VELİR OR VELER? EXPLORING ADULTS' IMPLICIT KNOWLEDGE OF AN IRREGULAR VERBAL PATTERN: THE TURKISH AORIST

ELISE ANNE CONSTANCE MICHON

BOĞAZİÇİ UNIVERSITY

2017

VELİR OR VELER? EXPLORING ADULTS' IMPLICIT KNOWLEDGE OF AN IRREGULAR VERBAL PATTERN: 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

Cognitive Science

by

Elise Anne Constance Michon

Boğaziçi University

2017

Velir or Veler? Exploring Adults' Implicit Knowledge of an Irregular Verbal Pattern: The Turkish Aorist

The thesis of Elise Anne Constance Michon

has been approved by:

Assoc. Prof. Mine Nakipoğlu
(Thesis Advisor)

Assoc. Prof. Arzucan Özgür
(Thesis Co-Advisor)

Assoc. Prof. Albert Ali Salah

Assist. Prof. Pavel Logačev

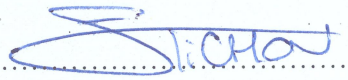
Assist. Prof. Serkan Şener
(External Member)

June 2017

DECLARATION OF ORIGINALITY

I, Elise Anne Constance Michon, 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 13/07/2017

ABSTRACT

Velir or Veler? Exploring Adults' Implicit Knowledge of

an Irregular Verbal Pattern: The Turkish Aorist

Morphological systems that possess irregular suffixation provide us important cues with respect to how words are processed and represented in the mind. This study attempts to find out how adult speakers of Turkish process and represent the aorist, an irregular pattern proven to be extremely challenging in acquisition, as it takes the form of -r for vowel-ending verbs (*uyu-r* 'sleeps'), -Ar for most monosyllabic verbs (*çal-ar* 'plays'), and -Ir for multisyllabic verbs (*çalış-ır* 'works') and 13 monosyllabic sonorant-ending verbs (*al-ır* 'takes'). In this study we have tested 90 native speakers with respect to their production of the aorist with 168 monosyllabic, sonorant-ending nonce roots (*vel-*, *rur-*, etc.). Participants were auditorily presented with sentences that contained nonce roots and instructed to orally conjugate the nonce root in a subsequent carrier sentence such as in *Ali biraz önce veldi. Ali yazın her gece vel____*. 'Ali velled a short while ago. Every summer night, Ali vel___.' Our results have shown that adult speakers of Turkish have opted for the use of -Ar with a rate of 86% (i.e., *vel-er* preferred over *vel-ir* by the overwhelming majority). To uncover the role of similarity in the use of –Ir, we have explored various similarity measures and have found that -Ir use appears to be correlated with the type frequency of existing irregulars sharing the same first consonant C_ _, the same first vowel _*V*_ and the consonantal template C_C. Rhyme, however did not appear to be a salient cue that language users rely on.

ÖZET

Velir veya Veler? Düzensizlik Gösteren bir Örüntüden Yetişkinlerin Örtük Bilgisini Ortaya Çıkarmak: Türkçe'de Geniş Zaman

Dağılımlarında düzensizlik gösteren biçimbirimsel sistemler, sözcüklerin zihinde nasıl işlendiği ve temsil edildiğine ilişkin önemli ipuçları sunar. Bu çalışma, eylem köküne eklenen eklerde düzensiz bir dağılım gösteren Türkçe Geniş Zaman ekinin ana dili Türkçe olan yetişkinler tarafından nasıl işlendiğini ve temsil edildiğini araştırmaktadır. Ünlüyle biten kökler için -r (*uyu-r*), ünsüzle biten tek heceli köklerin çoğu için -Ar (*çal-ar*), çok heceli kökler (*çalış-ır*) ve tınılı sesle biten 13 tek heceli kök için -Ir (*al-ır*) alan Geniş Zaman yapısının ediniminde yaşanan zorluklar bilinmektedir. Bu çalışmada anadili Türkçe olan 90 yetişkin katılımcının, Türkçe'de var olmayan, tınılı sesle biten 168 tek heceli kök ile (*vel-*, *rur-*, vb.) Geniş Zaman yapısı üretimi test edilmiş, katılımcılara, uydurulmuş bir eylem içeren tümceler dinletilmiş ve izleyen taşıyıcı tümcede bu eylem kökünü sesli olarak tamamlamaları istenmiştir (*Ali biraz önce veldi. Ali yazın her gece vel____* gibi). Bulgularımız, yetişkinlerin söz konusu eylemlerle -Ar ekini %86 oranında kullandığını göstermiştir (dolayısıyla, büyük çoğunluk *vel-ir* yerine *vel-er*'i tercih etmiştir). -Ir kullanımında, var olan eylemlerle benzerliklerin rol oynayıp oynamadığını ortaya çıkarmak için çeşitli benzerlik ölçütleri geliştirilmiş, dilde mevcut olan aynı ilk ünsüz C_ _, aynı ünlü _V_ ve aynı ünsüz örüntüsü C_C'yi paylaşan düzensizlik gösteren eylemlerin kullanım sıklığıyla –Ir kullanımının ilişkili olduğu, uyak ölçütünün, ise dil kullanıcıları açısından güvenilir bir ipucu oluşturmadığı gözlenmiştir.

ACKNOWLEDGMENTS

I would like to express my deep thanks and gratitude to all the people who took part in this research project. Namely:

Mine Nakipoğlu, for overcoming differences and delving into this research question with me. For giving me the opportunity, throughout all that transpired, to work on a multidisciplinary ground with a common language. You made this project possible by guiding me into linguistics, following the project closely season after season, taking the time to look at details, teaching me Turkish proverbs, deciphering dozens of tables full of figures, rewording my flowery English, always letting me explore other possibilities but not lose track of the main purposes of the study.

Arzucan Özgür, Pavel Logačev, Albert Ali Salah, Serkan Şener, for supporting us in making decisive choices at different steps of the project and providing insightful comments from their various perspectives.

The NLP team, for sharing progress of our projects week after week during winter months and advising me to follow Code Academy online tutorials to program better. In particular Orhan Bilgin, Çağıl Sönmez, and the user of the forum StackOverFlow named “lluvatar” for searching together a method compatible with the size of the corpus.

All the members of the Linguistics department, for welcoming the experiment stage into their space and enduring the many participants and cakes in the corridor.

Simge Topaloğlu and Noyan Dokudan, for recording one hundred stimuli so quickly.

Pınar Aldan, for helping me write so many things in proper Turkish, for listening to so many made-up Turkish sentences and for letting fun pierce the intense work periods.

All the students who participated in the experiment and kindly spent 20 minutes uttering non-existing-but-could-have-been Turkish verb forms.

And because a thesis would not be the same without a wide network of supporters far from university, my thanks go out also to:

The team of French Institute for Anatolian Research (IFEA), for giving me the chance to work on an interesting project and making it compatible with the load of a master thesis. Especially Anouck, Céline and Hülya for your sweet (in all senses) encouragements during writing.

Stanislas and Victor, for giving me a stable opportunity to teach and for the chance to see, week after week, as you gained self-confidence in mathematics and in the future, Ali, Satılmış, the brothers at Yaprak lahmacun, the family at Uzunçayır pilavcısı, and the personnel of Güney yemekhane, for your daily smiles in the campus or neighborhood and lovely help in sickness.

Shannon, Sima, Asmar, Amer and Nour, for making our home a kind of free oasis, even during crazy times for each of us, for laughing at my English and for preventing me from starving with innumerable oriental, zucchini and cheese meals.

İdil, Sinan, Semih, Elif, Gürcan, Burhan, Duygu, Emel, Zafer, Cem, Murat, Nurhan and others, for being my family in Istanbul (Türkiye şubesi).

Claudio, Hermine, Nino, Seda, Yvan, Elie and Mariam, for being heartfelt friends even from far away, always being touched and often in touch, endlessly in research.

And to my family, for your unconditional love and support wherever I am and whether you understand or not what is going on for me.

To God, present as always and everywhere, reminding us to do things and to "love with all your heart, and with all your soul, and with all your strength, and with all your mind; and your neighbor as yourself."

"When patterns are broken, new worlds emerge."

Tuli Kupferberg

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: DESCRIPTION OF THE TURKISH AORIST 7

2.1 Morphological properties of the Turkish aorist 7

2.2 Quantitative description: Frequencies in a web corpus 10

CHAPTER 3: THEORETICAL BACKGROUND 20

3.1 Connectionist and rule-based models 20

3.2 Related psycholinguistics experiments 25

3.3 Frequency considerations 28

CHAPTER 4: METHODOLOGY 30

4.1 Stimuli 30

4.2 Fillers 34

4.3 Carrier sentences 35

4.4 Participants 37

4.5 Procedure 38

CHAPTER 5: RESULTS 42

5.1 Default status: Is there a rule in the production of the aorist? 42

5.2 Irregularization errors: Is rhyme a relevant measure of similarity? 45

5.3 Alternative similarity measures: What makes a nonce more irregularized? 51

5.4 Summary 64

CHAPTER 6: DISCUSSION AND CONCLUSIONS 68

APPENDIX A: METHODOLOGY FOR FREQUENCY COUNTS 74

APPENDIX B: FREQUENCIES OF MONOSYLLABIC VERBS IN TURKISH 77

APPENDIX C: CONSONANT AND VOWEL PATTERN FREQUENCIES IN TURKISH 83

APPENDIX D: NONCE VERBS WITH CONTEXT AND CARRIER SENTENCES 93
APPENDIX E: INSTRUCTIONS DISPLAYED ON THE SCREEN DURING THE EXPERIMENT 100
APPENDIX F: ISLANDS OF RELIABILITY 102
REFERENCES 105

LIST OF TABLES

Table 1. Type and Token Frequency Counts according to Affixes Proceeding the Aorist 12

Table 2. Type and Token Frequency Counts according to Affixes Preceding the Aorist 13

Table 3. Examples of Aorist Forms Falling in Different Categories of Verb Root and Suffixes 15

Table 4. Type and Token Frequency Counts of the Suffixes -Ar and -Ir in the Aorist according to Verb Root Characteristics 15

Table 5. The 336 Possible CVS Syllables (Color Code Explained in the Text) 31

Table 6. Existing Monosyllabic Regular (Green) and Irregular (Red) Verbs in the Possible Syllables 32

Table 7. 168 Nonce Monosyllabic Verbs Chosen for the Study 34

Table 8. Breakdown of Counts and Rates of Aorist and Other Forms in Conditions A and B 41

Table 9. –Ir Use for the 168 Nonce Verbs 42

Table 10. Nonce Roots Never Produced with –Ir by the Participants 44

Table 11. Six Most Irregularized Nonce Verbs with the Number of -Ar and -Ir Answers 44

Table 12. –Ir Use according to the Rhyme of the Nonce Root 46

Table 13. Rhymes of Nonce Stems and their –Ir Use according to Type and Token Frequencies of Regular and Irregular Monosyllabics in Turkish 46

Table 14. Frequencies of Regular and Irregular Monosyllabics Rhyming in UL 47

Table 15. Frequencies of Regular and Irregular Monosyllabics Rhyming in ÜR, IR, İR, OR, IL, ÜL 48
Table 16. Frequencies of Regular and Irregular Monosyllabics Rhyming in UR, ÖR, ÖL 49
Table 17. Frequencies of Regular and Irregular Monosyllabics Rhyming in EL, AL, ER, OL, AR, İL 50
Table 18. –Ir Use according to the Final Consonant of the Nonce 52
Table 19. Frequencies of /r/ and /l/-ending Monosyllabic Verbs in Turkish 52
Table 20. Type and Token Counts of Trigram Patterns Corresponding to the Regular and Irregular Aorist Forms 53
Table 21. Type and Token Counts of Quadrigram Patterns Corresponding to the Regular and Irregular Aorist Forms 54
Table 22. -Ir Use according to the Vowel in the Nonce Root 55
Table 23. -Ir Use according to Backness/Frontness of the Vowel in the Nonce Root 56
Table 24. -Ir Use according to the Roundedness/Height of the Vowel in the Nonce Root 56
Table 25. Frequencies of Regular and Irregular Monosyllabics Sharing the Same Vowel 57
Table 26. Type and Token Counts of the Patterns V_V Corresponding to the Regular and Irregular Aorist Forms 58
Table 27. -Ir Use according to the First Consonant in the Nonce Root 59
Table 28. Frequencies of Regular and Irregular Monosyllabics Sharing the Same First Consonant 60
Table 29. –Ir Use according to the C_C Pattern in the Nonce Stem 62

Table 30. Frequencies of Regular and Irregular Monosyllabics Sharing the Same Pattern C_C 64

Table 31. –Ir Use according to Rhyme and First Consonant 65

Table 32. –Ir Use according to Pattern C_C and Vowel 66

LIST OF APPENDIX TABLES

Table B1. Token Counts of Aorist Forms of Monosyllabic Verbs in the BOUN Corpus (Irregulars in Grey) 77

Table C1. Type and Token Frequencies of the "Vowel-Consonant-Vowel" Patterns in the Entire Language, in the Verb Forms and in the Aorist Forms 83

Table C2. Type and Token Frequencies of the "Vowel-Sonorant-Harmonic Vowel" Patterns (where Sonorant Stands for /l/, /n/ or /r/) in the Entire Language, in the Verb Forms and in the Aorist Forms 85

Table C3. Type and Token Frequencies of the "Vowel-Sonorant#Harmonic Vowel" Patterns (where Sonorant Stands for /l/, /n/ or /r/) at a Word Boundary in the Entire Language, in the Verb Forms and in the Aorist Forms 87

Table C4. Type and Token Frequencies of the "Vowel-Sonorant-Harmonic Vowel-r" Patterns (where Sonorant Stands for /l/, /n/ or /r/) at Any Place in a Word in the Entire Language, in the Verb Forms and in the Aorist Forms 89

Table C5. Type and Token Frequencies of the "Vowel-Sonorant-Harmonic Vowel-r" Patterns (where Sonorant Stands for /l/, /n/ or /r/) in Final Position in a Word in the Entire Language, in the Verb Forms and in the Aorist Forms 91

Table D1. Nonce Monosyllabics Used as Stimuli in Condition A 93

Table D2. Nonce Monosyllabics Used as Stimuli in Condition B 96

Table D3. Nonce Multisyllabics Used as Fillers in Conditions A and B 99

LIST OF FIGURES

Figure 1. Distribution of the aorist in type and token with respect to the proceeding affix 12

Figure 2. Distribution of the aorist in type and token with respect to the preceding affix 13

Figure 3. Distribution of the aorist in type with respect to the verb root characteristics 17

Figure 4. Distribution of the aorist in token with respect to the verb root characteristics 17

Figure 5. Distribution of the suffixes -Ar and -Ir in type for consonant-ending verbs with respect to the verb root characteristics 18

Figure 6. Distribution of the suffixes -Ar and -Ir in token for consonant-ending verbs with respect to the verb root characteristics 18

CHAPTER 1

INTRODUCTION

Children learn to speak their native language without adults explicitly telling them the rules of grammar. Instead, from the huge amount of linguistic stimuli they hear, they gradually acquire an implicit knowledge of their native tongue and build abstractions at the levels of phonology, morphology, syntax, etc. In particular, to attain competence in morphology the child has to dissociate between words/roots that can stand on their own and bound forms, i.e., affixes that convey a particular grammatical function. Acquisition of morphology, in particular affixes/morphemes, is known to be less demanding when there is one-to-one mapping between an affix and a function but quite challenging when there is one-to-many mappings i.e., when a particular affix has several functions such as the English –s (marking possessive as in *John's*, marking plural as in *cat-s* or marking the 3rd person present tense as in *walk-s*).

An equally challenging process in acquisition is when a language resorts to the use of various forms to convey one particular function. This is the case in the Turkish aorist, the so-called unbounded (present*)* tense marker. When attached to vowel-ending verbs, the affix for the aorist assumes the shape of –r *(yürü-r* 'walks'), when attached to multisyllabic verbs, its surfaces as –Ir (*çalış-ır* 'works'); when attached to monosyllabic verbs, the outcome is sometimes –Ar (*çal-ar* 'plays') and at other times, –Ir (*al-ır* 'takes'). The variant behavior of the Turkish aorist presents a tremendous challenge for Turkish-speaking children (Nakipoğlu and Ketrez, 2006). Acquisition data clearly reveals that a representation of the pattern regarding the distribution of the affix is being built in the mind of the child up until the age of 8.

This finding then suggests that adults base their daily use and processing of the affix on this representation which of course, is not consciously available. To understand better the nature of this knowledge representation in adults, and the processing mechanisms involved in the production of the aorist, in this study we aim to examine the aorist forms produced by adult participants when presented with non-existent verbal roots, i.e. nonce roots, in Turkish.

If participants of a psycholinguistic experiment are asked to generate conjugated forms of a verb that exists in their language, they would just produce the correct answer, except errors made under time pressure or distraction. Similarly if participants are presented with a nonce root and asked to conjugate it in a regularly behaving verbal pattern such as the future tense in Turkish, they can easily generalize the pattern to items that they have never heard before, hence a nonce such as *vül-* would be produced as *vül-ecek*. But if asked to conjugate the verb in a pattern for which, to convey a particular function, the language makes use of a number of suffixes, several outcomes become possible. Furthermore when the distribution of the affixes is not predictable on phonological grounds, i.e. when there is irregularity, the issue becomes even more challenging.

As is well known, the most famous case of such an irregular verbal pattern is the English past-tense, which divides the verbs in two groups: regulars that take one of the variants of the suffix -(e)d, in a phonologically predictable manner i.e., [əd] as in *load-loaded*, [t] as in *walk-walked,* or [d] as in *rub-rubbed*; and irregulars that undergo various phonological processes, such as change of the stem vowel (*sing-sang*), change of the rhyme (*bring-brought*), or even no change at all (*hit-hit*). As reported in Albright and Hayes (2003), for instance, English-speaking participants when tested on nonce verbs such as *spling* in the past tense, produced the regular

form *splinged* with a rate of 51%, and the irregular forms *splung* and *splang* have also sprung up with a rate of 32% and 10%, respectively.

In cases like this where several answers can be entertained, the outcomes cannot be the result of a single rule. Such irregular patterns, where some verbs behave regularly and others irregularly, require further explanation. In the literature, mostly based on the English past tense, it has been proposed that regular and irregular verbs can be accounted for either by a single mechanism or by two separate mechanisms.

The Single-mechanism model (Rumelhart and McClelland, 1986; Plunkett and Marchman, 1993; Hare, Elman and Daugherty, 1995; Hahn and Chater, 1998; Nakisa, Plunkett and Hahn, 2000, among many others) proposes that English speakers rely on similarities between verbs to compute the past tense of regular and irregular verbs alike. Under this view, all nonce verbs would be conjugated according to similar verbs encountered thus far. In contrast, in the so-called Dual-mechanism model (Pinker and Prince, 1988; Marcus, Pinker, Ullman, Hollander, Rosen and Xu, 1992; Marcus, 1998; Pinker, 2001), the irregular past tense forms are stored in the lexicon with the verb stems, while regular verbs require a simple symbolic rule that requires the attachment of a suffix to the stem. Under this view, resemblance with existing items would affect only the roots (i.e. nonce forms) similar to irregulars, and not those similar to regulars. The competition between those models, still ongoing, has become the topic of more than 100 publications, and has often been called “The English past tense debate”, a good review of which can be found in the paper by Seidenberg and Plaut (2014).

These two models appear to be influenced by the linguistic input previously heard by the native speaker. In previous studies from both sides (Bybee and Moder,

1983; Prasada and Pinker, 1993), stimuli have been carefully controlled for their resemblance with existing verbs: nonces were closely selected so that the number of rhyming regulars and irregulars was low or high. But the authors had chosen rhyme as a predefined measure of similarity among nonces, regular and irregular verbs, and did not test it against other measures while this could be critical as noted by Hahn and Bailey (2005).

When producing the past tense form of nonce verb roots, it has been additionally observed that the existing verbs and their past tense forms can be more easily recalled since they are frequent. For instance, the English nonce *spling* triggers irregular answers such as *splung* supposedly because *spling* is similar to a certain number of existing verbs having such a past form, *swing, string, wring, sting, sling, fling, cling,* varying in frequency in the language. For these existing verbs similar to *spling*, the different items are referred to as "types", while the occurrences of these verbs are called "tokens". Then verbs similar to a nonce can be frequent in type when there are a lot of them, or in token when they represent a lot of occurrences in the language. If participants are sensitive to the exemplars they hear, one intuitively expects the token frequency of the verbs similar to the nonce to have more influence on its past form. Yet literature on the issue has actually suggested that generalizations to nonce stems are made on the basis of type and not token frequency (Bybee, 1995; Baayen, 2001; Bybee and Hopper, 2001).

The present study aims at investigating the way adults make use of rules, similarity and type/token frequency in the processing of an irregular verbal pattern in their native language. We examined the Turkish aorist to answer the following research questions:

1. When conjugating a new/nonce item in a morphological pattern that also embodies irregularity:
    a. Do speakers apply a rule?
    b. Do speakers rely on similar existing verbs?
2. If speakers rely on similarity with existing items, on which measures of similarity do they base their answers?
3. Does type or token frequency of existing verbs influence participants' answers?

Given the previous findings for the English past tense, we hypothesize that:

1. Both rules and similarity would matter in handling a morphological pattern in one's native language:
    a. If in the production of the Turkish aorist participants apply a rule, we expect most of the nonce items to be produced with the regular form.
    b. If participants rely on similarity with existing verbs, we expect some nonce items to be produced with the irregular form.
2. Rhyme can be the measure participants would mostly rely on to evaluate if a nonce resembles existing verbs: we expect rhyme to be the most salient among several measures of similarity between the stimuli and existing irregulars, such as sharing the same first consonant, the same vowel, the same last consonant, the same first and last consonants (the C_C pattern), or the same vowel and last consonant (rhyme).
3. Frequency, in particular type frequency enhances the effect of similarity: we expect more irregular forms to be produced for a nonce similar to a higher number of different irregular verbs (according to the above relevant measures of similarity).

The thesis is structured as follows: Chapter 2 is devoted to the structural description of the Turkish aorist where we explain the grammar rule governing the pattern. In the same chapter we also examine the frequency distribution of the aorist forms in Turkish by exploring a large online Turkish corpus. In Chapter 3, we summarize the available processing models for the English past tense and the previous findings concerning similarity and frequency in elicited past forms for nonce items, which will give us a background to understand and interpret the findings we obtained in our experiment. Chapter 4 presents the procedure of the experiment we have run on 90 native speakers of Turkish to test the behavior of the Turkish aorist. The results of the experiment will be laid out in Chapter 5. Finally in Chapter 6 we will discuss what our findings have taught us about the adult representation of the Turkish aorist, and more generally irregular patterns in the mind.

