the previous chapter suggest these models might be performing better due to having a bigger dataset.

4.1.3 Experiment results of the token frequency manipulation

After seeing type frequency manipulation failing to fully capture the regularization rates of humans, token frequency manipulation is tested. The aim here is to see whether having a higher token count for the distributions could affect the behavior of the model. The roots in the training dataset stayed the same, but the forms are manipulated. So, numbers 1-4 were added to only the forms in the training dataset as a way to artificially increase the token count. Example (31) shows some rows from the

[24] The labeling of overregularization and irregularization errors is problematic due to conflicting assumptions between the acquisition data and the training data. Reversing this labelling for the models would show a decreasing trend in overregularization for the model —similar to what is observed in the acquisition data.

token frequency manipulated training dataset for G3 and Table 26 below shows Pearson's coefficient score (r) and p-values:

(31) Some rows from the training dataset for G3

{gel, geldi, gelir}, {gel, 1geldi, 1gelir}, {gel, 2geldi, 2gelir}

Table 26. Pearson's Coefficient Score and P-values for All Models for Token Frequency Manipulation

Model Name	Regularization Type	Pearson's Coefficient (r)	P-Value
Syllable	Overregularization	0.005	0.99
Information	Irregularization	0.158	0.79
No empty Context	Overregularization	-0.85	0.06
	Irregularization	0.40	0.50
Rule Exception	Overregularization	0.34	0.56
Checker	Irregularization	0.76	0.13
Most Common	Overregularization	0.21	0.72
Feature in context	Irregularization	0.81	0.09

Judging only by the information on Tables 25 and 26, it seems that manipulation of type frequency steers the model into behaving more like children in overregularization, whereas manipulation of token frequency steers the model towards behaving more like children in irregularization (Based on r-scores). Although these patterns are not statistically significant, the shift across models could suggest that the kind of frequency manipulation applied to the input data plays a role in the type of regularization behavior the model exhibits. Figure 7 below shows the error rates of the syllable information model and the child acquisition experiment.

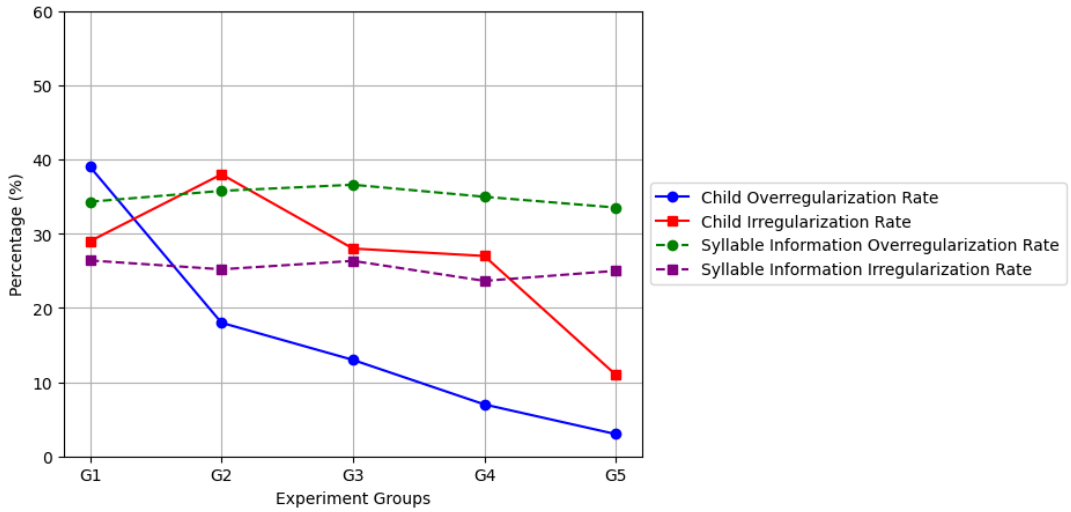

Figure 7. Child acquisition and syllable information model regularization rates on token frequency manipulation
(Nakipoğlu and Ketrez (2006, p.406)

Similar to the type frequency manipulation, the model's results do not fluctuate, but the increase of the irregularization rate in G5 suggests that the model shifts the default/elsewhere form from -Ar to -Ir. This could be because of the skewness of the training dataset.