CHAPTER 2

DESCRIPTION OF THE TURKISH AORIST

2.1 Morphological properties of the Turkish aorist

Used in the expression of generalizations, habits, hypotheses and volition, the Turkish aorist reveals itself in three variant affixes: -r, -Ar and -Ir. While the vowel-ending verbs are attached the affix –r as in (1), the affix that surfaces on the consonant-ending verbs is determined on the basis of whether the verb is monosyllabic or multisyllabic. In particular, multisyllabic verbs always surface with the affix -Ir i.e., [ır, ir, ur, ür] following the rules of Vowel Harmony in Turkish, as in (2).[1]

(1)	*uyu-*	'sleep'	*uyu-r*	's/he sleeps'
(2)	*kapat-*	'close'	*kapat-ır*	's/he closes'
	pişir-	'cook'	*pişir-ir*	's/he cooks'
	otur-	'sit'	*otur-ur*	's/he sits'
	düşün-	'think'	*düşün-ür*	's/he thinks'

The monosyllabic verbs, however, pose a certain challenge to the speakers of Turkish. Though most monosyllabic verbs take the affix -Ar as in (3), 13 verbs take the affix –Ir and are considered to be behaving irregularly as listed in (4).

[1] In Turkish the affix vowels are either I-type or A-type as in the Turkish aorist which surfaces as -Ir or -Ar. The I-type affixes follow the rules of high Vowel Harmony in Turkish and I stands for the corresponding high vowel.

(3)	*bak-*	'look'	*bak-ar*	's/he looks'
	seç-	'choose'	*seç-er*	's/he chooses'

(4)	*al-*	'take'	*al-ır*	's/he takes'
	bil-	'know'	*bil-ir*	's/he knows'
	bul-	'find'	*bul-ur*	's/he finds'
	dur-	'stop'	*dur-ur*	's/he stops'
	gel-	'come'	*gel-ir*	's/he comes'
	gör-	'see'	*gör-ür*	's/he sees'
	kal-	'stay'	*kal-ır*	's/he stays'
	ol-	'be'	*ol-ur*	's/he is'
	öl-	'die'	*öl-ür*	's/he dies'
	san-	'think'	*san-ır*	's/he thinks'
	var-	'arrive'	*var-ır*	's/he arrives'
	ver-	'give'	*ver-ir*	's/he gives'
	vur-	'hit'	*vur-ur*	's/he hits'

The irregularly behaving verbs are all monosyllabic roots ending in the sonorants /l/, /r/, or /n/. Furthermore they are highly frequent. Characterization of the irregularly behaving verbs on the basis of sonority is necessary but not sufficient, since not all sonorant-ending monosyllabic verbs fall under the -Ir category as can be observed in the examples given in (5) below.[2]

(5)	*öl-*	'die'	*öl-ür*	's/he dies'	(irregular)
	böl-	'divide'	*böl-er*	's/he divides'	(regular)
	gel-	'come'	*gel-ir*	's/he comes'	(irregular)
	del-	'drill'	*del-er*	's/he drills'	(regular)

Interestingly, the Turkish aorist pattern matches the English past tense pattern in that all verbs with multisyllabic roots are regular in the sense that they take the affix -Ir at all times and all irregulars are highly frequent just as in English (as reported in Pinker and Prince (1988)). Moreover, the final sounds of regular and irregular verbs

2 One could wonder how the so-called irregularity appeared in the Turkish aorist. Tekin (1995) (as cited in Kanık (2015)) states that "Originally the aorist suffix in Turkish was -r and the vowels before –r in the aorist suffix in Old and Middle Turkic are in fact a relic of the lost stem-final vowels in Proto-Turkic", and that "many monosyllabic Old Turkish verbs lost their older and perhaps original aorist forms in {-Ur} in the 11th century and took the new aorist forms in {-Ar}" [sic]. We also note that the verbs corresponding to the 13 irregular Turkish verbs in the other Turkic languages take a regular affix similar to -Ar. It is important to underline that our focus is not how aorist evolved in time in Turkish, but how it is processed by native speakers of Modern Turkish who have always been exposed to aorist with these 13 exceptionally behaving verbs.

can be exactly the same, both in English (*submit-submitted* is regular, but *hit-hit*, *sit-sat* are irregular) and in Turkish as observed in (5) above.

Nevertheless the two patterns cannot be considered as entirely similar: the irregular verbs in English depart in various ways from the regular -(e)d suffix, whereas the so-called irregulars in Turkish are monosyllabic but attached the affix that we see in multisyllabic verbs as well.

Acquisition of this intriguing distribution in the aorist has been studied by Nakipoğlu and Ketrez (2006). Using picture-prompted elicitation technique, the authors have tested 136 Turkish speaking children at various ages, and encountered a high rate of errors in children's productions mostly for sonorant-ending monosyllabic verbs. In particular sonorant-ending irregulars were found to be regularized yielding overregularization errors such as **bil-er* for *bil-ir 'knows'*, **al-ar* for *al-ır* 'takes' and regulars undergoing irregularization such as **bin-ir* for *bin-er* 'rides'**,** **çal-ır* for *çal-ar* 'plays/steals'. As for the rate of errors, 3-year-old children have been found to overregularize irregulars with a rate of 39% and irregularize regulars with a rate of 29%; 4-year-old children, however, make less overregularization errors, 18%, but more irregularization errors, 38%. The predominance of irregularizations over overregularizations persists in the following years till the disappearance of errors around 8 years of age. The reversal of the trend in errors interestingly parallels the development of verb vocabulary. At 3 years of age, the child encounters more aorist forms ending in -Ar (monosyllabic regulars and non-high-vowel-ending), as they represent up to 70% of aorist verbs in child-directed speech collected in the CHILDES Turkish corpora (Aksu-Koç and Slobin (1985) as cited in Nakipoğlu and Ketrez (2006)). With age, this proportion decreases as the children learn more verbs the majority of which are multisyllabic, so with aorist forms ending in -Ir.

2.2 Quantitative description: Frequencies in a web corpus

2.2.1 Methodology

Building on the insight that a corpus would reflect language use, we have extracted type (number of different verbs) and token (number of occurrences) frequency counts for the aorist from the BOUN Corpus (Sak, Güngör and Saraçlar, 2008), the biggest Turkish web corpus available. More particularly, we have worked on the version of this corpus made available by Zargan, an online Turkish Dictionary.[3]

The BOUN Corpus contains 1.337.898 word-forms and their stems, morphological analyses and frequencies. An example is given in (6) below:

(6) Word-form	Stem	Morphological analysis	Frequency
çöküntünün	Çöküntü	Noun+A3sg+Pnon+Gen[4]	309

We have performed several operations on the corpus to obtain the entries that we need for this study. Recall that the corpus initially consisted of 1.337.898 items. Here is what we have done to retrieve the aorist forms:

i. we have merged the lines concerning the same entity with and without apostrophe (which amounts to 1.260.882 items),
ii. we have selected the verb forms (748.051 items (59.33% of the lexicon)),
iii. then we have selected the aorist forms (only used in the affirmative contexts) [5] (44.972 items (6.01% of the verb forms)),
iv. we have examined the distribution of forms according to the suffixes occurring before and after the aorist,

3 http://st2.zargan.com/duyuru/Zargan_Linguistic_Resources_for_Turkish.html.
4 A3sg: Agreement 3rd person singular, Pnon: no Possessive Marker, Gen: Genitive
5 In negated contexts the Turkish aorist either appears as a null morpheme in 1st person singular and plural as in *git-me-m* 'I do not go' / *git-me-yiz* 'We do not go', or it surfaces in the form of –z with other person markers such as *git-me-z-sin(iz)*, 'You sg(pl) do not go' or *git-me-z-ler* 'They do not go'. For the purposes of this study we have restricted ourselves only with the aorist used in affirmative contexts.

v. finally we have created a compiled list of the aorist forms once the suffixes after the aorist (including person markers) have been stripped off. Hence, for instance *çal-abil-ir-dim, çal-abil-ir-iz, çal-abil-ir-cesine,* etc. are all compiled under the entry *çalabilir*.

A more detailed account of the methodology used in this study can be found in Appendix A.

2.2.2 Distribution of the aorist forms with respect to the suffixes preceding and/or proceeding the aorist

Examining the suffixes that proceed and in particular, precedes the Turkish aorist is crucial for our study as the variant of the aorist can differ with respect to the syllable size of the verb root. For instance, the monosyllabic *çal-* 'play/steal' takes the -Ar form yielding *çal-ar* in the aorist, but the attachment of the affix -Abil renders a multisyllabic stem, hence *çal-abil-* 'be able to play/steal' takes -Ir resulting in *çalabil-ir*. So a look at the distribution of the suffixes coming before the aorist is essential for our analysis since it directly increases the rate of verbs produced with -Ir. Table 1 and Table 2 below present a breakdown of the Turkish aorist forms and frequencies with respect to the proceeding and preceding affixes.

Table 1. Type and Token Frequency Counts according to Affixes Proceeding the Aorist

Suffixes after the aorist	Example on verb *çal-*	Type	Token
aorist + past	*çal-ar-dı*	7944	207412
aorist + cond	*çal-ar-sa*	8440	504001
aorist + ken	*çal-ar-ken*	3998	842814
aorist + mIş	*çal-ar-mış*	2046	17455
aorist + CAsInA	*çal-ar-casına*	644	7244
aorist + other	*çal-ar-SUFFIX*	181	1146
aorist + none	*çal-ar*	21719	3357211
Total		44972	4937283

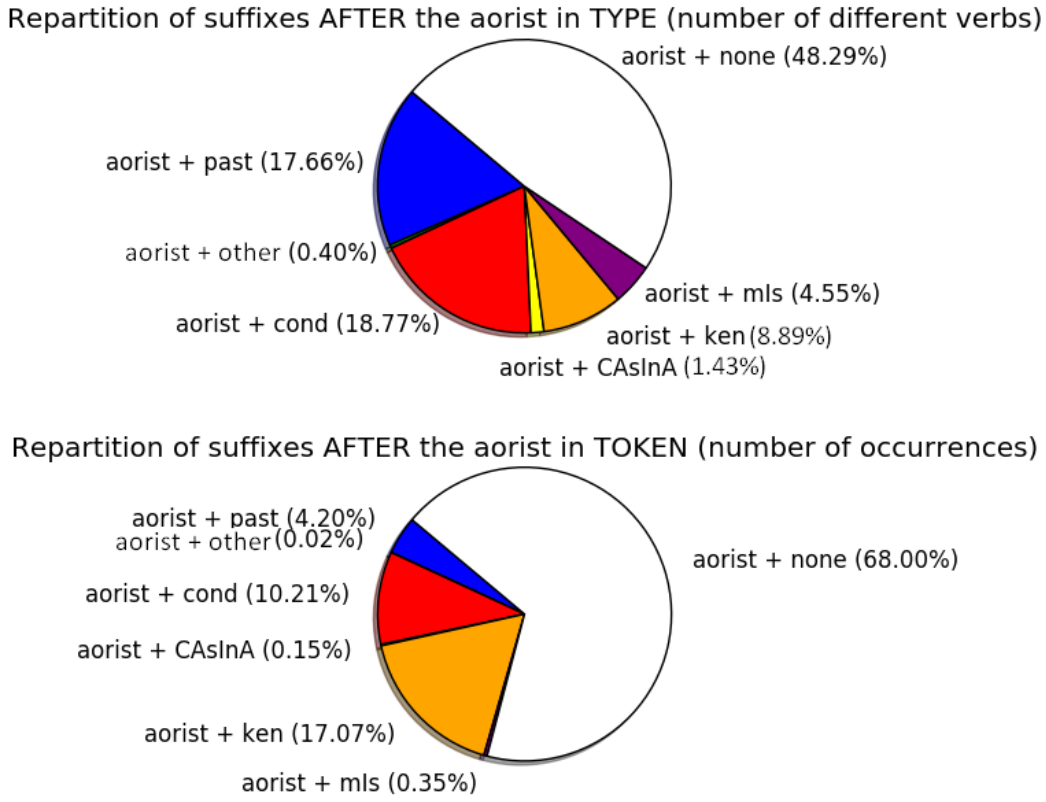


Figure 1. Distribution of the aorist in type and token with respect to the proceeding affix

Table 2. Type and Token Frequency Counts according to Affixes Preceding the Aorist

Suffixes before the aorist	Examples	Type	Token
Abil + aorist	*çal-abil-ir*	13583	925726
Iver + aorist	*çal-ıver-ir*	880	3111
passive + aorist	*çal-ın-ır*	6736	677169
causative + aorist	*çal-dır-ır*	6171	198192
none + aorist	*çal-ar*	17602	3133085
Total		44972	4937283

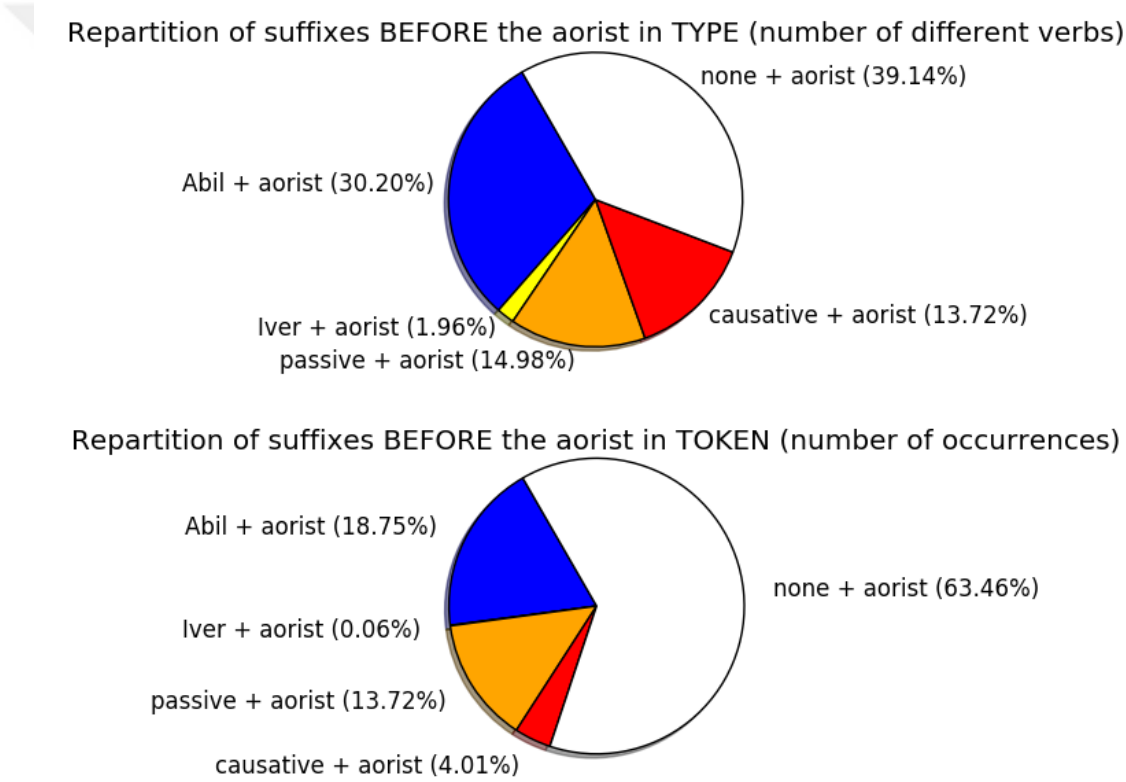


Figure 2. Distribution of the aorist in type and token with respect to the preceding affix

The token frequencies of monosyllabic verbs, of particular interest here, with and without preceding or proceeding suffix, can be found in Appendix B.

2.2.3 Distribution of the suffixes -Ar and -Ir

After these steps, we have looked at the distribution in type frequencies (number of different verbs) and token frequencies (number of occurrences) of the -Ar and -Ir forms of the aorist suffix. Since we are interested in finding out how many times a Turkish speaker actually hears -Ir or -Ar as the suffix of an aorist attached verb, and in particular, how many of these suffixes are preceded by a sonorant, we have used a classification a bit different from the usual one, reported in Table 4 below. For vowel-ending verbs, the suffix is only -r but we counted the aorist form of vowel-ending verbs in the -Ar column if the verb root ended in an A-vowel (i.e. /a/, /e/), and in the -Ir column if the root ended in an I-vowel (i.e. /i/, /ı/, /u/, /ü/) (e.g. *uyu-* 'sleep', the aorist of which is *uyu-r* 's/he sleeps', was counted in the -Ir category, as the verb contains the /u/ sound).

To reflect the frequency of -Ir and -Ar forms following one of the sonorants /l/, /r/ or /n/ at the end of the aorist forms, we counted the vowel-ending verbs in the *Sonorant* line if the consonant preceding the final vowel was a sonorant, and in the *Non-sonorant* line otherwise (e.g. *yürü-* 'walk', whose aorist is *yürü-r* 's/he walks', was counted in the Sonorant line).

We give in Table 3 below examples for all created categories, with light gray cells where we did not expect any aorist form according to the rules laid out in part 2.1 (monosyllabics take -Ar (see D and E) except 13 sonorant-ending monosyllabic irregulars taking -Ir (D, in darker grey), and multisyllabics take only -Ir (F, G)):

Table 3. Examples of Aorist Forms Falling in Different Categories of Verb Root and Suffixes

Verb root characterization				-Ar	-Ir
Vowel-ending	Monosyllabic		A	ye-r, de-r	yu-r (rare)
	Multisyllabic	Sonorant	B	yakala-r	yürü-r
		Non-sonorant	C	öde-r	uyu-r
Consonant-ending	Monosyllabic	Sonorant	D	dal-ar	gör-ür
		Non-sonorant	E	yap-ar	
	Multisyllabic	Sonorant	F		götür-ür
		Non-sonorant	G		çalış-ır

Table 4. Type and Token Frequency Counts of the Suffixes -Ar and -Ir in the Aorist according to Verb Root Characteristics

Verb root characterization				-Ar		-Ir		Total	
				Type	Token	Type	Token	Type	Token
V-ending	Total			695	576599	51	58262	746	634861
	Monosyllabic	Total	A	2	107930	1	99	3	108029
	Multisyllabic	Total		693	468669	50	58163	743	526832
		Sonorant	B	595	355666	20	22195	615	377861
		l		563	316135	3	146	566	316281
		r		18	19476	15	15573	33	35049
		n		14	20055	2	6476	16	26531
		Non-Sonorant	C	98	113003	30	35968	128	148971
C-ending	Total			303	961384	9386	3340775	9689	4302159
	Monosyllabic	Total		214	918427	17	791717	231	1710144
		Sonorant	D	45	131806	17	791717	62	923523
		l		14	13700	7	602213	21	615913
		r		15	81081	5	134241	20	215322
		n		16	37025	5	55263	21	92288
		Non-Sonorant	E	169	786621	0	0	169	786621
	Multisyllabic	Total		89	42957	9369	2549058	9458	2592015
		Sonorant	F	1	4	7976	2121451	7977	2121455
		l		0	0	5100	1431692	5100	1431692
		r		1	4	1737	288778	1738	288782
		n		0	0	1139	400981	1139	400981
		Non-Sonorant	G	88	42953	1393	427607	1481	470560
Total				998	1537983	9437	3399037	10530	4937020

Our frequency calculations in the corpus confirms that there are no monosyllabic non sonorant-ending verbs taking -Ir as stated in Part 2.1.[6] Overall, the -Ir suffix is much more frequent than the -Ar suffix both in token and type, due to the high number of consonant ending multisyllabic verbs. We have further observed that for vowel-ending verbs, a much higher number of aorist forms end in -Ar. We have also gained more insight in respect of the repartitioning of -Ar and -Ir suffixes in the case of monosyllabic sonorant-ending verbs, case to which the exceptions belong. The corpus counts 45 monosyllabic sonorant-ending verbs taking -Ar (rule) compared to 17 taking -Ir (exceptions), distributed in 131.806 occurrences with -Ar (rule) and 791.717 with -Ir (exceptions). In this case, the -Ar suffix is more frequent in type but less frequent in token.

6 However we have found:

- 4 extra cases of monosyllabic sonorant-ending verbs taking -Ir, more precisely ending in /n/ (E): according to the rule, only the 13 monosyllabic sonorant-ending exceptions take -Ir but in the corpus we find 17 verbs because colloquial verbs such as *yun*- 'wash' and some forms of *yen*- 'be eaten', passive of *ye*- 'eat', wrongly marked in their morphological analysis as active forms of *yen*- 'win' in the corpus
- 1 extra case of multisyllabic sonorant-ending verb taking -Ar, more precisely ending in /r/ (F)
- 88 extra cases of multisyllabic non sonorant-ending verbs taking -Ar (G): according to the rule, multisyllabic verbs should take -Ir, but some verbs are constructed from monosyllabic verbs and conjugated as such, even though they are now written in one word (e.g. *zannet*- 'think', *varsay*- 'assume', *vazgeç*- 'give up').

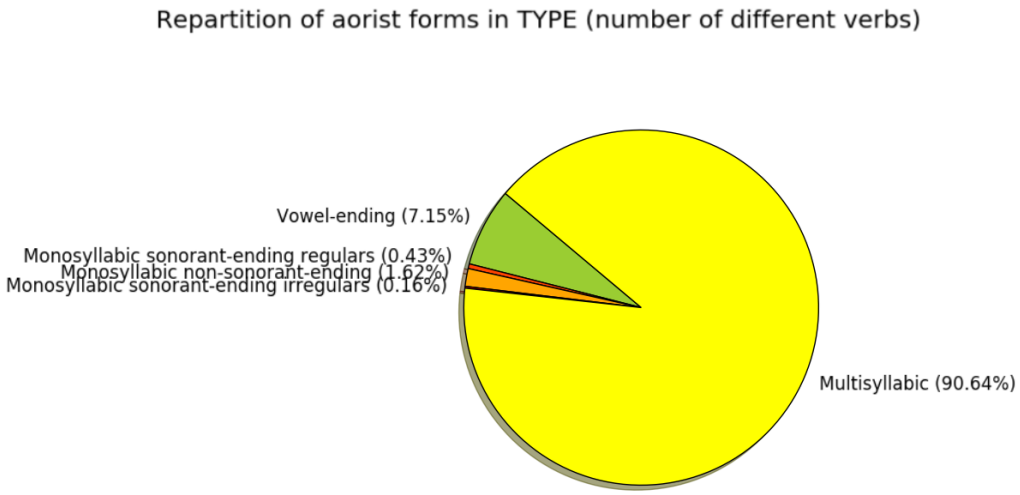


Figure 3. Distribution of the aorist in type with respect to the verb root characteristics

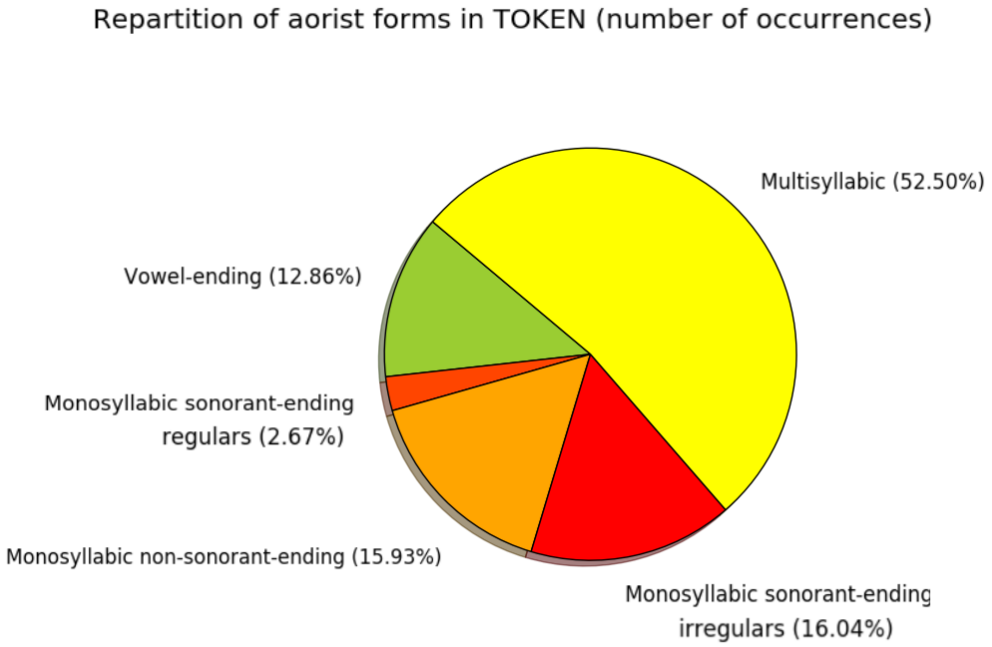


Figure 4. Distribution of the aorist in token with respect to the verb root characteristics

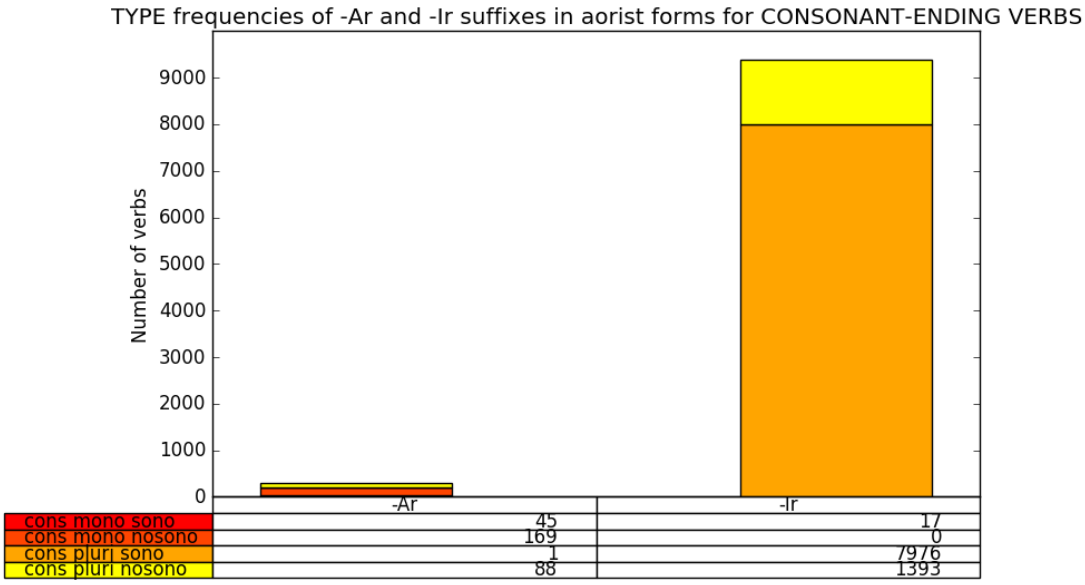


Figure 5. Distribution of the suffixes -Ar and -Ir in type for consonant-ending verbs with respect to the verb root characteristics

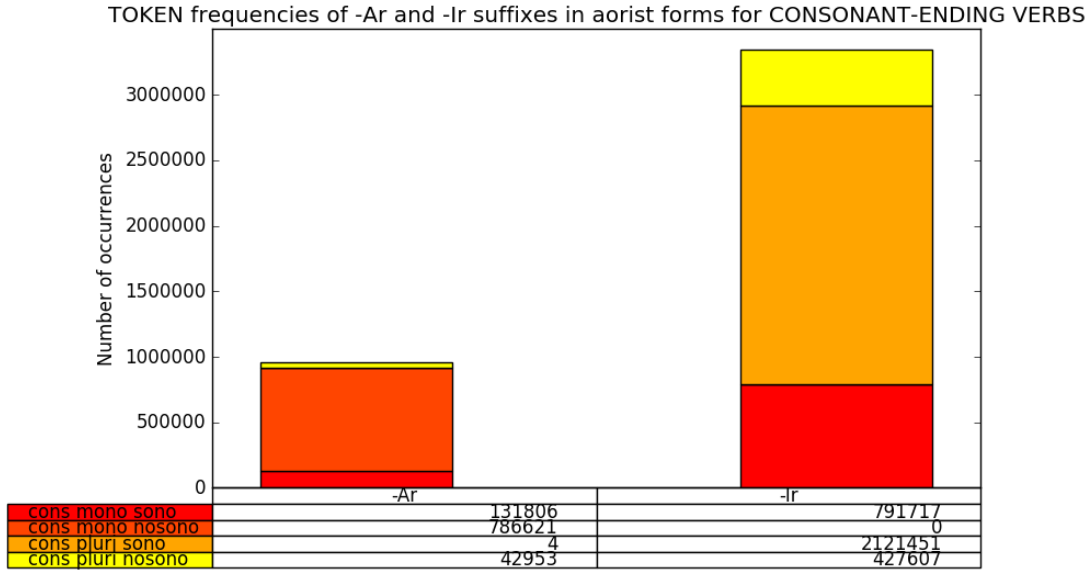


Figure 6. Distribution of the suffixes -Ar and -Ir in token for consonant-ending verbs with respect to the verb root characteristics

2.2.4 Distribution of patterns of consonants and vowels

If we assume that speakers rely on what they hear, we expect them to keep track of the frequencies of some patterns of consonants and vowels in the language. Hence with the intention of finding out which patterns native speakers hear the most and how much they hear them, we have looked into type frequency and token frequency for patterns, defined as such:

i. type frequency: the number of positions where a pattern occurs (can be several in a word and can be chained, e.g. “aCa” with C being any consonant occurs 3 times in the form *anlamamaca*

ii. token frequency: the number of times a pattern is heard (can be several in a word and can be chained, e.g. “aCa” is heard 3 x 10 = 30 times in all the occurrences of *anlamamaca* that occurs 10 times).

And with these two measures, we draw the list of the patterns native speakers hear the most for:

- the patterns of the type “vowel-consonant-vowel”
- the patterns of the type “vowel-sonorant-harmonic vowel”
- the patterns of the type “vowel-sonorant#harmonic vowel” (across a word boundary)
- the patterns of the type “vowel-sonorant-harmonic vowel-r” at any place
- the patterns of the type “vowel-sonorant-harmonic vowel-r” in final position

Relevant lists are given in Appendix C.

The following chapter, Chapter 3 presents theoretical and experimental findings on the processing of the English past tense, an extensively investigated irregular pattern.

CHAPTER 3

THEORETICAL BACKGROUND

3.1 Connectionist and rule-based models

The processing of irregularly behaving patterns has been mostly investigated through the example of the English past tense. This tense has proven to be productive very early on: at the age of 4, English-speaking children are able to produce the past tense form of a new verb given in a sentence, for example "Here is a man who knows how to *rick*; he did the same thing yesterday; he _____." (Berko, 1958). Various models have then strongly argued either about the existence of a united mechanism for all verbs, or about a separate treatment for regular and irregular verbs.

The Single-mechanism model of English past tense argues that every form, whether regular or irregular, should be generated by a unique mechanism. After all, English past tense is one paradigm associating to each verb root a unique output and conveys the same past aspect to any verb, regardless of the complexity of the transformation (from simple suffixation as in *change-changed,* to full suppletion as in *go-went*). According to Connectionism or Parallel Distributed Processing, the language user constantly computes statistics about heard exemplars and combines them to associate a past form to any verb, without reference to its regularity status. Regulars and irregulars would differ in their past-tense output due to a number of their properties i.e. other verbs in their neighborhood, phonetic properties, etc. Several implementations and refinements of this model (Plunkett and Marchman, 1993; Hare, Elman and Daugherty, 1995; Hahn and Chater, 1998; Nakisa, Plunkett and Hahn, 2000, among others) have been built on the seminal two-layer artificial neural network proposed by Rumelhart and Mc Clelland (1986).

One of the most characteristic features of such models - as noticed by Seidenberg and Plaut (2014) - is that they do not need any specific knowledge of English past tense. Instead, fed with pairs of the type “verb-past tense”, these models adjust their weights and tend to reflect general learning and processing mechanisms. Assuming a basic single mechanism, they focus on defining the details of the model and therefore propose a precise implementation in support of the theory.

On the other hand, the Dual-mechanism model has advanced the idea that two mechanisms are needed to account for the two types of forms. Since regulars and irregulars have a different behavior, they would be processed differently: regulars computed with a symbolic rule, irregulars retrieved from the lexicon. This rule-based model suggests that when a speaker produces the past form of a verb, if s/he finds it in the lexicon then this stored irregular form is chosen, otherwise it is computed directly by the rule, with no need to store this form. This view, sometimes called “Words-and-Rules”, has been developed by Pinker and Prince (1988), Marcus, Pinker, Ullman, Hollander, Rosen and Xu (1992), Marcus (1998) and Pinker (2001), among others.

The strength of the implementations of the Dual-mechanism model as noted by Seidenberg and Plaut (2014) is to integrate rules and abstract structures that present themselves more widely in the language, in phonology or syntax. Thus the model does not need to be provided with pairs of verb stems and correct past tense forms which is a very partial sample of English language, but can derive the past tense of a verb from rules with a wider scope.

These two lines of research have proven very productive during the last thirty years but none of them could entirely solve the research problem of how the productive use of morphemes possessing irregularity in distribution can be accounted

for (Mc Clelland and Patterson, 2002; Pinker and Ullman, 2002). At the heart of the discussion lies the observation that connectionist models using similarity between verbs generate inappropriate forms that native speakers actually never produce such as *smile-*smole* (overgeneration of irregulars); rule-based models, however, make categorical predictions, and provide regular forms that never surface in the language such as *come-*comed* (undergeneration of irregulars). The Single-mechanism model relies on the domain-general notion of analogy between items, used in other phenomena such as reading or object recognition. On the contrary, the Dual-mechanism model makes more use of properties specific to the domain of language.

The debate has also been fueled with new linguistic facts to explain. In particular, the application of the rule has proven to be sensitive to phonology and semantics. Bybee and Moder (1983) revealed the predominance of verbs starting with /s/+stop (/t/, /p/) in the verbs having an irregular past tense formed by vowel change (*spin-spun, stick-stuck*). Ramscar (2002) showed that when the nonce verb *sprink* was introduced in a story, the subjects preferred the regular past *sprinked* when the story was about eye problems and reminded them of the verb *blink* whose past is regular *blinked*; and they preferred the irregular past *sprank* when the story was related to drinking, priming the verb *drink* whose past is irregular *drank*. The Single-mechanism advocates propose to simply integrate these as additional constraints to their model relying uniquely on similarity. Meanwhile, the defenders of the Dual-mechanism model need to assume the influence of additional factors on the rule application for regulars, and they even accept that some kind of similarity must be at stake for the case of irregulars, note Seidenberg and Plaut (2014).

Recent years have seen an evolution of the mathematical frameworks (Bayesian Probability Theory, Machine Learning), allowing for new accounts of

language processing and acquisition, probabilistic in nature (Chater and Manning, 2006; Tenenbaum, Kemp, Griffiths and Goodman, 2011). In parallel, theoretical advances in the debate (and not only refining of the models' parameters or details) have indicated ways to disregard the existing dichotomy "rule vs. usage". For example, Pinker (2001) has suggested that calling a form the default is not a matter of frequency only: let us take the example of the suffix –s which is argued to be the default plural for nouns in English and German. Whereas in English more than 99 percent of the nouns form their plural by concatenation of –s, only 7 percent of the nouns in German are rendered plural this way. The default form must be the result of symbolic composition, not storage in memory, and appear in several cases of systematic generalization: -(e)d is argued to be the default marker for the English past tense because it is the form used for the unusual sounding nonce verbs even when they sound very different from existing verbs (*ploamph - ploamphed*), for verbs spontaneously created in the discourse (*sally-ride - sally-rided*) and in children's errors (*go - *goed*).

In 2003, Albright and Hayes proposed the Minimal Generalization Learner (MGL) model which proposes multiple stochastic rules (rules with assigned confidence scores) that are gradually constructed through a process of iterative generalization over pairs of forms "verb - past form". Arguing that "pure" analogy can capture effects irrelevant to the processing of the pattern and is not restrictive enough as a model of morphology, the model makes instead use of the emergence of structured similarity among verbs, grouped in "Islands of Reliability". The model generalizes from the exemplars by extracting rules, comparing them with one another to find the tightest rules, ending up with a small number of general rules and a large number of other less general rules. The past form for a given verb is thus determined

by the rule with the highest probability and the "Islands of Reliability" are the resulting groups of verbs to which the probabilistic rules apply.

For example, every English verb ending in a voiceless fricative (/f/, /th/, /s/, /sh/) has a regular past, so the model could build the rule from the initial data: "Suffix /-t/ attached to stems ending in voiceless fricatives". The verbs in the scope of this rule ending in voiceless fricatives (*miss, wish, laugh*, etc.) form an Island of Reliability for regular past forms. Since the rules are probabilistic, a particular verb can belong to several Islands of Reliability: for example, a nonce item such as *dize* belongs to both an Island of Reliability for regular pasts (verbs ending in [z]: *realize-realized, suppose-supposed*, etc.) and an Island of Reliability for irregular pasts (verbs with vowel change from [aɪ] to [o]: *rise-rose, ride-rode, dive-dove*). The MGL model has been successfully tested against an implementation of the Single-mechanism model called the Generalized Context Model (GCM) (Nosofsky, 1990).

Rather than bringing an end to the debate, Albright and Hayes (2003) opened a new research avenue since the proposed alternative is a model, necessarily making assumptions and differing from other models in the way that they conceptualize the data and the human processing mechanisms. For instance, Racz, Beckner, Hay and Pierrehumbert (2014) have retested the Minimal Generalization Learner against the Generalized Context Model. This time, analogy to individual instances (GCM) proves to be a better approximation of the behavior of subjects than recourse to abstract generalization (MGL) but each model explains a part of the observed variance. In the case of nonces such as *skast*, similar to verbs whose past form is identical to their infinitive form (*cut-cut, hurt-hurt*), the strong preference of participants for regularization (*skasted*) over irregularization (*skast*) is better predicted by GCM than by MGL. This happens because MGL, by its selection of the

highest probability to produce the output for a verb, penalizes rules supported by a low number of verbs in English. However, in the case of nonces similar to verbs that form their past form with a vowel change from [aɪ] or [i] to [o] (*drive-drove, weave-wove*), the MGL makes better predictions of the participants' answers, because some of the nonces such as *quine (quined)* or *sline (slined)* fall into an Island of Reliability wider for regulars (*line-lined, whine-whined*) than for irregulars (only *shine-shone*), while the GCM fails to capture this structured similarity.

3.2 Related psycholinguistics experiments

All these approaches proposed to model the processing of morphemes that embody some kind of irregularity, have a goal of reaching the gold standard set by human performance: a model that would represent the verbal processing mechanisms at stake in the English past tense should predict the past forms present in the language, but also human answers for non-existent verbs collected in psycholinguistics experiments. To gather these data to which the results of the models can be compared, Bybee and Moder (1983), Prasada and Pinker (1993), Albright and Hayes (2003) and Racz, Beckner, Hay and Pierrehumbert (2008) designed experiments with nonce verbs, where they have asked participants to rate the most plausible form between the regular and a proposed irregular past form, or to simply produce the past form of a given verb root.

In Bybee and Moder (1983), 34 participants heard 93 nonce verbs and had to write the associated past tense with the help of the sentence frame “Sam likes to _____. Yesterday he_____”. The proposed nonces varied in their phonological neighborhood proximity to the irregular verbs including the common /i/ - /Λ/ vowel change in their past tense (*string - strung, cling – clung, shrink – shrunk*).

The results reflected this gradient proximity: the more the nonce matched the verbs with this vowel change phonologically, the more irregular past tenses including the vowel change were elicited: thus, *spling* rendered *splung* with a rate of 44%, *shink* gave *shunk* with a rate of 24% and *sid* was produced as *sud* in the past tense with a rate of 7%. Generalization of irregular patterns was said to occur mostly on the basis of rhyme with existing items.

In another experiment that intended to contribute to the understanding of the issue, Prasada and Pinker (1993), chose to test 24 English students on 60 nonce verbs by giving definitions to the nonces and asking them to fill in frame sentences on a paper questionnaire. Using the mean of number of rhyming monosyllabic regulars or irregulars as a similarity measure, stimuli were designed to be at three distance levels - either prototypical, intermediate, or distant - either from existing regulars or from existing irregulars. Authors noted that given the high number and variety of regular English verbs, a nonce can only be "distant from existing regulars" if it is very dissimilar to all English words (e.g. *ploamph*), which could affect participants' answers. For nonces such as *spling*, the answers were classified as "suffixed" (regular past tenses such as *splinged*) or "non-suffixed" (irregular past tenses such as *splung*, but also spelling errors in the stem such as *spinged* or *spung*). "Non-suffixed" answers decreased with distance from irregular verbs: for example, "non-suffixed" answers represented 42% of the answers for a prototypical irregular nonce such as *cleed (cled)*, 33% for an intermediate irregular nonce such as *fring (frung)*, and 17% for a nonce distant from the irregulars such as *trisp (trusp)*. But this distance effect was not observed for regular verbs: "suffixed" answers did not increase with distance from regular verbs. This way, a prototypical regular nonce *greem*, an intermediate

regular nonce *smeej* and a nonce distant from regulars *frilg*, were all produced with a rate of 96% as regulars, respectively *greemed*, *smeejed* and *frilged*.

To test the Minimal Generalization Learner Model introduced earlier, Albright and Hayes (2003) also ran a rating and eliciting experiment based on their concept of "Island Of Reliability" (IOR), which are the clusters grouping verbs getting their past tense through the same probabilistic rule, as previously mentioned. Twenty students were auditorily presented with sentences that contained the infinitive form of 60 nonce verbs and were instructed to read aloud subsequent carrier sentences such as "I think I'd really enjoy to _____. My friend Sam _____ once, and he loved it.", orally filling the blanks with the appropriate form. Stimuli could belong to an IOR for regular items, for irregular items, for both regulars and irregulars, or for neither regulars nor irregulars. Similar to previous results, but formulated in terms of Island of Reliability instead of neighborhood similarity (distance between items used by Bybee and Moder (1983)) or prototypicality (distance used by Prasada and Pinker (1993)), irregular forms occurred more when the item belonged to an IOR for irregulars than when it did not: for example the irregular form *flep* was produced at 23% for the nonce *fleep*, belonging to an IOR for irregulars, i.e. the scope of the probabilistic rule deriving *bleed-bled* for example. Meanwhile *spuck* was produced at 7% for the nonce *spack*, which is not belonging to such an IOR. But importantly and contrary to Prasada and Pinker (1993), a comparable result was observed for regular verbs: regular forms were more frequent for items belonging to a group of verbs having regulars than for items that are not in such an IOR; for instance, the nonce *blafe*, belonging to an IOR for regulars, was regularized as *blafed* at 89% while the nonce *queed*, not belonging to such an IOR, gave *queeded* only at 52%.

Racz, Beckner, Hay and Pierrehumbert (2008) made use of the possibilities offered by the internet to design an online experiment with a forced choice between regular and irregular past tense forms. 316 English nonce verbs, belonging to 5 prototypical classes of irregulars, were visually presented in carrier sentences to as many as 111 participants. Irregularization rates were found to vary greatly among classes of nonces: 0.48 for the *drive – drove* class, 0.42 for the *sing – sang* class, 0.41 for the *keep – kept* class, 0.33 for the *burn – burnt* class, 0.17 for the *cut – cut* class.

3.3 Frequency considerations

Conjugating English verbs in the past tense is a frequent pattern in itself, as it occurs in 86% of the 1000 most common verbs. The irregularity in the pattern has emerged from the transition from a vowel change paradigm to a suffixation paradigm in history. English contains thousands of regular verbs and about 180 irregular verbs that are left-overs of the former system. Irregulars resist regularization diachronically thanks to their grouping in clusters and to their high frequency: according to Bybee (1982), out of the 30 most frequent past-tense forms 22 are irregulars. However, the distinction between regulars and irregulars may not be that clear-cut: the past form of 59% of the 181 irregulars end in /d/ or /t/, just like the regular past forms. Moreover, many irregular past tense forms are obtained by changing the vowel of the stem, which grants vowel change the status of a secondary rule, or quasi-rule, as labeled by Mc Clelland and Patterson (2002).

Irregulars are high-token-frequency (number of occurrences) low-type-frequency (number of verbs) items. Usage-based studies suggest that morphological patterns are extended on the basis of type frequency more than token frequency (Bybee, 1995; Baayen, 2001; Bybee and Hopper, 2001). But their opponents argue

that usage-based models should take into account the percentage of tokens and not types, to implement analogy with what is heard in the language (Pinker, 2001).

In children's lexicon, irregular verbs are so frequent in that they remain more frequent in tokens than regular verbs, even when their vocabulary considerably expands in regular verbs. Furthermore, in the study of Raez, Beckner, Hay and Pierrehumbert (2014), individuals with higher vocabulary scores happen to rate irregular forms as more plausible compared to individuals with lower vocabulary scores. Participants' strength of vocabulary could reflect in the number of stored instances of lower frequency irregular verbs and therefore their acceptance of new irregular forms.

The literature on the topic inspired us to design a psycholinguistic experiment where we have exposed a high number of participants to nonce items both auditorily and visually. We have tested a wide range of stimuli in order to check various measures of similarity, and provided frequencies we used both in type and token to be able to compare them. The ensuing Chapter lays out the procedure of our task to test adults' production of nonce aorist forms.

CHAPTER 4

METHODOLOGY

4.1 Stimuli

As in this study we are interested in finding out how adult native speakers of Turkish would produce non-existing verb-like items in the aorist and whether the outcome would reflect the application of a rule or the effect of analogy, we have created nonce verbs that mimic the existing verbs in Turkish.