The "no empty context" shows a similar situation to the type frequency manipulation experiment. This parameter is able to capture the first-increase-then-decrease pattern seen in the irregularization rate in the child experiment, but it captured on the overregularization, suggesting the model might be changing the default/elsewhere from between G1 and G2. Figure 8 below shows the regularization error rates for no empty context parameter and the child acquisition experiment.

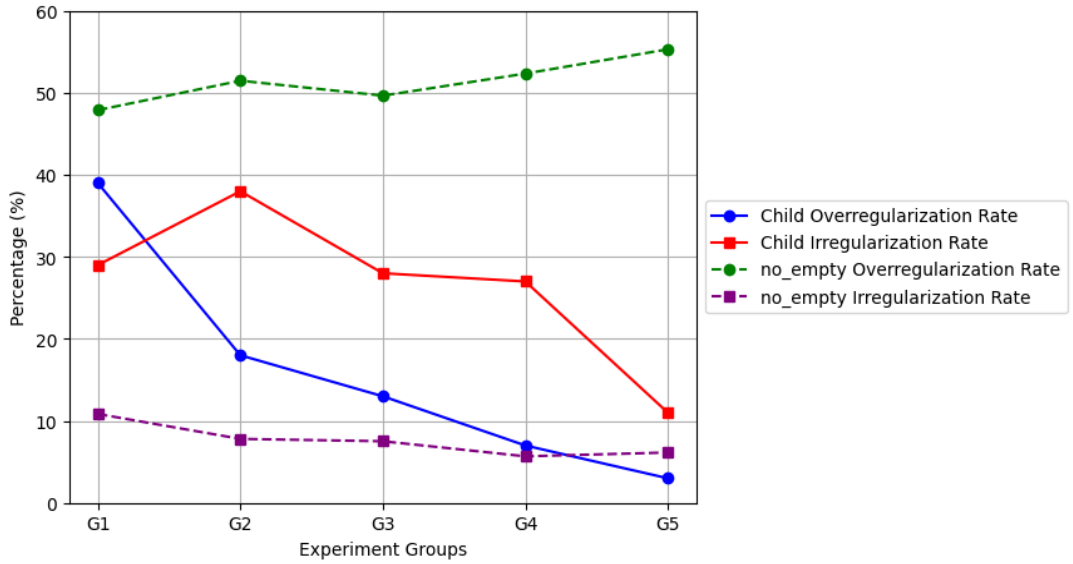

Figure 8. Child acquisition and syllable information and "no empty context" model regularization rates on token frequency manipulation
(Nakipoğlu and Ketrez (2006, p.406)

For the "rule exception checker" parameter, the results look similar to the syllable information but conversely to the type frequency manipulation. Figure 9 below shows the regularization error rates for the syllable information model and the child acquisition experiment.