Recall that all irregularly behaving verbs in the aorist are monosyllabic roots ending in sonorants, i.e., /l/, /r/, /n/, so we have designed our stimuli to include only monosyllabic /l/, /r/, /n/-ending nonce verbs. When we have further looked into the distribution of the sonorants in the 13 irregular verbs we have observed that the majority end in /l/ (7 out of 13 items) and, /r/ (5 out of 13 items). The only /n/-ending verb happens to be the infamous *san-* which is already problematic for native speakers of Turkish as 'san-ar' is widely accepted among speakers. Hence we have decided not to include any /n/-ending nonce-verbs and restricted ourselves with /r/ and /l/-ending verbs. Therefore the template we have constructed for the nonce verbs has been Consonant - Vowel - Sonorant (CVS). 21 consonants x 8 vowels x 2 sonorants give rise to 336 roots/syllables in Turkish as laid out in Table 5 below.

Table 5. The 336 Possible CVS Syllables (Color Code Explained in the Text)

	b	c	ç	d	f	g	h	j	k	l	m	n	p	r	s	ş	t	v	y	z
a	al	bal	cal	çal	dal	fal	gal	hal	jal	kal	lal	mal	nal	pal	ral	sal	şal	tal	val	yal	zal
a	ar	bar	car	çar	dar	far	gar	har	jar	kar	lar	mar	nar	par	rar	sar	şar	tar	var	yar	zar
e	el	bel	cel	çel	del	fel	gel	hel	jel	kel	lel	mel	nel	pel	rel	sel	şel	tel	vel	yel	zel
e	er	ber	cer	çer	der	fer	ger	her	jer	ker	ler	mer	ner	per	rer	ser	şer	ter	ver	yer	zer
i	il	bil	cil	çil	dil	fil	gil	hil	jil	kil	lil	mil	nil	pil	ril	sil	şil	til	vil	yil	zil
i	ir	bir	cir	çir	dir	fir	gir	hir	jir	kir	lir	mir	nir	pir	rir	sir	şir	tir	vir	yir	zir
ı	ıl	bıl	cıl	çıl	dıl	fil	gıl	hıl	jıl	kıl	lıl	mıl	nıl	pıl	rıl	sıl	şıl	tıl	vıl	yıl	zıl
ı	ır	bır	cır	çır	dır	fır	gır	hır	jır	kır	lır	mır	nır	pır	rır	sır	şır	tır	vır	yır	zır
o	ol	bol	col	çol	dol	fol	gol	hol	jol	kol	lol	mol	nol	pol	rol	sol	şol	tol	vol	yol	zol
o	or	bor	cor	çor	dor	for	gor	hor	jor	kor	lor	mor	nor	por	ror	sor	şor	tor	vor	yor	zor
ö	öl	böl	cöl	çöl	döl	föl	göl	höl	jöl	köl	löl	möl	nöl	pöl	röl	söl	şöl	töl	völ	yöl	zöl
ö	ör	bör	cör	çör	dör	för	gör	hör	jör	kör	lör	mör	nör	pör	rör	sör	şör	tör	vör	yör	zör
u	ul	bul	cul	çul	dul	ful	gul	hul	jul	kul	lul	mul	nul	pul	rul	sul	şul	tul	vul	yul	zul
u	ur	bur	cur	çur	dur	fur	gur	hur	jur	kur	lur	mur	nur	pur	rur	sur	şur	tur	vur	yur	zur
ü	ül	bül	cül	çül	dül	fül	gül	hül	jül	kül	lül	mül	nül	pül	rül	sül	şül	tül	vül	yül	zül
ü	ür	bür	cür	çür	dür	für	gür	hür	jür	kür	lür	mür	nür	pür	rür	sür	şür	tür	vür	yür	zür

From these 336 syllables, we have eliminated the following:

i. existing verb roots, regulars and irregulars (cf. Table 6), relying on the list of monosyllabic verbs provided in (Nakipoğlu and Üntak, 2008), (42 verbs):

Table 6. Existing Monosyllabic Regular (Green) and Irregular (Red) Verbs in the Possible Syllables

	b	**c**	**ç**	**d**	**f**	**g**	**h**	**j**	**k**	**l**	**m**	**n**	**p**	**r**	**s**	**ş**	**t**	**v**	**y**	**z**
a	al	bal	cal	çal	dal	fal	gal	hal	jal	kal	lal	mal	nal	pal	ral	sal	şal	tal	val	yal	zal
a	ar	bar	car	çar	dar	far	gar	har	jar	kar	lar	mar	nar	par	rar	sar	şar	tar	var	yar	zar
e	el	bel	cel	çel	del	fel	gel	hel	jel	kel	lel	mel	nel	pel	rel	sel	şel	tel	vel	yel	zel
e	er	ber	cer	çer	der	fer	ger	her	jer	ker	ler	mer	ner	per	rer	ser	şer	ter	ver	yer	zer
i	il	bil	cil	çil	dil	fil	gil	hil	jil	kil	lil	mil	nil	pil	ril	sil	şil	til	vil	yil	zil
i	ir	bir	cir	çir	dir	fir	gir	hir	jir	kir	lir	mir	nir	pir	rir	sir	şir	tir	vir	yir	zir
ı	ıl	bıl	cıl	çıl	dıl	fıl	gıl	hıl	jıl	kıl	lıl	mıl	nıl	pıl	rıl	sıl	şıl	tıl	vıl	yıl	zıl
ı	ır	bır	cır	çır	dır	fır	gır	hır	jır	kır	lır	mır	nır	pır	rır	sır	şır	tır	vır	yır	zır
o	ol	bol	col	çol	dol	fol	gol	hol	jol	kol	lol	mol	nol	pol	rol	sol	şol	tol	vol	yol	zol
o	or	bor	cor	çor	dor	for	gor	hor	jor	kor	lor	mor	nor	por	ror	sor	şor	tor	vor	yor	zor
ö	öl	böl	cöl	çöl	döl	föl	göl	höl	jöl	köl	löl	möl	nöl	pöl	röl	söl	şöl	töl	völ	yöl	zöl
ö	ör	bör	cör	çör	dör	för	gör	hör	jör	kör	lör	mör	nör	pör	rör	sör	şör	tör	vör	yör	zör
u	ul	bul	cul	çul	dul	ful	gul	hul	jul	kul	lul	mul	nul	pul	rul	sul	şul	tul	vul	yul	zul
u	ur	bur	cur	çur	dur	fur	gur	hur	jur	kur	lur	mur	nur	pur	rur	sur	şur	tur	vur	yur	zur
ü	ül	bül	cül	çül	dül	fül	gül	hül	jül	kül	lül	mül	nül	pül	rül	sül	şül	tül	vül	yül	zül
ü	ür	bür	cür	çür	dür	für	gür	hür	jür	kür	lür	mür	nür	pür	rür	sür	şür	tür	vür	yür	zür

ii. possible nonce forms that start with the letter 'j' as there are no verbs starting with this letter except for a few verbs derived from nouns or adjectives (16 syllables).

iii. syllables that have a meaning in Turkish, like *gür* since it is an adjective meaning 'abundant' (97 syllables).

iv. consonant-initial syllables that already have a meaning in Turkish once added the infinitive -mAk, *e.g. par-mak* since *parmak* is a noun form meaning 'finger' (1 syllable).

v. possible nonce forms that when conjugated with the *-Ar* or the *-Ir* forms of the aorist (expected answers of the participants), resemble vowel-ending verbs with -r allomorph of the aorist, e.g. *yür-ür*, because the vowel-ending verb *yürü-* 'walk' is conjugated as *yürü-r* 's/he walks', or *sul-ar*, because the verb *sula-* 'water' is conjugated as *sula-r* 's/he waters' in the aorist (12 syllables).

When all these items were removed what we were left with were 168 possible nonce verb stimuli as in illustrated below in Table 7:

Table 7. 168 Nonce Monosyllabic Verbs Chosen for the Study

	b	**c**	**ç**	**d**	**f**	**g**	**h**	**j**	**k**	**l**	**m**	**n**	**p**	**r**	**s**	**ş**	**t**	**v**	**y**	**z**
a		cal				gal								ral				val		zal
a										lar	mar			rar		şar				
e		cel					hel			lel			pel	rel		şel		vel		zel
e	ber		çer							ler	mer	ner		rer						
i	il	cil				gil	hil			lil		nil		ril		şil	til	vil	yil	
i	ir	cir			fir		hir					nir		rir	sir	şir	tir	vir	yir	
ı	ıl	bıl	cıl	çıl	dıl	fıl	gıl	hıl		lıl	mıl	nıl	pıl	rıl	sıl	şıl	tıl	vıl		zıl
ı	ır	bır	cır	çır	dır					lır	mır	nır		rır		şır		vır		zır
o		col	çol							lol	mol	nol				şol		vol		zol
o	or	cor		dor	for	gor						nor	por	ror				vor		
ö		cöl			föl		höl		köl	löl	möl	nöl	pöl	röl	söl	şöl	töl	völ	yöl	zöl
ö	bör	cör	çör	dör	för		hör			lör	mör	nör	pör	rör		şör		vör	yör	zör
u	ul	cul				gul	hul			lul	mul	nul		rul		şul		vul		zul
u	ur	cur	çur				hur			lur	mur		pur	rur		şur			yur	zur
ü	ül	bül	cül	çül	dül	fül		hül		lül		nül	pül	rül	sül	şül		vül		
ü	ür	cür			für					lür	mür	nür		rür		şür		vür		zür

We have randomly divided the 168 items into 2, to get two groups of 84 stimuli to be tested in Condition A and Condition B respectively (cf. Appendix Tables D1 and D2).

4.2 Fillers

To make the purpose of the experiment opaque to the participants and to ensure that the participants get involved in the task, 27 multisyllabic fillers were formed by extracting a list of frequently occurring consonant-initial and non-/l/ or /r/-ending

multisyllabic verbs from the corpus. Then we changed the first consonant of each verb randomly and constructed the 27 items listed in (1) below (cf. Appendix D3):

(1) *deliş-, derek-, dözle-, lelgele-, lestele-, lombala-, loşalt-, mekerle-, mitreş-, mıraşla-, moparla-, nayış-, nırılda-, raçmala-, rerpele-, seğerle-, şekele-, sepola-, süzenle-, tapsa-, tıpırda-, zankıla-, zasla-, zetiş-, zıprat-, zönet-, züzle-.*

We have expected the participants to conjugate these multisyllabic items just as the existing verbs, with the suffix -r for vowel-ending ones and the suffix -Ir for consonant-ending ones. The same set of fillers were used in both conditions.

4.3 Carrier sentences

For each nonce item, we have built a context sentence to introduce the verb and a carrier sentence to elicit it. An item was introduced and elicited in the same context across participants, with the context sentence presented auditorily and the carrier sentence in written form.

To create various and plausible contexts and to make the task as appealing and as natural as possible, we generated a pair of sentences with a Python program picking, for each item:

i. first a random context among these 7:

- 3 A-type contexts (where A stands for [a] or [e]):

1. … *V-mAyA gider* '… goes to Verb-NOM-DAT'
2. … *V-AcAk* '… Verb + Future tense'
3. … *V-mAk ister* '… wants to Verb-NOM'

- 3 I-type contexts (where I stands for [i] or [ɪ] in context 4, and [i], [ɪ], [u] or [ü] in contexts 5, 6):

4. … *V-mAyI sever* '… likes to Verb-NOM-ACC'
5. … *V-DI* '… Verb + Past tense'
6. … *V-mIş* '… Verb + Evidential '

- 1 one neutral context:

7. … … *V dedi* '… told …“Verb (Imperative)!”'

ii. a random character name among:
İdil, Ece, Özge, Elif, Gül, Aslı (feminine)
Ali, Emre, Can, Ozan, Cem, Mert (masculine)

iii. a random time marker for aorist among:

her sabah	'every morning'
her gün	'everyday'
her hafta	'every week'
her akşam	'every evening'
kışın her gece	'every winter night'
yazın her gün	'every summer day'
kışın akşamları	'in winter evenings'
yazın sabahları	'in summer mornings'
hep	'always'
her zaman	'always'
gece gündüz	'day and night'
iki günde bir	'every two days'
sabahları	'in the mornings'
akşamları	'in the evenings'
geceleri	'at nights'
yazları bazen	'sometimes in summer'
kışları sık	'often in winter'
geceleri genelde	'generally at nights'

iv. then extra elements to accompany the contexts that-is-to-say:

- for context 2 (future tense), a random time marker for future among the following forms:

yarın	'tomorrow'
haftaya	'next week'
bir sonraki gün	'the next day'
önümüzdeki hafta	'next week'
bu akşam	'this evening'

- for contexts 5, 6 and 7 (past tenses), a random time marker for past among the following forms:

dün	'yesterday'
önceki gün	'the day before'
demin	'a bit before'
az önce	'a bit before'
önceki gece	'last night'

- for context 7 (imperative), a random interlocutor whom the first character addresses among: *Esra, Yeşim, Merve (feminine), Sinan, Burak (masculine).*

We eventually built sentences made up of these blocks according to the scheme:

Context sentence: *Name + Verb conjugated in context.*

Carrier sentence: *Name* + A*orist-triggering adverb + Verb root to complete*

Looking at an example item, the nonce monosyllabic verb *dül-*, we obtain:

Context sentence: *Ali dülmeyi sever.* 'Ali likes to dül.'

Carrier sentence: *Ali kışın her gece dül__.* 'Ali every winter night dül___,'

where we expect participants to produce the monosyllabic nonce verb *dül-* conjugated in the aorist with the regular suffix -Ar *dül-er* or the irregular suffix -Ir *dül-ür*. Items were displayed in a subject-by-subject randomized order. For the complete list of sentences, cf. Appendix D.

The auditorily presented context sentences have been recorded by two Turkish native speakers of similar age, one woman and one man, on phone devices. Half of the context sentences designed for Conditions A and B were recorded by the female speaker, the other half by the male speaker. Sound volume has been equalized using the software Audacity.[7]

4.4 Participants

In order to get a sufficiently numerous and representative sample, 102 students at Boğaziçi University have been invited to participate in the experiment through e-mails sent by their instructors. The students did not know the goal of the experiment. Two participants did not complete the whole experiment because the program crashed during the testing phase and the students could not be reached later. Among the remaining 100 participants, 2 reported that their mother tongue is a language

7 http://www.audacityteam.org/

other than Turkish and 8 that they are bilinguals hence the data obtained from these 10 participants had to be removed from the study.

So in the 90 monolingual Turkish speakers' data we analyze, 46 participants were exposed to condition A and 44 participants to condition B. Due to the recruitment method, age was concentrated around 20 years old (Min. 18; Max. 43; Mean 21.57). Gender was roughly balanced across conditions by the experimenter during testing phase (52 women, 37 men and 1 other answer).[8] All participants had a good command of English, and 71 reported having knowledge of at least one other foreign language. The same experimenter, a native speaker of French learning Turkish for 3 years, welcomed all participants and gave them instructions in Turkish.

4.5 Procedure

The experiment was presented in the form of an Inflected Verb Elicitation Task realized in the laboratory, programmed with the software Open Sesame.[9]

Seated in front of a computer in an experimental room, participants were instructed on the screen to carefully listen to the sentences they would hear as they contained made-up verbs, then to complete the verb in the new sentence presented on the screen with a conjugated form. In order to elicit enough aorist answers without explicitly mentioning it, they were simply asked not to use the Turkish present continuous tense that could be a good fit as well in the proposed sentences.

The context sentence introducing each stimuli was heard while a fixation dot appeared on the black screen. Then a carrier sentence appeared on the screen that

8 The participants were recruited from Departments of Foreign Language Education (19), Translation (17), Computer Science (15), Western Languages and Literatures (14), Linguistics (11), Turkish Language and Literature (4), Psychology (3) and other (7).

9 http://osdoc.cogsci.nl/

contained an appropriate time marker implicitly encouraging the participants to produce orally the missing aorist form in the microphone and press a key on the keyboard to move to the next item. For detailed instructions refer to Appendix E.

Before the test, each participant was presented with 6 training items (nonce verbs from the other condition) in the presence of the experimenter, repeated if the experimenter did not hear any aorist form in the first 6 ones which suggest that the participant did not feel familiar yet with the task. Participants were encouraged to use the first form that came to their mind as fast as possible. They could either read the whole sentence or just the nonce inflected verb whichever was the easiest for them.

The experiment was self-paced and 2 pauses were inserted in between every 40 items in the test phase. In total, each participant was presented with 117 items (84 stimuli + 27 fillers + 6 training items). The task was followed by a quick computer questionnaire, to provide us with more explanatory factors to be explored post-hoc (gender, age, handedness, mother tongue, foreign language exposure). At the end of the experiment, participants were told that there were no good or bad answers and upon request provided with more explanation. The full experiment lasted 15 to 20 minutes.

For each participant, some of her/his 84 answers have been counted as "Other forms" and not further analyzed for the following reasons:

- Composed forms: aorist composed with another tense was produced, e.g. *cel* > *cel-irdi*
- Tense error forms: the right verb but a tense marker other than aorist was produced, e.g. *gal* > *gal-acak*
- Verb error forms: aorist but with a wrong verb root was produced, e.g. *lar* > *lar-lar*

- Harmony errors: vowel harmony was not respected, e.g. *cül* > *cül-ar*
- Missing forms: the space key was pressed too early before the end of their recording during the self-paced experiment (technical reason).

As can be seen below in Table 8, these other answers represent a small proportion of the total answers (15%). We report only 6 participants, 3 in each condition, for which less than half of the answers were available to analysis. These participants were included in the analysis because the instruction of the task asked participants to complete the verb according to the context. Since it did not explicitly ask to conjugate the verb in the aorist, giving non-aorist answers does not specifically reflect a misunderstanding of the task but could have been caused by strategies which are not the focus of our analysis (e.g. some participants reported to the experimenter that they deliberately introduced other tenses in their answers after producing too many aorist forms).

Table 8. Breakdown of Counts and Rates of Aorist and Other Forms in Conditions A and B

	Total		Cond A		Cond B	
	Count	**Rate**	**Count**	**Rate**	**Count**	**Rate**
Aorist forms	6391	85%	3347	87%	3044	82%
--Regular forms -Ar	5468	86%	2864	86%	2604	86%
--Irregular forms -Ir	923	14%	483	14%	440	14%
Other forms	1169	15%	517	13%	652	18%
--Composed forms	109	9%	61	12%	48	7%
--Tense error forms	411	35%	136	26%	275	42%
--Verb error forms	306	26%	58	11%	248	38%
--Harmony error forms	11	1%	6	1%	5	1%
--Missing forms	332	28%	256	50%	76	12%
Total forms	7560		3864		3696	
Number of nonces	168		84		84	
Number of participants	90		46		44	

We now turn in Chapter 5 to the analysis of the aorist forms produced by the 90 participants and the distribution of the -Ar and -Ir forms according to the properties of the nonces.

CHAPTER 5

RESULTS

In this chapter we will present the results of the experiment and discuss the implications of the findings obtained. Recall that if participants are found to resort to the use of the –Ar form, they can be argued to have formulated a rule for the Turkish aorist, considered -Ar as the default form and applied it across-the-board. If, however, they opt for the use of -Ir, then the issue of which nonce forms would trigger the use of –Ir becomes of crucial interest. In what follows, we will examine the results under two main sections. Section 5.1 lays out the general results and touches upon whether what we have observed is correlated with application of a rule. The proceeding sections investigate the effect of similarity and frequency in forms that have given rise to the -Ir use, for different measures of similarity: rhyme in Section 5.2, other variables in Section 5.3 and combinations of them in Section 5.4.

5.1 Default status: Is there a rule in the production of the aorist?

The results we have obtained in Table 9 show that overall, the so-called regular -Ar form has proven to be the dominant form, hence the default, as the participants have opted for the use of -Ar with a rate of 86% and the use of the so-called irregular -Ir with a rate of 14% on the 6391 aorist answers we obtained (cf. Table 8 for details).

Table 9. –Ir Use for the 168 Nonce Verbs

TOTAL	-Ir Use	# of nonces	Irregular forms	Total tokens
All nonces	14%	168	923	6391

More participants produced the -Ar form than the -Ir form for every tested nonce verb except one, i.e., *rur-* where *rur-ur* was produced by 21 participants and *rur-ar* by 15 participants. Furthermore, 6 nonce roots *nül-, pül-, şöl-, mol-, cil-, şil-* are observed to have always been produced with the -Ar form and never with the -Ir form. This suggests that the suffix -Ar has been considered to be the default form of the aorist for monosyllabic consonant-ending verbs, even for sonorant-ending ones.

Regarding the allomorphic distribution of the Turkish aorist, one would expect the -Ar marker to be used with monosyllabic nonce roots irrespective of whether the root ends in a sonorant or a non-sonorant ending consonant, given the fact that –Ar is the marker of aorist for the majority of monosyllabic verbs in Turkish. Recall from Chapter 2 that in Turkish out of 232 monosyllabic verbs 219 are conjugated with –Ar. Therefore that the overwhelming majority of the participants converged on the choice of –Ar as the marker of aorist (with a rate of 86%) is not surprising.

We also remind the reader that Nakipoğlu and Ketrez (2006) have observed that both types of errors, -Ar use for -Ir taking monosyllabic verbs (overregularization) and -Ir use for -Ar taking monosyllabic verbs (irregularization), are produced by Turkish children from 2 to 8 years of age. Irregularization errors are more prevalent overall in the acquisition process, but overregularizations are predominant in the youngest group (mean age 3.5) and contrary to irregularization errors, even pervade vowel-ending verbs (**okar* instead of *oku-r* for the verb *oku-* 'read', and *uyar* instead of *uyu-r* for the verb *uyu-*'sleep'). The default status of -Ar may not be as clear-cut as that of –(e)d for English, as -Ir can be considered to be the default aorist marker for multisyllabic verbs and the non-sonorant-ending monosyllabic verbs. More investigation using multisyllabic and non-sonorant-ending

monosyllabic nonces along with sonorant-ending monosyllabic nonces could give us more insight regarding its default status.

We give below for comparison the 6 roots that have never been irregularized in Table 10 and the 6 most irregularized items with their regular and irregular forms and the number of participants who opted for one or the other in Table 11.

Table 10. Nonce Roots Never Produced with –Ir by the Participants

Nonce	Regular form	# of answers
nül	nül-er	37
pül	pül-er	38
şöl	şöl-er	38
mol	mol-ar	34
cil	cil-er	33
şil	şil-er	37

Table 11. Six Most Irregularized Nonce Verbs with the Number of -Ar and -Ir Answers

Nonce	Regular form	# of answers	Irregular form	# of answers
rur	rur-ar	15	rur-ur	21
vel	vel-er	19	vel-ir	18
cel	cel-er	22	cel-ir	17
ur	ur-ar	23	ur-ur	16
vör	vör-er	22	vör-ür	15
cur	cur-ar	25	cur-ur	16

In what follows, we will turn to the discussion of the so-called irregularization errors. Though they are few in number, we believe that they tend to present a picture of to what extent analogy appears to be affecting the outcome.

5.2 Irregularization errors: Is rhyme a relevant measure of similarity?

Our goal in the following analysis is to test the characteristics of the stems that have given rise to the use of –Ir forms, the so-called irregularizations. As we will lay out in the ensuing paragraphs, some nonce forms have triggered the use of –Ir more than the others, hence yielded a high rate of irregularization.

To account for the variation in the irregularization rates, Chi-squared tests of independence were run on R[10] and p-values were adjusted for multiple comparisons with the Bonferroni correction. We also put the experimental results in relation with statistics on aorist drawn from the BOUN Corpus (cf. Chapter 2), focusing on the frequencies of monosyllabic verbs in the aorist forms without extra preceding or proceeding suffix (cf. Appendix B) and the frequencies of some relevant patterns of consonants and vowels (cf. Appendix C). In the tables given in this section, we consistently use the grey color to indicate the cells discussed in the text, in some cases the highest frequency between the frequency for regulars and the frequency for irregulars we compared, and in other cases the variables that seem to vary along the -Ir use.

We first examine to what extent rhyme, the measure of similarity chosen a priori in English past tense studies, determines the -Ir use in the Turkish aorist. There are 16 possible rhyme templates as laid out in Table 12. The differences with respect to -Ir use among these 16 possible rhymes for our nonce roots proves to be significant (χ-squared = 246.0527, df = 15, p-value < 0.001).

[10] https://www.r-project.org/

Table 12. –Ir Use according to the Rhyme of the Nonce Root

RHYME	-Ir Use	# of nonces	Irregular forms	Total
UR	31%	11	129	411
EL	22%	8	69	315
ÜR	20%	10	75	382
IR	20%	12	88	430
ÖR	19%	15	114	591
AL	16%	5	31	198
UL	16%	11	65	415
İR	15%	11	63	427
ER	13%	6	30	226
OR	12%	9	41	335
IL	12%	18	82	669
OL	9%	8	28	301
ÖL	9%	15	51	595
AR	7%	4	10	142
İL	5%	11	20	415
ÜL	5%	14	27	539

Looking at the frequencies of rhyming monosyllabic verbs in Turkish and comparing the type and token frequencies of regulars and irregulars, we can distinguish 9 potential profiles of rhyme sets as given in the table below. As Table 13 illustrates only 4 of these potential sets are realized.

Table 13. Rhymes of Nonce Stems and their –Ir Use according to Type and Token Frequencies of Regular and Irregular Monosyllabics in Turkish

In Token **In Type**	**Irregular > Regular**	**Irregular = Regular**	**Irregular < Regular**
Irregular > Regular	UL (16%)		
Irregular = Regular	UR (31%), ÖR (19%), ÖL (9%)		
Irregular < Regular	EL (22%), AL (16%), ER (13%), OL (9%), AR (7%), İL (5%)		ÜR (20%), IR (20%), İR (15%), OR (12%), IL (12%), ÜL (5%)

We observe a great variability in these four profiles, suggesting that frequencies of existing verbs could explain participants' answers for only a limited number of rhymes. Indeed, if the type or token frequency of existing rhyming monosyllabics explained the -Ir use, irregularization should be high when there are more irregulars than regulars rhyming, and low when there are more regulars than irregulars rhyming, which is not always the case in our results.

To gain more insight on the effect of rhyme, we look more closely at these 4 profiles of rhyme sets. As laid out in Table 14, the rhyme set UL is the only one for which there is a rhyming irregular (*bul-* 'find' that gives *bul-ur*) but no rhyming regular (i.e., **cul-ar*), a situation that can potentially trigger a lot of irregularized nonce forms. But even though irregulars represent 100% of types and tokens rhyming in UL, the experimental -Ir use for nonces in this rhyme set is only moderate (16%).

Table 14. Frequencies of Regular and Irregular Monosyllabics Rhyming in UL

		Existing verbs with the same rhyme							
	Experiment	**Regular**		**Irregular**		**Total**		**Rate**	
RHYME	**-Ir Use**	**Type**	**Token**	**Type**	**Token**	**Type**	**Token**	**Type**	**Token**
UL	**16%**	0	0	1 bul	15262 15262	1	15262	100%	100%

On the contrary, there are 6 rhyme sets for which there is no rhyming irregular but 1 or 2 rhyming regulars such as ÜR in *sür-er 'drives/lasts'*, IR in *kır-ar 'breaks'*, İR in *gir-er* 'enters', or finally OR in *sor-ar* 'asks'. As shown in Table 15, for these rhyme sets irregulars represent 0% of the existing verbs in types and tokens. Therefore we expect a low -Ir use but this does not appear to be the case: these rhyme sets yield very different -Ir use, from 5% to 20%; the nonces for the rhymes ÜR and IR are

even more irregularized on average than the nonces for the rhyme UL, just mentioned above.

Table 15. Frequencies of Regular and Irregular Monosyllabics Rhyming in ÜR, IR, İR, OR, IL, ÜL

		Existing verbs with the same rhyme							
	Experiment	Regular		Irregular		Total		Rate	
RHYME	-Ir Use	Type	Token	Type	Token	Type	Token	Type	Token
ÜR	**20%**	2	6280	0	0	2	6280	0%	0%
		sür	6205						
		dür	75						
IR	**20%**	1	1250	0	0	1	1250	0%	0%
		kır	1250						
İR	**15%**	1	20000	0	0	1	20000	0%	0%
		gir	20000						
OR	**12%**	2	8117	0	0	2	8117	0%	0%
		sor	7633						
		yor	484						
IL	**12%**	2	2316	0	0	2	2316	0%	0%
		kıl	2131						
		yıl	185						
ÜL	**5%**	1	735	0	0	1	735	0%	0%
		gül	735						

Two other interesting cases appear when there are both rhyming regulars and irregulars in the lexicon. For 3 rhyme sets, there is as much rhyming regulars as rhyming irregulars (in type) but rhyming irregulars are more frequent (in token), so the type rate of irregulars is 50% but their token rate is higher. Here the higher the frequency of rhyming irregulars in type or token, the higher the observed -Ir use. For example, as the Table 16 clearly lays out, the rhyme UR has regular and irregular exemplars (i.e., *kur-ar* 'establishes' and *dur-ur* 'stops') and they differ from each other in respect of token frequency (4049 regular forms for 11398 irregular forms).

Table 16. Frequencies of Regular and Irregular Monosyllabics Rhyming in UR, ÖR, ÖL

		Existing verbs with the same rhyme							
	Experiment	Regular		Irregular		Total		Rate	
RHYME	-Ir Use	Type	Token	Type	Token	Type	Token	Type	Token
UR	**31%**	2	4049	2	11398	4	15447	50%	74%
		kur	4043	Dur	8094				
		bur	6	Vur	3304				
ÖR	**19%**	1	234	1	24154	2	24388	50%	99%
		ör	234	gör	24154				
ÖL	**9%**	1	490	1	5367	2	5857	50%	92%
		böl	490	öl	5367				

Finally, we are interested in the 6 rhyme sets presented in Table 17 for which rhyming regulars exceed rhyming irregulars in type, but in contrast rhyming irregulars are more frequent in token. For example the rhyme EL, although it has only one irregular (*gel-ir* 'comes') and 2 regulars (*del-er* 'drills', *çel-er* 'deviates') in type, has more irregular tokens (4212) than regular ones (270). This ambiguous situation yields very disparate -Ir uses (from 5% to 22%) that do not seem to be correlated with type or token frequencies. Hence, for instance participants heavily irregularized nonces rhyming in EL (22%) but not nonces rhyming in İL (5%) whereas the rhyme İL is in the same situation of having a single very frequent irregular (*bil-ir* 'knows') and 3 less frequent regulars (*sil-er* 'deletes', *il-er* 'knots', *dil-er* 'slices').

Table 17. Frequencies of Regular and Irregular Monosyllabics Rhyming in EL, AL, ER, OL, AR, İL

		Existing verbs with the same rhyme							
	Experiment	Regular		Irregular		Total		Rate	
RHYME	-Ir Use	Type	Token	Type	Token	Type	Token	Type	Token
EL	**22%**	2	270	1	4212	3	4482	33%	94%
		del	232	gel	4212				
		çel	38						
AL	**16%**	3	2291	2	69709	5	72000	40%	97%
		dal	1219	Al	46422				
		sal	799	kal	23287				
		çal	273						
ER	**13%**	4	2905	1	48774	5	51679	20%	94%
		er	2380	ver	48774				
		ser	256						
		ger	248						
		yer	21						
OL	**9%**	3	1063	1	239904	4	240967	25%	100%
		sol	904	Ol	239904				
		yol	159						
		dol	0						
AR	**7%**	3	1918	1	1998	4	3916	25%	51%
		sar	1917	var	1998				
		kar	1						
		yar	0						
İL	**5%**	3	1362	1	36313	4	37675	25%	96%
		sil	1164	bil	36313				
		il	189						
		dil	9						

We conclude that, contrary to what has been assumed/ postulated for the English past tense, rhyme does not explain consistently the recourse to the irregular form for the Turkish aorist. In what follows, we will attempt to tease apart the effect of variables other than rhyme that may play a role in the use of the –Ir form.

In their paper entitled after the important research question “What makes words sound similar?” Hahn and Bailey (2005) answer that it is the number of mismatching phonemes (onset, nucleus, coda) and even sub-phonemic features

(labiality, stopness, etc.) that matters. Also of interest, Hartshorne and Ullman (2006), in their re-analysis of children's overregularization errors (**holded* instead of *held*) in relation with similarity measures based on token frequency and monosyllabic verbs, found stronger correlations of overregularization rates with the final-coda measure than with the rhyme measure. Even though it could be due to a floor effect, since final-coda is included in rhyme so an overregularized verb has fewer rhyme neighbors than final-coda neighbors, the authors explain it in such a way that it could also be the case that overregularized verbs are generalized not from regulars similar in their entire rhyme, but from regulars similar in their coda. So starting by investigating measures of similarity at the phonemic level, we will successively test the effect of:

i. Final consonant of the nonce root
ii. Vowel of the nonce root
iii. First consonant of the nonce root
iv. C_C Pattern of the nonce root (first and last consonant)

5.3 Alternative similarity measures: What makes a nonce more irregularized?

5.3.1 Final consonant of the nonce root

Recall that in constructing the nonce roots we have restricted ourselves with the use of /l/ or /r/ as the final consonants of the CVC roots. As laid out in Table 18 below, we have observed that /r/-ending nonces have yielded –Ir use, hence irregularizations, significantly more than /l/-ending nonces (χ-squared = 79.4109, df = 1, p-value < 0.001).

Table 18. –Ir Use according to the Final Consonant of the Nonce

Final Consonant	-Ir Use	# of nonces	Irregular forms	Tokens
r	19%	78	550	2944
l	11%	90	373	3447

At first sight, one tends to think that this outcome may be correlated with the corpus frequency of /r/ and /l/-ending aorist attached monosyllabic verbs in Turkish. Neither the type nor the token frequency of the monosyllabic verbs in the language, however, appear to be correlated with the participants' answers, since /l/-ending irregular verbs are more frequent in type and token than /r/-ending irregulars as illustrated in Table 19.

Table 19. Frequencies of /r/ and /l/-ending Monosyllabic Verbs in Turkish

/r/-ending verbs				**/l/-ending verbs**			
Regular		**Irregular**		**Regular**		**Irregular**	
Type	**Token**	**Type**	**Token**	**Type**	**Token**	**Type**	**Token**
16	44753	5	86324	15	8527	7	370767
Gir	20000	ver	48774	kıl	2131	ol	239904
Sor	7633	gör	24154	dal	1219	al	46422
Sür	6205	dur	8094	sil	1164	bil	36313
Kur	4043	vur	3304	sol	904	kal	23287
Er	2380	var	1998	sal	799	bul	15262
Sar	1917			gül	735	öl	5367
Kır	1250			böl	490	gel	4212
Yor	484			çal	273		
Ser	256			del	232		
Ger	248			il	189		
Ör	234			yıl	185		
Dür	75			yol	159		
Yer	21			çel	38		
Bur	6			dil	9		
Kar	1			dol	0		
Yar	0						

What this brings to mind is that the frequency of trigrams /VrV/ and /VlV/ may differ and impact participants' answers. As illustrated in Table 20, the sequences /Vru, Vrü, Vrı, Vri/ corresponding to irregular forms (i.e., /örü/ present in *gör-ür* 'sees') are generally more frequent in type and token than /Vra, Vre/ corresponding to regular forms (i.e., /öre/ in *ör-er* 'knits'). As for sequences including the liquid/ sonorant /l/, we see that /Vla, Vle/ present in regulars (i.e., /ola/ in *sol-ar* 'fades') are more frequent than /Vlu, Vlü, Vlı, Vli/ present in irregulars (i.e., /olu/ in *ol-ur* 'is'). This tendency may be giving rise to a higher use of -Ir after /r/ than after /l/.

Table 20. Type and Token Counts of Trigram Patterns Corresponding to the Regular and Irregular Aorist Forms

Regular	**Type**	**Token**	**Irregular**	**Type**	**Token**
üre	6385	1078938	ürü	8807	956358
öre	2713	1043721	örü	5881	1179019
ura	8004	923661	uru	11918	2629747
ora	2728	703818	oru	13321	2768599
ire	15231	1105808	iri	26747	2567948
ere	18954	4045182	eri	96985	12515197
ira	14196	1189403	irı	22985	1409920
ara	35131	9244353	arı	98892	11307560
üle	7911	570151	ülü	4448	339727
öle	930	68022	ölü	2016	571984
ula	19745	2297231	ulu	10318	2598510
ola	8299	5513476	olu	5341	2010995
ile	67846	7538577	ili	43766	4575921
ele	42537	4155343	eli	24772	4032934
ila	39981	3331901	ilı	17182	2456780
ala	50328	4902995	alı	27693	3840091

Frequencies of corresponding quadrigrams shown in Table 21 confirm our expectations of finding more -Ir forms after /r/ than after /l/: most of the sequences /Vrur, Vrür, Vrır, Vrir/ are more frequent than /Vrar, Vrer/ in type and token while the sequences /Vlar, Vler/ are more frequent than /Vlur, Vlür, Vlır, Vlir/. For example, /örür/ is more frequent than /örer/, supporting the irregular *zör-ür* over the regular *zör-er* for the nonce z*ör-* irregularized at 31% by the participants, but /olar/ is more frequent than /olur/, this time supporting the regular *zol-ar* over the regular *zol-ur* irregularized at 3% by the participants.

Table 21. Type and Token Counts of Quadrigram Patterns Corresponding to the Regular and Irregular Aorist Forms

Regular	**Type**	**Token**	**Irregular**	**Type**	**Token**
ürer	101	17840	ürür	431	82528
örer	17	4437	örür	42	28444
urar	94	34863	urur	896	78315
orar	104	9286	orur	32	6088
irer	460	122006	irir	1853	84803
erer	87	42638	erir	1016	123140
ırar	417	33812	ırır	1966	52600
arar	3101	1239280	arır	281	21944
üler	3229	228714	ülür	359	37427
öler	56	2341	ölür	24	9069
ular	5789	726825	ulur	551	63547
olar	2371	2114722	olur	177	388649
iler	26390	2619574	ilir	15933	1051883
eler	24841	2420011	elir	2862	1378678
ılar	18042	1271196	ılır	1851	189123
alar	26496	2495636	alır	196	127404

Here we can hypothesize that the choice of -Ar or -Ir depend on the high frequency of the trigrams /Vru, Vrü, Vrı, Vri/ on one hand, and /Vla, Vle/ on the other hand. This does not exclude other factors such as more fine-grained frequency measures, to be tested in further studies: indeed, the presence of plural suffixes *-lar* and *-ler* (the most frequent suffix in Turkish) which partially overlap with the aorist affix could make the regular aorist form in -Ar more readily available for production after /l/-ending nonces. We now look at the vowel of the nonce stem to investigate its potential influence on the -Ir use.

5.3.2 Root vowel

Among the eight vowels that constitute the nucleus of the CVC we have observed a significant difference with respect to the –Ir use. While the high, back, rounded /u/ in the root rendered the most –Ir use, high, front, unrounded /i/ yielded the least, as illustrated in Table 22 (χ-squared = 93.387, df = 7, p-value < 0.001).

Table 22. -Ir Use according to the Vowel in the Nonce Root

VOWEL	-Ir use	# of nonces	Irregular forms	Total
u	23%	22	194	826
e	18%	14	99	541
ı	15%	30	170	1099
ö	14%	30	165	1186
a	12%	9	41	340
ü	11%	24	102	921
o	11%	17	69	636
i	10%	22	83	842

Thus when the features of the vowels are taken into consideration, the difference between front and back vowels in rendering the use of –Ir forms, hence irregularizations, appears to be significant (χ-squared = 15.4717, df = 1, p-value <

001) as illustrated in Table 23. More precisely, the -Ar suffix was always dominant but at different rates according to the backness of the stem vowel. For items containing back vowels (/u/, /o/, /ı/, /a/) the -Ir suffix surfacing as [-ur] or [-ır] was preferred over the -Ar suffix [-ar] by 16%. For items containing front vowels (/ü/, /ö/, /i/, /e/), the -Ir suffix surfacing as [-ür] or [-ir] was preferred over the -Ar suffix [-er] to a lesser extent, by 13%. For example, the nonce containing a back vowel *zur-* was irregularized as *zur-ur* with a rate of 32% while its front version the nonce *zür-* was irregularized as *zür-ür* only with a rate of 17%.

Table 23. -Ir Use according to Backness/Frontness of the Vowel in the Nonce Root

VOWEL	-Ir Use	# of nonces	Irregular forms	Total
back (u, o, ı, a)	16%	78	474	2901
front (ü, ö, i, e)	13%	90	449	3490

We have, however, observed no significant difference between rounded and unrounded vowels (χ-squared = 1.0884, df = 1, p-value = 0.4765), nor between high and non-high vowels with respect to -Ir vs. -Ar use (χ-squared = 1.3908, df = 1, p-value = 0.4765) as laid out in Table 24 below.

Table 24. -Ir Use according to the Roundedness/Height of the Vowel in the Nonce Root

VOWEL	-Ir use	# of nonces	Irregular forms	Total
rounded (ü, ö, u, o)	14%	93	530	3569
unrounded (i, e, ı, a)	15%	75	393	2822
high (ü, u, i, ı)	14%	98	549	3688
non-high (ö, o, e, a)	15%	70	374	2703

To address the question of whether the root vowel somehow influences the –Ir uses, we have looked into the ratio of regular/irregular types and tokens that consist of a

particular root vowel in monosyllabic verbs in Turkish. Table 25 below presents this information. The vowel /u/ for instance has 3 exemplars which behave irregularly i.e., *dur/durur, vur/vurur, bul/bulur*) as opposed to 14 exemplars that behave regularly such as *kur/kurar, sun/sunar, etc.* Hence we have observed type frequency to be 18% and token 33%.

Table 25. Frequencies of Regular and Irregular Monosyllabics Sharing the Same Vowel

		Corpus					
	Experiment	Verbs sharing the same vowel					
VOWEL		Type			Token		
		Regular	Irregular	Rate	Regular	Irregular	Rate
	-Ir Use						
u	**23%**	14	3	18%	54133	26660	33%
e	**18%**	30	2	6%	275422	52986	16%
ı	**15%**	21	0	0%	11578	0	0%
ö	**14%**	17	2	11%	12542	29521	70%
a	**12%**	55	4	7%	147605	112434	43%
o	**11%**	25	1	4%	35479	239904	87%
ü	**11%**	20	0	0%	22396	0	0%
i	**10%**	24	1	4%	48681	36313	43%

The -Ir use seems to increase with the type frequency of irregular Turkish monosyllabic verbs sharing the same vowel: the more existing irregular forms there are with the same vowel, the more irregularized a nonce would be.[11]

In an attempt to uncover whether any particular feature of the root vowel, i.e., backness vs. frontness has any effect on the –Ir use, we have also examined the type and token frequencies of Vowel-(Any)Consonant-Vowel (VCV) sequences in the entire language. All the possible VCV templates are listed in Table 26 below. As it is

11 A careful look into the distribution reveals however an exception, as there are four roots in Turkish with the vowel /a/ (*al-*, *kal- var-* and *san-*) that require the use of –Ir, however a low –Ir use, only 7%, is at issue in the case of roots that have the vowel /a/.

clear from the Table 26, both the type and token frequency rates of VCV templates are higher when the second vowel is a high vowel corresponding to –Ir, compared to when it is a non-high one corresponding to –Ar, regardless of the backness or frontness of the first vowel of V_V (except for the token frequency of the pattern öCe). These patterns cannot be responsible for the different -Ir uses observed according to backness.

Table 26. Type and Token Counts of the Patterns V_V Corresponding to the Regular and Irregular Aorist Forms

VOWEL	Regular	Type	Token	Irregular	Type	Token
u	uCa	64977	8635979	uCu	82987	16000198
o	oCa	35674	8800267	oCu	54753	12033980
ı	ıCa	106049	10524662	ıCı	236350	20936663
a	aCa	355760	49394054	aCı	379983	51093444
ü	üCe	35332	6643054	üCü	50424	8215233
ö	öCe	19286	5434041	öCü	20719	3474056
i	iCe	158165	21107365	iCi	288000	34247749
e	eCe	246693	25061126	eCi	376657	45976368

Hence, we can conclude that similarity of the vowel matters to irregularization: the higher the number of monosyllabic irregulars in Turkish sharing the same vowel for a nonce, the higher its -Ir use. We can also say that an inspection of vowel pattern frequencies in the whole language does not enable us explain the higher –Ir use we have seen in some nonces containing a back vowel. In the next section we will examine whether the initial consonant of the nonce has any role in triggering the –Ir use.

5.3.3 First consonant of the nonce root

Even though there is variability in items starting with the same consonant, some first consonants have triggered significantly more –Ir use than others: the -Ir use between the categories of nonces defined by their first consonant is significantly different (χ-squared = 108.3059, df = 19, p-value < 0.001) as laid out in Table 27. For example, for the nonce verbs the initial consonants of which is /v/, we have encountered an -Ir use of 25%.