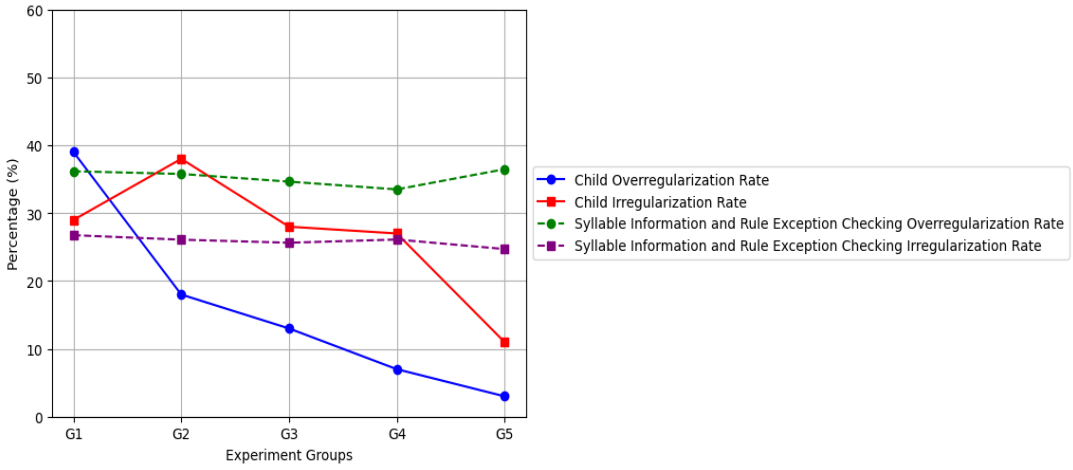

Figure 9. Child acquisition and syllable information and "rule exception checker" model regularization rates on token frequency manipulation
(Nakipoğlu and Ketrez (2006, p.406)

What is noteworthy here is that the overregularization rate (Ar>Ir) is steadily dropping along with the irregularization rate; however, in G5, there is a stark increase in overregularization rates, but a slight decrease in irregularization rates. This could be the effect of the "rule exception parameter". Since this parameter works on the basis of removing a rule if its context overlaps with that of a generalized rule, a root could be marked as an exception, which removes every other occurrence of that root as an exception, which could affect the generalization of these rules. Figure 10 below shows the regularization error rates for syllable information with the "most common feature in context" parameter and the child acquisition experiment.

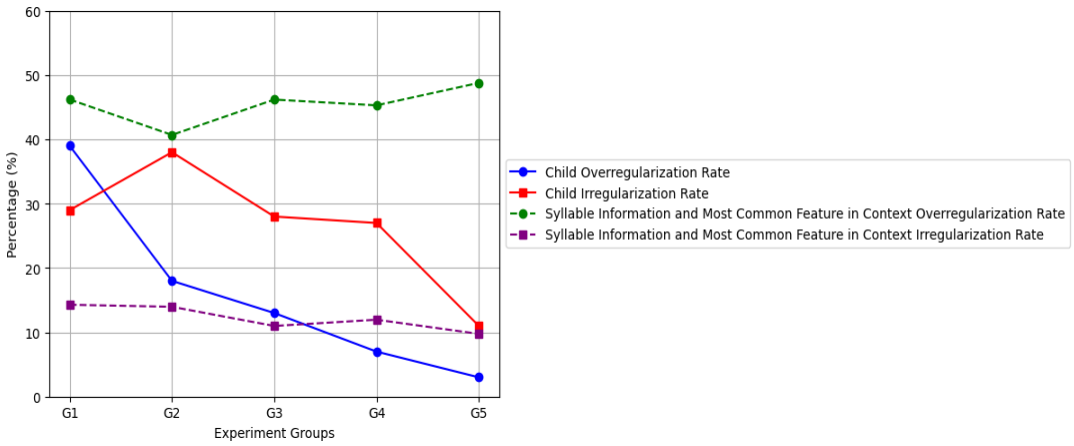

Figure 10. Child acquisition and syllable information and "most common feature in context" model regularization rates on token frequency manipulation (Nakipoğlu and Ketrez (2006, p.406)

This model slowly prefers to use Ar>Ir as the training dataset increases. This suggests the model slowly shifts from favoring -Ir to favoring -Ar for monosyllabics.

It should also be noted that the added parameters, which helped the model learn all rule-governed forms, are not implemented based on the developmental patterns observed in children. Rather, these parameters are implemented to solve the

models' failure to segment the forms in Turkish. Though these parameters are not based on patterns observed in children, the models started to have error rates similar to a child in rule-governed forms and monosyllabic roots ending with a non-sonorant consonant. Furthermore, when additional parameters that aim to have more specific context information, the drop in irregularization rates for monosyllabic roots ending in sonorants are observed. These suggest that even though the parameters are not implemented based on developmental evidence, the results suggest that they might translate into developmentally based parameters.

Overall, the results of the models cannot be argued to show strong correlation with the developmental trajectory outlined in Nakipoğlu and Ketrez (2006). The Baseline learner, which stores no phonological context, invariably applies -Ir. Although the model correctly extracted morphemes, the system's context construction showed the empty context made the model overapply the rule with the highest confidence. Introducing syllable structure solved the harmony problem and made the model correctly produce the multi-syllabic and vowel ending roots, all of which are rule-governed. Yet there was an abundance of errors with monosyllabic roots.

Since syllable information failed to capture the distribution of the monosyllabic roots, two parameters aimed to narrow the context have been implemented. A pivotal change occurs once every rule is required to encode at least one shared phonological feature. The "no empty context" and "most common feature in context" models were able to capture monosyllabic distribution better than the syllable information model. Furthermore, the advantages of having at least one {feature: value} pair for each context, show a convergence with the learning trajectory. These models posit -Ar as the default for monosyllabic verbs, despite the

global frequency bias favouring -Ir. Moreover, the “no empty context” model showed the first-increase-then decrease pattern in irregularization error rate. As for the “most common feature in context” model, the increase on overregularization rate and the decrease on irregularization rate for G5 (See Figures 6 and 10) suggest the reflection of the patterns seen across the distribution which further shows the skew of the training data does not undermine the matching.

Attempts to minimize the bleeding effect of the contrasting rules by pruning rules that conflict with other competitors does not yield marginal gains. There seems to be no perceived effect of this parameter. This suggests that the useful generalizations are sometimes discarded alongside spurious ones, and overall stability suffers.

Two broader conclusions follow. First, context specificity, rather than sheer rule count, is the decisive factor in preventing runaway generalizations: once a minimal but phonologically meaningful anchor is in place, the learner can negotiate the trade-off between default rules in a manner broadly comparable to that observed in children. Second, input distribution exerts a powerful influence on the shape of the learning path. When the training set is re-balanced, the learner readily flips its default suffix, confirming that the observed preference for -Ar in certain models is not just an artefact of the algorithm.

Together, these findings show that while context-specific anchoring and corpus balance steer these models toward behaving child-like, the absence of corrective feedback leaves entrenched rules unchallenged. Models consistently overregularized lexically-specified -Ir forms, while also wrongfully irregularizing rule-governed -Ar forms.

4.2 Future work

These results show that the implementation of syllable information to the baseline model allows the model to learn rule-governed forms fully. The mistakes for this model come from the lexically conditioned forms and the broad context problem, where the context information fails to capture necessary features for models to predict a form. Additional parameters were implemented to solve this broad context problem, but it persisted. In these lexically conditioned allomorphic forms, the model shows some parallels to child acquisition data, where type frequency is the determining factor in selection. There are several features that could be incorporated or changed to explore what features impact the model the most and what features make the results in line with human behavior. These additions not only help make the model's performance better but also show what features are important to morphological rule learning.

Future work could benefit from a fuller comparison between the model's behavior on regular and irregular forms, rather than focusing solely on error patterns. Analyzing both correct and incorrect productions across form types would provide a more comprehensive understanding of the model's generalization behavior. This comparison would also help clarify whether the model systematically treats rule-governed and exception forms differently, or whether its output reflects broader distributional patterns in the input. While the current results highlight variability in error rates and lack strong statistical correlations, the model did succeed in producing rule-governed forms without error under certain parameter settings. This suggests that the core morphological regularities were captured, and future analyses could build on this by exploring the model's treatment of both irregular and regular forms in more depth.

Context narrowing solved some problems but adding reinforcement in place of corrective feedback for word generation, where the output would also have an influence on the selected rule, is likely to yield better results. Not only would reinforcement allow the model to reanalyze the rule, but it would also allow the model to posit exceptions. Hiller and Fernández (2016) explain that there is a long-term effect of corrective feedback. Though time cannot be implemented, the results can still be checked to see whether the error rates behave similarly to the children in later stages of acquisition data (Nakipoğlu & Ketrez, 2006). This could also provide a useful benchmark for checking the performance of the model. If corrective feedback leads to noticeably better results, it could provide a way to create computational models with better accuracy. Additionally, adding a way to get the URs (underlying representation) of the rules might yield better results on word generation, as the model would deal with fewer rules and have more features to decide. Finally, the current parameters, though not developmentally motivated, yield error patterns resembling those of children, future models could implement incorporating parameters based on the observed developmental patterns. Using such parameters would help both segmentation accuracy and the model's behavior plausibility.