Table 27. -Ir Use according to the First Consonant in the Nonce Root

FIRST CONSONANT	-Ir Use	# of nonces	Irregular forms	Total
v	25%	13	126	499
None (vowel-starting verb)	19%	9	64	344
b	19%	5	35	188
y	19%	5	37	194
r	16%	15	89	570
c	15%	14	81	524
g	15%	5	29	192
n	15%	12	70	452
z	15%	10	57	376
l	14%	13	71	502
h	13%	9	47	359
m	13%	10	50	375
f	12%	7	31	253
p	11%	7	30	266
ç	10%	7	27	262
d	10%	5	20	194
k	10%	1	4	42
ş	7%	13	37	494
s	6%	4	10	157
t	5%	4	8	148

More generally, we have found that the percentage of irregularizations appears to be correlated with corpus statistics, i.e., the more a nonce has existing irregular verbs

(type) or irregular verb occurrences (token) starting with the same first consonant (if it has some), the more it is irregularized by participants, as illustrated in Table 28. Also, the higher the proportion of irregulars among all the monosyllabic verb tokens starting with the same consonant (token rate), the higher the irregularity rate.[12]

Table 28. Frequencies of Regular and Irregular Monosyllabics Sharing the Same First Consonant

		Corpus						
	Experiment	Same first consonant						
		Regular		Irregular			Rate	
FIRST CONSONANT	**-Ir Use**	**Type**	**Token**	**Type**	**Token**	**List**	**Type**	**Token**
v	**25%**	0	0	3	54076	var,ver,vur	100%	100%
b	**19%**	17	33152	2	51575	bil,bul	11%	61%
y	**19%**	24	73195	0	0		0%	0%
None	**19%**	0	0	0	0		0%	0%
r	**16%**	0	0	0	0		0%	0%
c	**15%**	2	94	0	0		0%	0%
z	**15%**	0	0	0	0		0%	0%
g	**15%**	10	33585	2	28366	gel,gör	17%	46%
n	**15%**	0	0	0	0		0%	0%
l	**14%**	0	0	0	0		0%	0%
m	**13%**	0	0	0	0		0%	0%
h	**13%**	0	0	0	0		0%	0%
f	**12%**	0	0	0	0		0%	0%
p	**11%**	2	275	0	0		0%	0%
k	**10%**	29	36266	1	23287	kal	3%	39%
ç	**10%**	11	19786	0	0		0%	0%
d	**10%**	21	89559	1	8094	dur	5%	8%
ş	**7%**	2	1032	0	0		0%	0%
s	**6%**	39	55733	1	40727	san	3%	42%
t	**5%**	13	3168	0	0		0%	0%

12 This holds true if we neglect *san-*, the only irregular verb ending in /n/, for which we tested no /n/-ending nonce.

Consulting the frequencies of the irregulars starting with the same consonant seems compatible with the fact that the participants had to read the nonce on the screen to complete it. For example when reading a monosyllabic nonce, starting with the consonant /v/, they could be looking up in their lexicon for the "letter /v/" and see that all existing monosyllabics starting with /v/ are irregulars, which may give rise to the production of more irregular forms. We have to be cautious with this interpretation though as there are nonce forms such as the /y/-starting and /r/-starting forms which have given rise to irregularizations, though they do not have neighbors starting with the same consonant.

Thus, the use of the -Ir form for nonces starting with different first consonant seems to correlate with the type number of existing irregular verbs starting with the same consonant, just as it was also the case for the type number of existing irregular verbs sharing the same vowel in Section 5.3.2. We will now examine the effect of the pattern formed by the first and last consonants.

5.3.4 C_C Pattern (first and last consonant)

Although we did not have enough items in each category to make comparisons between patterns (some patterns were tested only for one item), the irregularity rate was found significantly different between categories of nonces sharing the same first and last consonant, even after a continuity correction was applied since more than 20% of the counts in the categories were less than 5 (χ-squared = 221.902, df = 38, p-value < 0.001), as laid out in Table 29.

Table 29. –Ir Use according to the C_C Pattern in the Nonce Stem

PATTERN C_C	-Ir Use	# of nonces	Irregular forms	Total
v_r	35%	5	70	201
y_r	25%	3	28	114
z_r	25%	4	35	142
l_r	22%	6	49	226
_r	21%	5	41	193
r_r	20%	8	59	291
b_l	19%	2	14	75
b_r	19%	3	21	113
c_r	19%	6	45	232
p_r	19%	3	20	106
v_l	19%	8	56	298
h_r	18%	3	21	120
g_r	17%	1	7	41
n_l	17%	6	40	232
f_r	16%	4	23	144
m_r	16%	6	37	230
_l	15%	4	23	151
g_l	15%	4	22	151
n_r	14%	6	30	220
c_l	12%	8	36	292
ç_r	12%	4	19	148
ş_r	12%	6	27	228
d_r	11%	3	13	118
y_l	11%	2	9	80
h_l	10%	6	26	239
k_l	10%	1	4	42
r_l	10%	7	30	279
d_l	9%	2	7	76
m_l	9%	4	13	145
z_l	9%	6	22	234
l_l	8%	7	22	276
ç_l	7%	3	8	114
f_l	7%	3	8	109
s_r	7%	1	3	44
p_l	6%	4	10	160
s_l	6%	3	7	113
t_r	6%	1	2	33
t_l	5%	3	6	115
ş_l	4%	7	10	266

Turning to frequencies in the Turkish language in Table 30, we find again that the more the nonce has existing irregulars in number (type) with the same pattern C_C (if any), the more it is irregularized (e.g., */v_r/, /b_l/, /g_r/, /g_l/, /k_l/, /d_r/*), similar to what was happening for first consonant and vowel. And the least irregularized nonces among those sharing the same pattern C_C with existing verbs (e.g., */d_l/, /s_r/, /ç_l/, /s_l/*) are found to be sharing it with no irregulars but with several regulars.[13] However, one must not forget that some nonces that do not share their first and last consonant with any irregulars (e.g., */y_r/, /b_r/, /y_l/*) are also highly irregularized and that many highly irregularized nonces (for example the top three *rur-*, *vel-*, *cel-*) do not share their pattern C_C with any existing monosyllabic verbs. For those, the first consonant only seems to have played a more important role.

The importance of the pattern C_C could reflect the fact that the item to be completed is presented visually, allowing the participant to perceive the nonce and make analogies based on the global shape of the word. This could be all the more relevant since the visual difference between a nonce and an irregular can be as small as a dot or an umlaut (*bıl- / bil-*, *bül- / bul-*, *vür- / vur-*, *gor- / gör-*).

13 The token counts seem to follow the same order except for the irregular verb *gel* 'come': we note that its number of tokens in the aorist is quite low in the corpus whereas the verb is actually very frequent in the language. It could be caused by using a corpus based on the compilation of websites, and not an oral corpus where the verb expressing the action of "coming" may be much more frequent; but this does not sound very convincing since it also has some frequent abstract usages such as *anlamına gelir* 'it means' or *bana gelir* 'it seems to me'.

Table 30. Frequencies of Regular and Irregular Monosyllabics Sharing the Same Pattern C_C

Experiment		Corpus: Same first and last consonants									
PATTERN C_C		Regular			Irregular			Total		Rate	
	-Ir Use	Type	Token	List	Type	Token	List	Type	Token	Type	Token
v_r	**35%**	0	0		3	54076	var,ver,vur	3	54076	100%	100%
y_r	**25%**	3	505	yar,yer,yor	0	0		3	505	0%	0%
b_r	**19%**	1	6	bur	0	0		1	6	0%	0%
b_l	**19%**	1	490	böl	2	51575	bil,bul	3	52065	67%	99%
g_r	**17%**	2	20248	ger,gir	1	24154	gör	3	44402	33%	54%
g_l	**15%**	1	735	gül	1	4212	gel	2	4947	5%	85%
k_l	**10%**	1	2131	kıl	1	23287	kal	2	25418	5%	92%
y_l	**11%**	2	344	yıl,yol	0	0		2	344	0%	0%
d_r	**11%**	1	75	dür	1	8094	dur	2	8169	5%	99%
d_l	**9%**	4	1460	dal,del,dil,dol	0	0		4	1460	0%	0%
s_r	**7%**	4	16011	sar,ser,sor,sür	0	0		4	16011	0%	0%
ç_l	**7%**	2	311	çal,çel	0	0		2	311	0%	0%
s_l	**6%**	3	2867	sal,sil,sol	0	0		3	2867	0%	0%

5.4 Summary

Let us wrap up the previous findings by looking at couples of variables simultaneously:

i. First consonant C + Rhyme VC

ii. Pattern C_C + Vowel V

Table 31. –Ir Use according to Rhyme and First Consonant

Rhyme		First consonant																			
Vowel	Last	v	y		b	r	z	n	c	g	l	m	h	f	p	d	ç	k	ş	s	t	Total
u	r	vur	21 %	41 %		58 %	32 %		39 %		31 %	26 %	28 %		26 %	dur	29 %		15 %			31 %
e	l	49 %				18 %	16 %		44 %	gel	14 %		17 %		14 %				7 %			22 %
ı	r	24 %		28 %	20 %	33 %	20 %	18 %	13 %		27 %	24 %				21 %	8 %		15 %			21 %
ö	r	41 %	29 %		18 %	7 %	31 %	21 %	19 %	gör	20 %	16 %	13 %	31 %	21 %	7 %	11 %		12 %			20 %
ü	r	38 %		13 %		8 %	17 %	18 %	26 %		26 %	11 %		19 %					16 %			19 %
u	l	31 %		24 %	bul	13 %	8 %	31 %	6 %	14 %	13 %	21 %	10 %						3 %			16 %
a	l	12 %		al		13 %	12 %		5 %	35 %								kal				15 %
i	r	36 %	23 %	14 %		25 %		13 %	9 %				12 %	8 %					5 %	7 %	6 %	14 %
e	r	ver			18 %	14 %		11 %			15 %	16 %					7 %					14 %
ı	l	20 %		30 %	24 %	20 %	3 %	13 %	19 %	6 %	8 %	9 %	15 %	12 %	8 %	11 %	7 %		5 %	5 %	5 %	12 %
o	r	35 %		11 %		13 %		3 %	8 %	17 %				4 %	9 %	8 %						12 %
o	l	9 %		ol			3 %	37 %	5 %		3 %	0 %					5 %		8 %			9 %
ö	l	10 %	7 %	öl		5 %	14 %	16 %	10 %		12 %	5 %	18 %	6 %	3 %			10 %	0 %	8 %	5 %	9 %
a	r	var				7 %					9 %	5 %							8 %			7 %
ü	l	10 %		5 %	12 %	2 %		0 %	8 %		3 %		3 %	5 %	0 %	7 %	8 %		3 %	5 %		5 %
i	l	11 %	16 %	3 %	bil	5 %		3 %	0 %	3 %	4 %		3 %						0 %		5 %	5 %
Total		25 %	19 %	19 %	18 %	16 %	16 %	15 %	15 %	15 %	14 %	13 %	13 %	12 %	12 %	11 %	11 %	10 %	7 %	6 %	5 %	14 %

We observe in Tables 31 and 32:

- Higher irregularization rates for /v/, /b/- and non-consonant-starting nonces: the similarly starting verbs have higher number of irregulars in type (2 or 3) than any other consonant-starting verbs (1 or 0 except for /g/)
- Higher irregularization rates for nonces rhyming in UR: 2 existing irregular verbs rhyme in UR whereas for other rhymes (apart from the rhyme AL) only one or zero irregular verb exists.
- Quite high irregularization rates around the existing irregulars in general but not so much around *bil*-and *ol*-[14], or *var*-and *kal*-.[15]
- Some highly irregularized nonces out of the two-dimensional gradients and the attraction points created by existing irregulars, potentially confounded by some idiosyncratic properties of the language.[16]

14 At first sight it is surprising because these two are used to create many other verbs like *yapabil* or *kaybol* that take -Ir, both because they are multisyllabic and because they are based on these roots.

15 Our way of generating stimuli did not allow us to test nonces close to *kal* along these dimensions.

16 Some consonants exhibit variability in their pronunciation according to a particular dialect of Turkish used in a particular province, such as /k/ sometimes pronounced /g/, or /g/ sometimes

Table 32. –Ir Use according to Pattern C_C and Vowel

	VOWEL								
PATTERN C_C	u	e	ı	ö	a	ü	o	i	Total
v_r	vur	ver	24 %	41 %	var	38 %	35 %	36 %	35 %
z_r	32 %		20 %	31 %		17 %			25 %
y_r	21 %			29 %				23 %	24 %
_r	41 %		28 %			13 %		14 %	24 %
l_r	31 %	15 %	27 %	20 %	9 %	26 %			21 %
r_r	58 %	14 %	33 %	7 %	7 %	8 %	13 %	25 %	21 %
c_r	39 %		13 %	19 %		26 %	8 %	9 %	19 %
v_l	31 %	49 %	20 %	10 %	12 %	10 %	9 %	11 %	19 %
b_r		18 %	20 %	18 %					19 %
p_r	26 %			21 %			9 %		19 %
b_l	bul		24 %			12 %		bil	18 %
h_r	28 %			13 %				12 %	18 %
g_r				gör			17 %		17 %
n_l	31 %		13 %	16 %		0 %	37 %	3 %	17 %
m_r	26 %	16 %	24 %	16 %	5 %	11 %			16 %
_l	24 %		30 %	öl	al	5 %	ol	3 %	16 %
f_r				31 %		19 %	4 %	8 %	16 %
g_l	14 %	gel	6 %		35 %			3 %	15 %
n_r		11 %	18 %	21 %		18 %	3 %	13 %	14 %
ç_r	29 %	7 %	8 %	11 %					14 %
c_l	6 %	44 %	19 %	10 %	5 %	8 %	5 %	0 %	12 %
d_r	dur		21 %	7 %			8 %		12 %
ş_r	15 %		15 %	12 %	8 %	16 %		5 %	12 %
y_l				7 %				16 %	12 %
h_l	10 %	17 %	15 %	18 %		3 %		3 %	11 %
r_l	13 %	18 %	20 %	5 %	13 %	2 %		5 %	11 %
k_l				10 %	kal				10 %
z_l	8 %	16 %	3 %	14 %	12 %		3 %		9 %
d_l			11 %			7 %			9 %
m_l	21 %		9 %	5 %			0 %		9 %
l_l	13 %	14 %	8 %	12 %		3 %	3 %	4 %	8 %
f_l			12 %	6 %		5 %			8 %
s_r								7 %	7 %
ç_l			7 %			8 %	5 %		7 %
p_l		14 %	8 %	3 %		0 %			6 %
s_l			5 %	8 %		5 %			6 %
t_r								6 %	6 %
t_l			5 %	5 %				5 %	5 %
ş_l	3 %	7 %	5 %	0 %		3 %	8 %	0 %	4 %
Total	24 %	19 %	16 %	14 %	12 %	11 %	10 %	10 %	14 %

pronounced /dj/ (written /c/). Therefore *gal* could be highly irregularized because it sounds like a version of the irregular *kal* 'stay', or *cel* because it sounds like a version of the irregular *gel* 'come'. Also, the nonce *nol* gives *nolur* when conjugated irregularly and sounds very much like the frequently used phrase *n'olur*, the contraction of *ne olur* 'what happens'.

The type frequency of irregular verbs seems to enter into account: for example the irregular *vur-* is at the conjunction of the highest type frequency of irregulars starting with a same first consonant (3) and the highest type frequency of irregulars rhyming (2). It is also at the conjunction of the highest type frequency of irregulars sharing the same first and last consonant (3) and the highest type frequency of irregulars sharing the same vowel (3).[17] A tentative account of the results in terms of Islands of Reliability (Albright and Hayes, 2003) is provided in Appendix F. Although rules derived from the experimental data at the phonemic level were not precise enough to yield conclusions, we believe it suggests that a model deriving rules at the sub-phonemic level could reach interesting results.

In summary, we have found that for the Turkish aorist, regular forms in -Ar are dominant for nearly all nonces. As for irregular forms in -Ir, they do not seem to reflect the frequencies of rhyming regular and irregular verbs in Turkish. Importantly, forms in -Ir are used more with nonces ending in /r/ than with nonces ending in /l/, which appears to follow the type or token frequencies of the corresponding trigrams /VlV/ and /VrV/ and even quadrigrams /VlVr/ and /VrVr/ in the language. Furthermore, participants' recourse to -Ir forms seems to vary with the type frequency of existing irregulars sharing the same Vowel, First Consonant and Pattern C_C, if there are any. It should be underlined that these variables appear to follow the same trends but are proposed as potential correlations and not causes. We now turn in Chapter 6 to the discussion of the presented results.

17 But this reasoning also has limits:

- why g-starting verbs are not more irregularized since 2 irregulars start in /g/?
- why nonces rhyming in *AL* are not more irregularized since 2 irregulars rhyme in *AL*?
- why vowel-/l/ verbs are not more irregularized since 3 irregulars are of type Vowel-/l/? It could be explained if there was a separate entry for each vowel, and not a common entry "vowel"
- why verbs containing the vowel /a/ are not more irregularized since 3 irregulars contain the vowel /a/?

CHAPTER 6

DISCUSSION AND CONCLUSIONS

The experimental study presented in this thesis opens a path for future research as it provides the first behavioral results for adults for the aorist pattern in Turkish. Complemented by the acquisition data, it helps us draw a more complete picture of the representation of this pattern in native speakers' brains.

Recall that the acquisition work on the aorist (Nakipoğlu and Ketrez, 2006) demonstrates that at 3 years of age, children learning Turkish make more overregularization errors choosing -Ar for -Ir taking monosyllabic verbs (**öl-er* instead of *öl-ür*). From 4 years onwards, children make more irregularization errors using -Ir for -Ar taking verbs (**bin-ir* instead of *bin-er*), and do so till the complete acquisition of the irregular pattern around 8 years of age. At the adult stage when tested on nonce roots that resemble sonorant ending monosyllabic verbs in Turkish for nonce items, the present results show that the -Ar form is dominant (86%) overall and for all nonces but one.

We observe that in Turkish the suffix -Ar is used in more monosyllabic verbs than -Ir, and also that it is generalized to a wider range of situations for nonces by adults. Children make more overregularization errors using -Ar for -Ir taking verbs in their youngest productions: all these arguments contribute to the claim that -Ar is the default form of the Turkish aorist. Therefore, adults could parsimoniously resort to a rule assigning -Ar by default to the verbs, but this rule seems more probabilistic than deterministic since -Ir forms surfaced for at least one participant for 162 out of 168 nonces.

When we examined the -Ir use in more detail, we have observed that –Ir forms crucially do not seem to be driven by the rhyme with existing items -which was used as a predefined measure of similarity in previous studies on the English past tense. Among the five variables we have looked at Last Consonant, Vowel of the root, First consonant and C_C Pattern appear to play a more active role in relating nonces with existing features of the Turkish language compared to the rhyme variable.

In particular, -Ir use for nonces seem to be correlated with the type frequency of existing irregulars sharing the same vowel _V_ , the same first consonant C__ or the same C_C pattern: for a given nonce, if any such irregulars exist in the language, the higher the number of types, the higher the -Ir use. Interestingly, we have observed a correlation between high -Ir use, high type frequency of irregular verbs but no relation between frequency of regular verbs and lower -Ir use.

The Last Consonant of the nonces seem to have a decisive influence on the outcome: the -Ir forms appear significantly more with nonces ending in /r/ (19%) than with nonces ending in /l/ (11%). In this case, participants' answers do not reflect the frequencies of existing monosyllabic verbs, but the type and token frequencies of patterns /VrV/ and /VlV/ in Turkish. Finally, the frontness of the root Vowel has also impacted the results: -Ir is used more with nonces including back vowels [u, o, ı, a] (16%) than nonces including front vowels [ü, ö, i, e] (13%). This tells us that [-ar] and [-er] are preferred in general, but [-ur, -ır] appear to win the competition over [-ar] more often than [-ür, -ir] over [-er].

Our study takes place among recent studies investigating irregular patterns other than the English past tense. A first example is the case of Romance languages,

such as Italian or Portuguese, where verb stems consist of a verb root, to which is added a theme vowel, -a for 1st conjugation, -e for 2nd, and -i for 3rd conjugation. The first person present tense indicative is ambiguous as to which conjugation the verb belongs to in that it does not display the theme vowel.

Say and Clahsen (2002) tested Italian nonce first person present forms that rhymed either almost exclusively with regulars, almost exclusively with irregulars, or with no Italian verbs. After presenting the nonce verb in first person, they asked participants to complete the test sentences with a participle form and that way to choose a theme vowel and a conjugation class for the verb. Whether the nonce was rhyming with existing verbs or not, the 1st conjugation was predominantly chosen, revealing the existence of a default form. Meanwhile, 2nd and 3rd conjugations were produced only for nonces highly similar to existing 2nd and 3rd conjugation verbs, suggesting recourse to analogy for the irregular verbs.

Questioning the nature of stimuli typology based on rhyme - predefined measure of similarity - Verissimo and Clahsen (2014) tested first person forms in Portuguese, this time constructing some nonce forms similar to 1st, 2nd or 3rd conjugation, according to the confidence scores attributed by the Minimal Generalization Learner model of Albright and Hayes (2003) applied to the Portuguese verb lexicon. A nonce first person such as *acuo* could then elicit 1st conjugation *acuar*, 2nd conjugation *acuer* or 3rd conjugation *acuir*. The study yielded the same results as for Italian, again that for every nonce the 1st conjugation was preferred and the higher its similarity with the 2nd or 3rd conjugation according to reliability scores, the higher the use of 2nd or 3rd conjugations.

Finally, a recent experiment by Nemeth, Janacsek, Turi, Lukacs, Peckham, Szanka, Gazso, Lovassy and Ullman (2015) investigates Hungarian, an agglutinative

language with complex morphology just as Turkish. In Hungarian, the past tense for regular verbs is formed with suffixation (*hamoz-hamozott* 'peel-peeled'), while the past tense for irregular verbs requires phonological changes in the stem before suffixation (*vastagszik-vastagodott* 'thicken-thickened'). Thus there are both regular items attached the affix –ott and irregular items undergoing a phonological change before being attached the same affix. Participants who are asked to conjugate existing regular and irregular verb stems as quickly and as accurately as possible, proved to be more accurate and quicker in producing regular past forms than irregular past forms. This suggests a clear distinction between regulars and irregulars, it has however only been established for existing verbs and not tested on nonce items.

Our findings, as well as those in Italian, Portuguese and Hungarian, seem to substantiate the theory of the Dual-mechanism, where regulars are computed by a rule and irregular verbs by similarity. A tentative look at our data from the perspective of "Islands of Reliability" concept introduced by Albright and Hayes (2003), has not shed more light on the -Ir use for nonce verbs. However, in this study, the variables we examined were based on sound (= letter) change, but results indicate that sound feature-oriented variables (backness, labiality, etc.) could reasonably be looked at and that investigating more fine-grained frequency counts in the language is definitely worthy.

A path worth exploring to make broader conclusions is to extend our investigation to Turkish multisyllabic verbs, examining frequencies of existing multisyllabic verbs as well as pattern frequencies in these verbs in the language, and testing multisyllabic nonce verbs. In the acquisition of aorist, a growing lexicon

gives rise to more -Ir use as the child gets to observe more -Ir exemplars in her language input. Verbs also grow bigger in size, and become multisyllabic due to the attachment of suffixes proceeding the aorist (causative, passive, etc.), so even more -Ir exemplars are heard. Inspecting multisyllabics would complete our present results for monosyllabics to demonstrate that at the adult level in Turkish there is clear dissociation between the requirements of the monosyllabic and multisyllabic roots.