As for the word generation side, the algorithm could be complemented by the same phonological information and constraints that are present in Turkish. This would allow word generation to produce forms better, as the current algorithm picks out the form based on the context information of the provided generalized rules.

CHAPTER 5

CONCLUSION

This study evaluates the performance of a supervised morphological learner model, adapted from Albright and Hayes (2002), in predicting the distribution of the Turkish aorist marker and compares its output to developmental patterns found in child language acquisition. By applying the model to the aorist form in Turkish, a language characterized by its rich morphophonological interactions and lexically conditioned allomorphy, this research aimed to examine the extent to which a morphological learner could account for both rule-governed and irregular morphological patterns, as well as the extent to which the model could approximate human-like behavior.

The results show that while the baseline model was able to generate rules with high reliability and confidence scores, it failed to capture the correct surface forms for roots; instead, it picked the rule with the highest confidence score. The implementation of the syllable information parameter significantly improved the model's ability to generate forms for fully regular, rule-governed, roots, suggesting that syllable information provides a helpful structural cue in morphological learning. However, the model continued to struggle with monosyllabic verbs ending in sonorants, which are the very forms that pose acquisition challenges for children.

To address these issues, three additional parameters were independently tested: the "no empty context" parameter, the "rule exception checker", and the "most common feature in context". Among these, the "no empty context" and "most common feature in context" parameters showed the most consistent improvement in the model's word generation performance. Notably, both exhibited trends that seemingly mirrored children's overregularization behaviors, although neither fully

captured the developmental shift observed in the acquisition data. This suggests that while these parameters enhance the model's ability to generalize, they may still lack the nuanced mechanisms required to mimic developmental patterns.

Further experiments manipulating type and token frequencies showed that input distributions influence model behavior. Token frequency manipulation captured the first-increase-then-decrease pattern in irregularization error rates in the child experiment in the models' overregularization errors. Overall, these patterns align with some findings in acquisition research, indicating that input skew can shape morphological generalizations. However, the lack of statistically significant correlations in most models suggests that frequency alone cannot account for the full trajectory of morphological development.

Overall, this study demonstrates that rule-based learners, when extended with linguistically motivated parameters, can approximate certain aspects of child morphological behavior. At the same time, the limitations encountered, particularly with forms exhibiting root-conditioned allomorphy, highlight the challenges of modelling acquisition in the presence of irregularities. The findings suggest that future work could benefit from integrating additional learning mechanisms, such as corrective feedback or reimagining the rules' weight, and having a bigger longitudinal dataset.

In conclusion, while this research cannot adequately claim that the supervised learner fully replicates human morphological acquisition, it showed the models can behave similarly in certain ways when linguistically motivated parameters are used and showcased the implementation of syllable information could capture fully rule-governed forms in a language with vowel harmony.

REFERENCES

Ahyad, H., & Becker, M. (2020). Vowel unpredictability in Hijazi Arabic monosyllabic verbs. *Glossa: A Journal of General Linguistics*, *5*(1). https://doi.org/10.5334/gjgl.814

Albright, A. (2002). Islands of reliability for regular morphology: Evidence from Italian. *Language*, *78*(4), 684–709. https://doi.org/10.1353/lan.2003.0002

Albright, A., & Hayes, B. (2002). Modeling English past tense intuitions with minimal generalization. *In Proceedings of the ACL-02 Workshop on Morphological and Phonological Learning.* (pp. 58–69). https://doi.org/10.3115/1118647.1118654

Albright, A., & Hayes, B. (2006). Modelling productivity with the gradual learning algorithm. In G. Fanselow, C. Féry, M. Schlesewsky, & R. Vogel (Eds.), *Gradience in grammar* (pp. 185–204). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199274796.003.0010

Belth, C., Payne, S., Beser, D., Kodner, J., & Yang, C. (2021). *The greedy and recursive search for morphological productivity*. arXiv. https://doi.org/10.48550/arXiv.2105.05790

Bilgin, O. (2016). *Frequency effects in the processing of morphologically complex Turkish words*. [Master's Thesis, Boğaziçi University]. doi.org/10.13140/RG.2.2.20856.44809