Another interesting point to investigate can be sources of individual variability such as age, handedness and gender. Hartshorne and Ullman (2006) have importantly pointed out that this last factor impacts children's errors: girls overregularize (*hold-*holded* instead of *held*) more than boys, and they do so all the more for items with a high number of similar-sounding regulars. This could be due to the fact that girls tend to generalize by having neighboring regulars more strongly stored in the associative lexical memory, while boys are more likely to generalize by overapplying the rule. In preliminary tests, gender does not appear to be a relevant explanatory variable for our results though, since we did not observe a significant difference of the -Ir use between female (15%) and male participants (14%) (χ-squared = 0.8032, df = 1, p-value = 0.3701).

A follow-up to the present study could test the retrieval times for regular and irregular aorist forms, as well as the reaction times to produce aorist forms for nonces varying in their similarity to existing verbs. Since our current results seem to support the theses of the Dual-mechanism model to account for the Turkish aorist, we hypothesize that different reaction times should be observed for regular and irregular forms. The former would be computed by the rule attaching the default -Ar to the stem, and the latter derived by similarity from irregular items in the lexicon.

To conclude, the results of the present study, we hope will contribute to the existing literature on the issue, and more importantly broaden the debate on irregularity and representation of forms by inspecting a pattern with different properties than the English past tense. Such cross-linguistic work constitute a gold standard against which predictions of upcoming computing models can be tested. Following the approach of Verissimo and Clahsen (2014) for Portuguese, an implementation of the Minimal Generalization Learner model could now be proposed and compared to similarity-driven models on their predictions of human performance for nonces, with the aim of rebuilding the mind's response to irregular patterns from their output.

APPENDIX A

METHODOLOGY FOR FREQUENCY COUNTS

We have worked with the BOUN Corpus (Sak, Güngör and Saraçlar, 2008), the biggest Turkish web corpus available, and more particularly with the version of this corpus made available by Zargan Ltd. on the following page: http://st2.zargan.com/duyuru/Zargan_Linguistic_Resources_for_Turkish.html.

It contains 1,337,898 word-forms and their stems, morphological analyses and frequencies, presented as such:

Word-form	Stem	Morphological analysis	Frequency
e.g. çöküntünün	çöküntü	Noun+A3sg+Pnon+Gen[3]	309

Importantly, the BOUN corpus is originally case-sensitive and contains some items whose morphological analysis could not be computed automatically. For these reasons, in the file available on Zargan website, all items have been normalized to lowercase and the items marked as "undef" have been discarded.

We perform several treatment operations on the corpus: we use a first script proposing a solution to merge the lines concerning the same entity with and without apostrophes, still compatible with the high number of lines of the file, and obtain a list of 1 260 882 items.

Then we select only the verb forms and obtain a list of 748 051 items. It should be noted that this number is only approximate since our selection of verb forms consists in filtering the items with the label "Verb" in their morphological analysis and this method can create false positives such as *dalıcı* analyzed as a noun derived from a verb. Since we make our further calculations on the aorist forms only,

for which the morphological marking is consistent, we decide to keep this number of verb forms, keeping in mind that it is only indicative.

Then we filter the aorist positive forms by selecting the aorist forms (containing the label “Aor”) then discarding the negative forms (containing the label “Neg”), the morphological analysis being precise and robust enough in this case. In an additional fifth column, we compute the aorist form without the suffixes that can appear after the -r/Ar/Ir aorist suffix (person markers, past -DI, conditional -sA, adverbial “while” -ken, narrative -mIş, adverbial “as if” -CAsInA, copula -DIr and combinations like narrative + adverbial “as if” -mIşçAsInA).

For this purpose, again we make use of the labels of the morphological analysis, and regular expressions to take into account the different forms of person markers e.g. *n(i|ı|u|ü)z$* for the four forms of the 2nd person plural suffix -nIz. We also use linguistic knowledge to strip the suffixes in the correct order, since the person marker is not present in combination with -ken, -mIş or -CAsInA, and present after the past, conditional or aorist suffix, but before the copula -DIr. The positive aorist list obtained contains 44 972 items.

We can describe further the aorist in the corpus, by looking at the repartition of forms according to the suffixes occurring before and after the aorist. Starting with the suffixes occurring after the aorist, and regardless of the suffixes occurring beforehand, we count the number of aorist forms followed respectively by the past -DI, the conditional -sA, the adverbial “while” -ken, the narrative -mIş, the adverbial “as if” -CAsInA, the copula -DIr and no extra suffix (verb forms ending with the aorist). Since we inspect the presence of these suffixes in this order, we note that a combination of two suffixes such as the narrative + adverbial “as if” -mIşçAsInA

will be counted under the category of aorist occurring followed by the narrative -mIş, which means our counts reflect the suffix directly following the aorist.

Turning to the suffixes occurring before the aorist, and this time regardless of the suffixes occurring afterwards, we count the number of aorist forms preceded by the ability -Abil, the adverbial “hastily” -Iver, the adverbial “ever since” -Agel, the passive, the causative and no suffix (aorist directly attached to the verb root). Again, our inspection of the suffixes follows the cited order (e.g. the causative + ability on the form of -DIrAbil for example, counted under the category of aorist occurring after ability), so our counts reflect the suffix directly preceding the aorist.

We also decide to look at the list once the suffixes after the aorist (including person markers) have been striped: merging the lines representing the same aorist form, we obtain a list of 10 530 items. 2 984 of these items are forms where the aorist is directly attached to a verb root.

APPENDIX B

FREQUENCIES OF MONOSYLLABIC VERBS IN TURKISH

Table B1. Token Counts of Aorist Forms of Monosyllabic Verbs in the BOUN Corpus (Irregulars in Grey)

Verb root	(Prefix +) Aorist (+ Suffix)	Aorist (+ Suffix)	Only Aorist
al	101083	80739	46422
bil	55852	39468	36313
bul	34465	22470	15262
dur	19657	15163	8094
gel	45564	30058	4212
gör	66041	32557	24154
kal	44899	37561	23287
ol	558510	383019	239904
öl	15488	9069	5367
san	42741	41915	40727
var	4913	2801	1998
ver	129857	78792	48774
vur	6702	4930	3304
aç	34402	17206	12981
ağ	58	57	35
ak	873	815	49
an	2700	675	246
art	21108	16021	9141
as	2043	1141	697
aş	5686	3431	2127
at	26713	18755	10138
az	115	13	5
bak	41958	32631	15634
ban	25	22	13
bas	5902	3644	2641
bat	1243	1158	393
bay	25	20	2
bez	182	45	41
bık	189	126	111
biç	1513	1373	1275
bin	2089	1483	431

bit	11684	11114	8495
boğ	1384	545	380
boz	8119	3345	2774
böl	2031	681	490
bur	30	7	6
burk	99	30	22
bük	683	485	434
büz	92	10	10
cay	206	88	48
coş	263	143	46
çak	425	116	19
çal	3391	2022	273
çarp	2732	1763	1245
çat	277	244	189
çek	29727	19445	11292
çel	90	77	38
çık	50466	16976	2715
çırp	282	157	76
çiz	2641	1434	239
çöz	8764	3309	2490
çök	1857	1672	1210
dal	1843	1616	1219
de	145512	107731	56292
del	517	291	232
deş	70	61	39
dik	1292	397	181
dil	134	17	9
din	492	214	199
dit	14	13	13
diz	683	281	185
doğ	6178	5386	4180
dol	3390	12	0
don	872	439	336
doy	247	199	164
dök	3762	1568	987
dön	13053	9361	1621
döv	1801	1457	1044
duy	15388	12030	10172
dür	99	80	75

dürt	65	49	41
düş	22196	19901	12554
düz	104	18	16
eğ	1014	207	83
ek	3291	2946	2696
em	739	463	310
er	3925	3393	2380
es	205	0	0
eş	58	53	45
et	354794	246383	178597
ez	1540	939	849
geç	23729	17286	2972
ger	789	315	248
gez	5445	3768	573
gir	41850	33749	20000
git	44255	34913	6948
giy	4322	2934	1642
göç	96	78	22
göm	1177	523	409
gül	3015	2111	735
güt	185	142	36
iç	5694	3755	2267
il	227	212	189
in	12796	7241	3592
it	1882	1257	960
iv	37	26	21
kaç	3935	3309	550
kak	110	76	68
kalk	10003	9186	5614
kan	604	17	15
kap	2233	0	0
kar	7	1	1
kas	377	244	223
kat	14868	5017	3877
kay	1217	462	24
kaz	352	256	158
kert	5	3	2
kes	7216	1743	668
kıl	4177	2814	2131

kır	4668	2003	1250
kırk	2	0	0
kırp	281	234	167
kıs	546	260	138
kıy	469	305	228
kız	2077	1436	1316
kok	1003	983	800
kon	5045	1013	494
kop	1608	1461	1340
kork	6237	5706	5111
koş	3239	2459	511
kov	781	530	389
koy	15334	11776	6702
kur	15895	6465	4043
kus	296	200	135
küs	448	353	311
on	1	0	0
ov	90	47	20
oy	321	250	216
ölç	2064	221	79
öp	2569	2364	1906
ör	443	322	234
ört	1022	691	597
öt	319	287	236
öv	764	653	230
piş	506	502	274
pus	2	1	1
saç	583	399	306
sağ	103	74	46
sal	2703	895	799
sap	450	254	151
sar	4491	2193	1917
sark	381	254	167
sars	1476	540	409
sat	9206	4789	2968
sav	140	135	124
say	25140	4380	2985
seç	15588	7654	3680
sek	711	684	601

ser	648	405	256
serp	343	94	68
sev	16314	15002	10804
sez	204	64	14
sıç	24	22	14
sığ	742	552	510
sık	1619	1382	1004
sız	404	264	197
sik	65	45	38
sil	3586	1536	1164
silk	81	69	62
sin	520	168	139
sok	4800	3157	2287
sol	1066	982	904
sor	17922	13485	7633
soy	576	373	187
sök	1241	470	329
sön	1120	691	563
söv	144	124	95
sun	13058	8842	6910
sus	1398	1091	857
sün	56	45	45
sür	40627	17958	6205
sürç	25	22	15
sürt	49	38	23
süs	3	3	3
süz	2052	1332	1254
şiş	339	316	285
şaş	956	906	747
tak	2251	1899	1251
tap	396	362	225
tart	461	370	279
taş	248	210	33
tat	554	161	94
tep	426	282	244
tık	37	0	0
tın	29	29	29
tırs	30	28	21
toz	313	303	280

tut	13902	4621	505
tüt	328	288	178
tüy	34	33	29
uç	1363	1177	263
um	30914	30516	30349
uy	1604	282	229
ürk	228	208	189
üz	2692	143	35
yağ	1899	1505	758
yak	4244	3349	2826
yan	4210	4079	3293
yap	199153	109591	55616
yar	93	0	0
yas	631	629	523
yat	5537	5270	3003
yay	3344	681	473
yaz	20168	10385	987
ye	2148	288	286
yen	8454	4729	1966
yer	104	79	21
yet	1147	916	125
yığ	316	82	46
yık	3838	1783	1146
yıl	342	276	185
yırt	679	411	294
yit	87	83	73
yol	238	205	159
yont	117	102	82
yor	1379	644	484
yum	368	257	161
yut	1365	660	480
yüz	1039	747	208

APPENDIX C

CONSONANT AND VOWEL PATTERN FREQUENCIES IN TURKISH

Table C1. Type and Token Frequencies of the "Vowel-Consonant-Vowel" Patterns in the Entire Language, in the Verb Forms and in the Aorist Forms

Pattern VCV	**In the whole language**		**In the verb forms**		**In the aorist forms**	
	Type	**Token**	**Type**	**Token**	**Type**	**Token**
aCa	355760	49394054	224895	18380922	3510	407197
aCe	36100	6962237	3074	48239	130	1299
aCi	99330	11020661	35938	1190450	3760	471426
aCı	379983	51093444	191748	18333300	4034	620436
aCo	9292	1019312	184	2913	10	64
aCö	2132	145574	45	94	0	0
aCu	12238	1895549	3566	165306	129	4664
aCü	2240	356101	5	11	0	0
eCa	25754	4509564	2413	75337	107	3612
eCe	246693	25061126	159514	10104036	2330	399624
eCi	376657	45976368	180897	13993756	6002	879456
eCı	23	641	0	0	0	0
eCo	7142	1152608	335	13381	16	440
eCö	562	115066	29	64	0	0
eCu	1775	175602	206	688	7	11
eCü	942	157441	0	0	0	0
iCa	39888	8501688	3321	198461	102	9315
iCe	158165	21107365	65482	4782516	1565	182589
iCi	288000	34247749	118160	7670150	11957	1243262
iCı	35	1033	0	0	0	0
iCo	30947	3651337	19777	2647477	12	32
iCö	742	50802	0	0	0	0
iCu	682	57417	0	0	0	0
iCü	1374	111872	0	0	0	0
ıCa	106049	10524662	54136	4222085	1630	131487
ıCe	3334	66192	2415	10079	286	947
ıCi	317	132242	60	253	0	0

ıCı	236350	20936663	103626	7726824	3806	316683
ıCo	20040	1986618	17523	1922789	7	109
ıCö	148	20181	0	0	0	0
ıCu	127	16028	0	0	0	0
ıCü	69	5054	0	0	0	0
oCa	35674	8800267	13419	4206184	584	197202
oCe	8410	1309474	108	643	5	5
oCi	10490	1837353	5	5	0	0
oCı	15	252	0	0	0	0
oCo	10407	1488707	223	2440	8	44
oCö	203	23434	0	0	0	0
oCu	54753	12033980	25616	4228800	621	327350
oCü	1622	283261	0	0	0	0
öCa	764	117809	0	0	0	0
öCe	19286	5434041	10130	741153	396	49387
öCi	375	130769	0	0	0	0
öCı	6	255	0	0	0	0
öCo	350	35001	0	0	0	0
öCö	11	177	0	0	0	0
öCu	70	1108	38	107	2	4
öCü	20719	3474056	11613	992937	407	70126
uCa	64977	8635979	34020	2766375	1134	156371
uCe	4634	461814	815	4626	94	632
uCi	3225	443920	0	0	0	0
uCı	75	4537	0	0	0	0
uCo	4859	968622	4022	794507	0	0
uCö	85	1527	0	0	0	0
uCu	82987	16000198	34307	6378486	1542	238810
uCü	828	638849	0	0	0	0
üCa	4331	373140	80	226	5	10
üCe	35332	6643054	17517	1512216	693	74035
üCi	2504	531694	54	324	2	7
üCı	15	608	0	0	0	0
üCo	3508	620859	2710	535521	0	0
üCö	181	5271	0	0	0	0
üCu	1117	144035	0	0	0	0
üCü	50424	8215233	23502	2412381	1202	142675

Table C2. Type and Token Frequencies of the "Vowel-Sonorant-Harmonic Vowel" Patterns (where Sonorant Stands for /l/, /n/ or /r/) in the Entire Language, in the Verb Forms and in the Aorist Forms

Pattern VSV	In the whole language		In the verb forms		In the aorist forms	
	Type	Token	Type	Token	Type	Token
ala	50328	4902995	26984	2023868	620	41128
alı	27693	3840091	14651	2227150	214	116087
ana	23392	2988231	13520	1134190	496	42867
anı	42568	5815627	21706	1541121	1118	162696
ara	35131	9244353	13214	3291617	266	33189
arı	98892	11307560	29465	1875883	802	119924
ele	42537	4155343	19542	1033009	434	21330
eli	24772	4032934	12331	1917122	178	37524
ene	16231	2044229	10019	689165	373	34100
eni	28999	4191062	14672	877253	703	74777
ere	18954	4045182	7094	1885523	94	101082
eri	96985	12515197	26247	2023420	1238	284561
ile	67846	7538577	35121	2117976	558	75792
ili	43766	4575921	25149	1620273	7056	888062
ine	24779	3566802	9490	416314	97	6788
ini	74613	6873028	28748	1801727	1620	140198
ire	15231	1105808	9310	522386	420	37895
iri	26747	2567948	14192	975278	2924	189194
ıla	39981	3331901	20760	2017784	640	70458
ılı	17182	2456780	5554	508421	726	125024
ına	21810	2952582	9850	802347	90	6776
ını	66870	6787071	27992	2371621	728	44673
ıra	14196	1189403	9842	550674	470	22528
ırı	22985	1409920	13007	568071	1770	106256
ola	8299	5513476	3150	3815714	95	160159
olu	5341	2010995	2908	1223980	142	277830
ona	3969	505621	2152	96502	92	2890
onu	10175	3466126	2981	674818	65	12385
ora	2728	703818	485	23291	26	6323
oru	13321	2768599	9319	1595752	71	9900
öle	930	68022	393	43209	18	1184
ölü	2016	571984	842	52089	35	6434
öne	5883	2518331	3683	281221	99	17353

önü	2402	730451	1513	169685	45	7475
öre	2713	1043721	1181	152332	29	8603
örü	5881	1179019	3036	545473	84	43618
ula	19745	2297231	12349	1219006	389	57472
ulu	10318	2598510	4662	1300791	316	73944
una	5676	1256281	2468	463676	80	15926
unu	17020	3468474	7489	1440374	236	47658
ura	8004	923661	5345	313214	188	16199
uru	11918	2629747	7102	720334	620	79680
üle	7911	570151	4180	246975	170	20610
ülü	4448	339727	1761	140006	184	24829
üne	3200	639206	1347	87819	46	5435
ünü	6288	1213573	1785	365210	130	22535
üre	6385	1078938	4173	274850	154	12910
ürü	8807	956358	6786	477630	511	47643

Table C3. Type and Token Frequencies of the "Vowel-Sonorant#Harmonic Vowel" Patterns (where Sonorant Stands for /l/, /n/ or /r/) at a Word Boundary in the Entire Language, in the Verb Forms and in the Aorist Forms

Pattern VSV	In the whole language		In the verb forms		In the aorist forms	
	Type	Token	Type	Token	Type	Token
ala	0	0	0	0	0	0
alı	0	0	0	0	0	0
ana	0	0	0	0	0	0
anı	0	0	0	0	0	0
ara	0	0	0	0	0	0
arı	0	0	0	0	0	0
ele	1575	450177	1225	368518	40	11626
eli	3121	636537	1000	275569	62	9979
ene	3644	156985	3096	123181	125	6633
eni	5770	1050509	2418	186357	243	19193
ere	3511	698581	2767	547975	76	22430
eri	8038	2839053	2957	1031062	136	120491
ile	1262	219711	1027	98356	38	3177
ili	2848	573530	1028	291966	57	41701
ine	1949	117019	1484	66976	71	6046
ini	3877	310352	1203	51461	103	10481
ire	3405	364523	3093	325722	99	28825
iri	5183	571234	3526	427918	184	31273
ıla	950	187999	758	29685	54	3553
ılı	1110	674451	750	38389	77	5541
ına	1386	302060	1178	244111	43	5742
ını	1934	250417	824	65977	95	16958
ıra	3367	285758	3227	275112	118	11292
ırı	4098	346155	3267	293310	202	21612
ola	1540	3814551	1467	3749098	35	157013
olu	2211	1243795	1057	754325	71	250637
ona	743	113511	244	13454	23	1379
onu	4654	704900	516	68986	13	2467
ora	490	49296	364	22539	21	6302
oru	1429	330738	564	40550	14	1445
öle	407	45333	388	43199	18	1184
ölü	634	63712	519	50113	21	6302
öne	486	170079	476	67610	14	3866

önü	544	514011	332	82257	8	437
öre	1063	165520	937	150856	20	8479
örü	2267	548053	922	341647	24	37583
ula	1341	209924	1260	145691	20	11580
ulu	1542	545518	525	98894	48	18755
una	1541	384599	1368	343511	49	13399
unu	2394	560951	1253	240143	82	23599
ura	2973	231447	2789	209066	79	12466
uru	4515	631883	3121	500832	159	32868
üle	379	23635	281	10369	13	829
ülü	897	102275	211	14403	11	488
üne	958	178472	873	72327	30	5146
ünü	1464	591852	752	251209	57	17478
üre	1076	97091	1002	89421	39	6622
ürü	2046	213754	1354	107603	81	8778

Table C4. Type and Token Frequencies of the "Vowel-Sonorant-Harmonic Vowel-r" Patterns (where Sonorant Stands for /l/, /n/ or /r/) at Any Place in a Word in the Entire Language, in the Verb Forms and in the Aorist Forms

Pattern VSVr	In the whole language		In the verb forms		In the aorist forms	
	Type	Token	Type	Token	Type	Token
alar	26496	2495636	11981	788005	156	5425
alır	196	127404	195	127401	81	75499
anar	739	114871	378	97528	20	4041
anır	2359	193290	1942	185334	821	128769
arar	3101	1239280	565	32436	44	9804
arır	281	21944	268	21872	97	14288
eler	24841	2420011	9359	391118	111	3660
elir	2862	1378678	2194	1075563	119	31127
ener	1396	373280	272	45560	11	1655
enir	1624	95806	1228	91499	555	63940
erer	87	42638	85	42632	21	1989
erir	1016	123140	975	123060	675	84740
iler	26390	2619574	4426	356509	69	23041
ilir	15933	1051883	13963	998813	6947	882618
iner	489	56923	82	18635	16	4292
inir	675	48255	339	19352	102	14492
irer	460	122006	199	71498	7	16840
irir	1853	84803	1214	69180	512	48213
ılar	18042	1271196	4011	279258	41	6401
ılır	1851	189123	1575	180030	670	124412
ınar	530	91445	55	33411	11	354
ınır	1045	128590	589	41639	121	19337
ırar	417	33812	207	31296	7	1241
ırır	1966	52600	1406	45178	619	31048
olar	2371	2114722	60	1522077	14	615
olur	177	388649	109	386197	42	242589
onar	576	29130	392	10115	25	1081
onur	418	55680	175	2333	10	49
orar	104	9286	101	9276	14	5112
orur	32	6088	30	6081	9	3972
öler	56	2341	26	1487	6	489
ölür	24	9069	24	9069	6	5367
öner	1105	235465	669	82389	24	9883

önür	5	197	0	0	0	0
örer	17	4437	17	4437	6	234
örür	42	28444	36	28424	10	20653
ular	5789	726825	691	79828	40	3606
ulur	551	63547	547	63543	227	46513
unar	63	20185	47	20133	7	5069
unur	194	43006	193	43005	86	30444
urar	94	34863	94	34863	8	4049
urur	896	78315	765	56695	249	34542
üler	3229	228714	411	33441	18	717
ülür	359	37427	359	37427	156	24528
üner	163	19937	19	11401	1	13
ünür	308	34877	161	26388	59	14303
ürer	101	17840	92	17139	16	3814
ürür	431	82528	402	27066	174	16486

Table C5. Type and Token Frequencies of the "Vowel-Sonorant-Harmonic Vowel-r" Patterns (where Sonorant Stands for /l/, /n/ or /r/) in Final Position in a Word in the Entire Language, in the Verb Forms and in the Aorist Forms