Hiller, S., & Fernandez, R. (2016). A data-driven investigation of corrective feedback on subject omission errors in first language acquisition. *In Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning.* 105–114. https://doi.org/10.18653/v1/K16-1011

Jun, J., & Albright, A. (2017). Speakers' knowledge of alternations is asymmetrical: Evidence from Seoul Korean verb paradigms. *Journal of Linguistics*, *53*(3), 567–611. https://doi.org/10.1017/s0022226716000293

Kabak, B. (2011). Turkish vowel harmony. In M. Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (1–24). Wiley. https://doi.org/10.1002/9781444335262.wbctp0118

Kapatsinski, V. (2010). Velar palatalization in Russian and artificial grammar: Constraints on models of morphophonology. *Laboratory Phonology*, *1*(2), 361-393. https://doi.org/10.1515/labphon.2010.019

Kirov, C., & Cotterell, R. (2018). Recurrent neural networks in linguistic theory: Revisiting Pinker and Prince (1988) and the past tense debate. *Transactions of the Association for Computational Linguistics*, *6*, 651–665. https://doi.org/10.1162/tacl_a_00247

Kuo, J. (2020). *Evidence for base-driven alternation in Tgdaya Seediq* [Master's Thesis, UCLA]. https://escholarship.org/uc/item/8gv0c68g

McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology, 27*(4), 1546–1571. https://doi.org/10.1044/2018_AJSLP-17-0100

Mikheev, A. (1997). Automatic rule induction for unknown-word guessing. *Computational Linguistics*, *23*(3), 405–423.

Nakipoğlu, M., & Ketrez, F. (2006, January). *Children's overregularizations and irregularizations of the Turkish aorist*. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th annual Boston University Conference on Language Development.* (399–410). Cascadilla Press.

Nakipoğlu, M., & Michon, E. (2020). Abstraction vs. analogy in the Turkish aorist. In A. Gürer, D. Uygun-Gökmen, & B. Öztürk (Eds.), *Studies in Language Companion Series* (pp. 14–38). https://doi.org/10.1075/slcs.215.01nak

Nakisa, R., Plunkett, K., & Hahn, U. (2000). Single- and dual-route models of inflectional morphology. In P. Broeder & J. Murre (Eds.), *Models of language acquisition* (pp. 201–222). https://doi.org/10.1093/oso/9780198299899.003.0010

Odden, D. (2013). *Introducing phonology* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139381727

Oseki, Y., Sudo, Y., Sakai, H., & Marantz, A. (2019). Inverting and modeling morphological inflection. *Proceedings of the 16th workshop on computational research in phonetics, phonology, and morphology* (pp. 170–177). https://doi.org/10.18653/v1/w19-4220

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*, 73–193. https://doi.org/10.1016/0010-0277(88)90032-7

Ravid, D. (2019, December 23). First-language acquisition of morphology. In *Oxford research encyclopedia of linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.603

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1986). *Parallel distributed processing, volume 1: Explorations in the microstructure of cognition: Foundations*. MIT Press. https://doi.org/10.7551/mitpress/5236.001.0001

Sak, H., Güngör, T., & Saraçlar, M. (2008). Turkish language resources: Morphological parser, morphological disambiguator and web corpus. In B. Nordström & A. Ranta (Eds.), *Advances in natural language processing (GoTAL 2008)* (pp. 417–427). Springer. https://doi.org/10.1007/978-3-540-85287-2_40

Sak, H., Güngör, T., & Saraçlar, M. (2010). Web corpus. Boğaziçi University Turkish Language Resources Portal (TULAP). https://tulap.cmpe.boun.edu.tr/handle/20.500.12913/68

Strik, O. (2014). Explaining tense marking changes in Swedish verbs: An application of two analogical computer models. *Journal of Historical Linguistics*, *4*(2), 192–231. https://doi.org/10.1075/jhl.4.2.02str

Veríssimo, J., & Clahsen, H. (2014). Variables and similarity in linguistic generalization: Evidence from inflectional classes in Portuguese. *Journal of Memory and Language*, *76*, 61–79. https://doi.org/10.1016/j.jml.2014.06.001