Pattern VSVr	In the whole language		In the verb forms		In the aorist forms	
	Type	Token	Type	Token	Type	Token
alar	3461	522955	2018	151390	16	3005
alır	18	54803	17	54800	17	54800
anar	7	3386	3	2820	3	2820
anır	472	89186	349	85066	347	85061
arar	9	311802	3	6877	3	6877
arır	26	10160	25	10155	25	10155
eler	3015	528038	1598	100329	17	2892
elir	30	118011	29	5393	27	5259
ener	35	40435	1	519	1	519
enir	400	59831	268	57313	268	57313
erer	5	1800	5	1800	5	1800
erir	351	62451	327	62407	323	62402
iler	7034	719699	2301	158412	6	1203
ilir	4165	682564	3381	653018	2700	638917
iner	29	19947	3	3398	3	3398
inir	31	19363	29	6173	29	6173
irer	5	56386	2	13626	2	13626
irir	401	43516	182	38147	182	38147
ılar	5363	470416	2437	166645	4	4915
ılır	524	121447	400	115816	400	115816
ınar	29	40839	3	239	3	239
ınır	38	30283	37	17404	37	17404
ırar	3	723	2	716	2	716
ırır	396	29968	198	25010	198	25010
olar	189	265815	3	474	2	472
olur	12	209852	12	209852	12	209852
onar	7	2017	3	701	3	701
onur	4	30295	0	0	0	0
orar	2	3497	2	3497	2	3497
orur	3	3183	2	3177	2	3177
öler	6	486	2	407	1	405
ölür	1	3381	1	3381	1	3381
öner	7	20750	1	358	1	358

önür	1	190	0	0	0	0
örer	1	183	1	183	1	183
örür	3	10700	2	10691	2	10691
ular	1078	203152	251	44674	4	2576
ulur	112	38296	112	38296	112	38296
unar	7	3106	2	3091	2	3091
unur	33	27677	32	27676	32	27676
urar	2	2495	2	2495	2	2495
urur	66	38862	65	26487	65	26487
üler	559	100686	165	15364	1	22
ülür	81	22794	81	22794	81	22794
üner	13	5176	1	13	1	13
ünür	18	13116	16	8555	16	8555
ürer	5	4097	4	3442	4	3442
ürür	51	13006	51	13006	51	13006

APPENDIX D

NONCE VERBS WITH CONTEXT AND CARRIER SENTENCES

Table D1. Nonce Monosyllabics Used as Stimuli in Condition A

Tested nonce	**Context sentence (heard)**	**Carrier sentence (seen)**
bıl	Ali bu akşam bılacak.	Ali yazları bazen bıl_____.
bır	Emre bırmayı sever.	Emre kışın akşamları bır_____.
bör	Cem bir sonraki gün börecek.	Cem gece gündüz bör_____.
çer	Ali çermeye gider.	Ali yazları bazen çer_____.
cır	Cem Sinan'a cır dedi.	Sinan iki günde bir cır_____.
çır	Elif çırmayı sever.	Elif sabahları çır_____.
col	Özge colmak ister.	Özge iki günde bir col_____.
cöl	Cem az önce cöldü.	Cem kışın akşamları cöl_____.
çol	Aslı haftaya çolacak.	Aslı geceleri genelde çol_____.
cör	Ali cörmeye gider.	Ali her hafta cör_____.
çör	Emre Sinan'a geçen sene çör dedi.	Ondan beridir Sinan her akşam çör_____.
cul	Mert Merve'ye geçen kış cul dedi.	Ondan beridir Merve gece gündüz cul_____.
cül	Ozan önceki gün cüldü.	Ozan gece gündüz cül_____.
cur	Özge önceki gün curdu.	Özge hep cur_____.
dor	Cem dormaya gider.	Cem her akşam dor_____.
dör	Ali önümüzdeki hafta dörecek.	Ali yazları bazen dör_____.
dül	Gül dülmeyi sever.	Gül kışın her gece dül_____.
fir	Ece yarın firecek.	Ece her akşam fir_____.
fül	Can Merve'ye geçen yaz fül dedi.	Ondan beridir Merve her hafta fül_____.
für	Gül az önce fürmüş.	Gül her gün für_____.
gıl	Aslı gılmayı sever.	Aslı yazın sabahları gıl_____.
gor	İdil önümüzdeki hafta goracak.	İdil hep gor_____.
hel	Ozan önceki gece helmiş.	Ozan sabahları hel_____.
hir	Elif dün hirdi.	Elif yazın her gün hir_____.
hıl	Can Yeşim'e geçen yaz hıl dedi.	Ondan beridir Yeşim geceleri genelde hıl_____.
höl	Elif önceki gece hölmüş.	Elif geceleri höl_____.
hör	Aslı hörmeyi sever.	Aslı kışın akşamları hör_____.

hul	Özge yarın hulacak.	Özge geceleri hul_____.
hül	Elif hülmek ister.	Elif sabahları hül_____.
hur	Mert dün hurmuş.	Mert her sabah hur_____.
ir	Emre irmeyi sever.	Emre her gün ir_____.
ıl	Can ılmak ister.	Can yazın sabahları ıl_____.
köl	Elif yarın kölecek.	Elif her sabah köl_____.
lar	Mert larmak ister.	Mert hep lar_____.
ler	Ozan Yeşim'e ler dedi.	Yeşim iki günde bir ler_____.
lil	Gül lilmeye gider.	Gül yazları bazen lil_____.
lol	Emre önceki gece lolmuş.	Emre kışın her gece lol_____.
löl	Cem önceki gece lölmüş.	Cem sabahları löl_____.
lör	İdil önceki gece lördü.	İdil hep lör_____.
lul	Cem demin luldu.	Cem kışın akşamları lul_____.
lur	Ece lurmaya gider.	Ece kışın her gece lur_____.
lür	Özge lürmeye gider.	Özge geceleri lür_____.
mar	Gül bu akşam maracak.	Gül kışları sık sık mar_____.
mer	Ozan Esra'ya geçen yaz mer dedi.	Ondan beridir Esra iki günde bir mer_____.
mıl	Cem mılmaya gider.	Cem iki günde bir mıl_____.
mır	Mert Burak'a mır dedi.	Burak yazları bazen mır_____.
mör	Ozan Burak'a mör dedi.	Burak akşamları mör_____.
mul	Ece önceki gün muldu.	Ece her zaman mul_____.
nir	Emre Yeşim'e nir dedi.	Yeşim sabahları nir_____.
nıl	Emre nılmak ister.	Emre sabahları nıl_____.
nol	Elif dün noldu.	Elif hep nol_____.
nor	Can normak ister.	Can akşamları nor_____.
nul	Cem önceki gece nulmuş.	Cem iki günde bir nul_____.
or	İdil ormayı sever.	İdil akşamları or_____.
pel	Can pelmek ister.	Can sabahları pel_____.
pıl	Ozan önceki gece pılmış.	Ozan kışın akşamları pıl_____.
rer	İdil rermek ister.	İdil her sabah rer_____.
ril	Cem haftaya rilecek.	Cem her gün ril_____.
rıl	Aslı rılmaya gider.	Aslı her sabah rıl_____.
rör	Ali bir sonraki gün rörecek.	Ali kışın her gece rör_____.
rür	Aslı Sinan'a geçen sene rür dedi.	Ondan beridir Sinan her zaman rür_____.
şel	Özge Sinan'a geçen yaz şel dedi.	Ondan beridir Sinan akşamları şel_____.
sir	Cem önceki gün sirdi.	Cem sabahları sir_____.

şıl	Ece Sinan'a şıl dedi.	Sinan her akşam şıl_____.
şır	Mert önceki gün şırdı.	Mert her akşam şır_____.
şol	Ozan yarın şolacak.	Ozan gece gündüz şol_____.
şöl	Elif şölmeye gider.	Elif kışın her gece şöl_____.
şör	Özge az önce şördü.	Özge her sabah şör_____.
şur	İdil haftaya şuracak.	İdil iki günde bir şur_____.
tıl	Can tılmayı sever.	Can yazın her gün tıl_____.
töl	İdil tölmeyi sever.	İdil kışın her gece töl_____.
ül	Ece Burak'a geçen sene ül dedi.	Ondan beridir Burak her zaman ül_____.
ur	İdil urmayı sever.	İdil her sabah ur_____.
val	Gül Yeşim'e geçen kış val dedi.	Ondan beridir Yeşim kışları sık sık val_____.
vil	Ozan vilmeye gider.	Ozan yazın her gün vil_____.
vir	Ozan önceki gece virmiş.	Ozan iki günde bir vir_____.
vır	Cem az önce vırmış.	Cem gece gündüz vır_____.
vor	Özge az önce vormuş.	Özge her hafta vor_____.
vür	Ece vürmeyi sever.	Ece yazın sabahları vür_____.
yöl	Emre Yeşim'e geçen yaz yöl dedi.	Ondan beridir Yeşim kışın her gece yöl_____.
yör	Özge demin yördü.	Özge kışları sık sık yör_____.
zal	Ali bir sonraki gün zalacak.	Ali gece gündüz zal_____.
zöl	Cem demin zölmüş.	Cem geceleri genelde zöl_____.
zür	Ece önceki gün zürdü.	Ece geceleri genelde zür_____.

Table D2. Nonce Monosyllabics Used as Stimuli in Condition B

Tested nonce	Context sentence (heard)	Carrier sentence (seen)
ber	Özge Sinan'a ber dedi.	Sinan yazın her gün ber_____.
bül	Gül Merve'ye geçen yaz bül dedi.	Ondan beridir Merve hep bül_____.
cal	Elif önümüzdeki hafta calacak.	Elif geceleri cal_____.
cel	Emre yarın celecek.	Emre kışın akşamları cel_____.
cil	Özge Merve'ye geçen sene cil dedi.	Ondan beridir Merve yazın sabahları cil_____.
cir	Elif yarın cirecek.	Elif akşamları cir_____.
cıl	Ali Sinan'a geçen yaz cıl dedi.	Ondan beridir Sinan geceleri cıl_____.
çıl	Mert az önce çılmış.	Mert her zaman çıl_____.
cor	Özge dün cordu.	Özge her akşam cor_____.
çül	Aslı dün çülmüş.	Aslı kışın akşamları çül_____.
cür	Elif cürmeyi sever.	Elif hep cür_____.
çur	Emre Sinan'a çur dedi.	Sinan hep çur_____.
dıl	Özge dılmaya gider.	Özge geceleri dıl_____.
dır	Can önceki gece dırmış.	Can her akşam dır_____.
fıl	Cem Merve'ye fıl dedi.	Merve yazın her gün fıl_____.
föl	İdil bu akşam fölecek.	İdil kışın akşamları föl_____.
for	Mert Yeşim'e for dedi.	Yeşim kışları sık sık for_____.
för	Ozan önümüzdeki hafta förecek.	Ozan kışın akşamları för_____.
gal	Can galmayı sever.	Can geceleri genelde gal_____.
gil	Ali gilmeye gider.	Ali her sabah gil_____.
gul	Aslı haftaya gulacak.	Aslı her akşam gul_____.
hil	Gül hilmeye gider.	Gül her akşam hil_____.
il	Mert Merve'ye geçen kış il dedi.	Ondan beridir Merve gece gündüz il_____.
ır	Ece önceki gün ırdı.	Ece geceleri genelde ır_____.
lel	Elif haftaya lelecek.	Elif kışın her gece lel_____.
lıl	İdil lılmak ister.	İdil kışları sık sık lıl_____.
lır	Gül lırmayı sever.	Gül akşamları lır_____.
lül	Ozan önümüzdeki hafta lülecek.	Ozan her sabah lül_____.
mol	Mert molmaya gider.	Mert yazın sabahları mol_____.
möl	Emre mölmeyi sever.	Emre iki günde bir möl_____.
mur	Gül dün murmuş.	Gül her akşam mur_____.
mür	Mert mürmeyi sever.	Mert gece gündüz mür_____.

ner	İdil nermeyi sever.	İdil her sabah ner_____.
nil	Özge nilmek ister.	Özge her gün nil_____.
nır	Ali nırmak ister.	Ali yazın her gün nır_____.
nöl	Elif önceki gün nöldü.	Elif yazın sabahları nöl_____.
nör	Ece dün nörmüş.	Ece her gün nör_____.
nül	Mert nülmeye gider.	Mert kışın her gece nül_____.
nür	Gül Sinan'a geçen kış nür dedi.	Ondan beridir Sinan hep nür_____.
pöl	Mert pölmeye gider.	Mert sabahları pöl_____.
por	Ali pormak ister.	Ali her sabah por_____.
pör	Özge Merve'ye geçen sene pör dedi.	Ondan beridir Merve kışın akşamları pör_____.
pül	Ece bu akşam pülecek.	Ece akşamları pül_____.
pur	Özge demin purmuş.	Özge iki günde bir pur_____.
ral	Emre ralmayı sever.	Emre yazın sabahları ral_____.
rar	Gül Merve'ye rar dedi.	Merve akşamları rar_____.
rel	Elif relmeye gider.	Elif kışları sık sık rel_____.
rir	Emre rirmek ister.	Emre geceleri genelde rir_____.
rır	Ozan Yeşim'e rır dedi.	Yeşim yazın her gün rır_____.
röl	Gül yarın rölecek.	Gül her akşam röl_____.
ror	Ozan rormak ister.	Ozan iki günde bir ror_____.
rul	Ece önceki gece ruldu.	Ece geceleri rul_____.
rül	Ozan bu akşam rülecek.	Ozan kışları sık sık rül_____.
rur	Mert dün rurdu.	Mert kışın akşamları rur_____.
şar	Ali yarın şaracak.	Ali gece gündüz şar_____.
şil	Can şilmeye gider.	Can geceleri şil_____.
şir	Cem şirmek ister.	Cem geceleri şir_____.
sıl	Emre az önce sıldı.	Emre kışları sık sık sıl_____.
söl	Emre bir sonraki gün sölecek.	Emre hep söl_____.
sül	Ozan bir sonraki gün sülecek.	Ozan gece gündüz sül_____.
şul	Cem bir sonraki gün şulacak.	Cem sabahları şul_____.
şül	Can önceki gece şüldü.	Can yazın sabahları şül_____.
şür	Emre Sinan'a şür dedi.	Sinan geceleri şür_____.
til	Elif tilmek ister.	Elif her zaman til_____.
tir	Mert Merve'ye geçen kış tir dedi.	Ondan beridir Merve kışları sık sık tir_____.
ul	İdil haftaya ulacak.	İdil iki günde bir ul_____.
ür	Aslı az önce ürmüş.	Aslı sabahları ür_____.
vel	Ali demin veldi.	Ali kışın her gece vel_____.

vıl	Ece Yeşim'e vıl dedi.	Yeşim sabahları vıl_____.
vol	Can Merve'ye vol dedi.	Merve iki günde bir vol_____.
völ	Can völmek ister.	Can geceleri genelde völ_____.
vör	Mert dün vördü.	Mert geceleri vör_____.
vul	Cem vulmak ister.	Cem her akşam vul_____.
vül	Aslı önceki gün vüldü.	Aslı kışları sık sık vül_____.
yil	Ece yilmeyi sever.	Ece yazları bazen yil_____.
yir	Ali dün yirmış.	Ali geceleri yir_____.
yur	Emre yarın yuracak.	Emre gece gündüz yur_____.
zel	Elif bu akşam zelecek.	Elif kışları sık sık zel_____.
zıl	Elif Sinan'a geçen sene zıl dedi.	Ondan beridir Sinan kışın her gece zıl_____.
zır	Elif zırmak ister.	Elif iki günde bir zır_____.
zol	Can önceki gece zoldu.	Can yazın sabahları zol_____.
zör	Elif az önce zörmüş.	Elif yazları bazen zör_____.
zul	Aslı demin zulmuş.	Aslı geceleri genelde zul_____.
zur	Ali Yeşim'e geçen kış zur dedi.	Ondan beridir Yeşim hep zur_____.

Table D3. Nonce Multisyllabics Used as Fillers in Conditions A and B

Tested nonce	Context sentence (heard)	Carrier sentence (seen)
deliş	İdil Burak'a geçen kış deliş dedi.	Ondan beridir Burak geceleri deliş_____.
derek	Cem derekmeyi sever.	Cem iki günde bir derek_____.
dözle	Ali önceki gün dözledi.	Ali akşamları dözle_____.
lelgele	Aslı bu akşam lelgeleyecek.	Aslı sabahları lelgele_____.
lestele	İdil yarın lesteleyecek.	İdil yazın her gün lestele_____.
lombala	Cem az önce lombaladı.	Cem her akşam lombala_____.
loşalt	Ali Yeşim'e loşalt dedi.	Yeşim yazın her gün loşalt_____.
mekerle	Emre önceki gece mekerlemiş.	Emre yazın sabahları mekerle_____.
mitreş	Ozan önceki gece mitreşmiş.	Ozan her zaman mitreş_____.
mıraşla	Mert Yeşim'e mıraşla dedi.	Yeşim her hafta mıraşla_____.
moparla	Ozan önceki gece moparladı.	Ozan kışları sık sık moparla_____.
nayış	Cem nayışmayı sever.	Cem her zaman nayış_____.
nırılda	Emre bir sonraki gün nırıldayacak.	Emre hep nırılda_____.
raçmala	Aslı Merve'ye raçmala dedi.	Merve her hafta raçmala_____.
rerpele	Elif bir sonraki gün rerpeleyecek.	Elif iki günde bir rerpele_____.
seğerle	Ece demin seğerlemiş.	Ece her gün seğerle_____.
şekele	Aslı bu akşam şekeleyecek.	Aslı gece gündüz şekele_____.
sepola	Elif sepolamak ister.	Elif akşamları sepola_____.
süzenle	Can Yeşim'e süzenle dedi.	Yeşim akşamları süzenle_____.
tapsa	Gül Sinan'a tapsa dedi.	Sinan her gün tapsa_____.
tıpırda	Ali tıpırdamayı sever.	Ali sabahları tıpırda_____.
zankıla	Mert zankılamayı sever.	Mert her akşam zankıla_____.
zasla	Elif Sinan'a geçen kış zasla dedi.	Ondan beridir Sinan yazın her gün zasla_____.
zetiş	Elif bu akşam zetişecek.	Elif geceleri genelde zetiş_____.
zıprat	Özge bu akşam zıpratacak.	Özge yazın sabahları zıprat_____.
zönet	Emre yarın zönetecek.	Emre hep zönet_____.
züzle	Ozan bir sonraki gün züzleyecek.	Ozan her hafta züzle_____.

APPENDIX E

INSTRUCTIONS DISPLAYED ON THE SCREEN DURING THE EXPERIMENT

(At the beginning)

Deneyimize hoş geldiniz!

Devam etmek için "Space" tuşuna basın.

Bu deneyde, size kulaklıkla cümleler okunacak.
Bu cümlelerde uydurulmuş fiiller var.
Cümleleri dikkatlice dinleyin.

Devam etmek için "Space" tuşuna basın.

Daha sonra ekranda beliren cümleyi, istenen fiille tamamlayın.
Tamamladıktan sonra "Space" tuşuna basın.

Cümleleri tamamlarken –iyor eki kullanmak yasak!

Bir deneyelim mi?

Devam etmek için "Space" tuşuna basın.

(After the training)

Bravo, artık hazırsınız! Deneye başlayabiliriz.

Deney sırasında küçük molalar verilecek.
Moladan sonra hazır olduğunuz zaman devam edebilirsiniz.

Devam etmek için "Space" tuşuna basın.

Cümleleri tamamladığınızdan emin olduktan sonra
"Space" tuşuna basmayı unutmayın.

Hatırlayın: -iyor kullanmak yasak!

Haydi başlayalım!

Devam etmek için "Space" tuşuna basın!

(During the test at the break)

Küçük bir mola verelim...

Devam etmek için "Space" tuşuna basın!

(After the test, for the questionnaire)

Bitmek üzere! Önce size birkaç soru sormak istiyoruz.

Kimliğinizi ortaya çıkaracak hiçbir kişisel bilgi alınmayacak.
İsminiz, ses kayıtlarınızla ya da verilerinizle eşleştirilmeyecektir.

Devam etmek için "Space" tuşuna basın.

Yaşınız:

Cinsiyetiniz:

Esas olarak hangi elinizi kullanırsınız? Sol/Sağ

Bölümünüz:

Anadiliniz:

Tek anadilim Türkçe
Anadilim Türkçe ama en az bir başka anadilim daha var
Anadilim Türkçe dışında bir dil

Öğrendiğiniz yabancı diller:

Sizce bu deney ne hakkındaydı?

(Last screen, after the questionnaire)

Bitti!
Desteğiniz için çok teşekkür ederiz.

Bu deneyde çekimlediğiniz fiiller uydurulmuş sözcükler
olduğu için doğru ya da yanlış cevap yok.
Tüm cevaplar bizim için çok faydalı.

Deneyin sonuçlarını öğrenmek için
Dilbilimi Bölümü'nün internet sayfasında
yayınlanan sunumları takip edebilirsiniz.
Devam etmek için "Space" tuşuna basın.

APPENDIX F

ISLANDS OF RELIABILITY

In Albright and Hayes (2003), the Minimal Generalization Learner model iteratively discover phonological rules that apply to a certain cluster called “Island Of Reliability” and attribute it a confidence level by simply dividing the number of hits over the scope of the rule (the number of verbs this rule effectively applies to / the number of verbs this rule could apply to). This raw confidence score is then slightly adjusted to penalize rules based on a small number of forms and non-uniform distributions. Then the output is assigned the score of the best rule that derives it. For example, we get:

Past tenses for *gleed* derived by the rule-based model

Output	Rule	Hits/Scope	Raw confidence	Adjusted confidence	Hits/Failures
gleeded	$\varnothing \rightarrow$ əd / [X {d, t} ___]$_{[+\text{past}]}$	1146/1234	0.929	0.872	*want*, *need*, *start*, *wait*, *decide*, *etc.* / **get*, **find*, **put*, **set*, **stand*, etc.
gled	i $\rightarrow$ ɛ / [X {l, r} ___ d]$_{[+\text{past}]}$	6/7	0.857	0.793	*read*, *lead*, *bleed*, *breed*, *mislead*, *misread* / **plead*
glode	i $\rightarrow$ o / [X C ___ [+cons]]$_{[+\text{past}]}$	6/184	0.033	0.033	*speak*, *freeze*, *weave*, *interweave*, *bespeak* / **leak*, **teach*, **leave*, etc.
gleed	No change / [X {d, t} ___]$_{[+\text{past}]}$	29/1234	0.024	0.014	*shed*, *spread*, *put*, *let*, *set*, *cut*, *hit*, *beat*, *shut*, *hurt*, *cost*, *cast*, *burst*, *split*, etc. / **get*, **want*, **need*, etc.

In our study we do not have a model iteratively discovering rules and we do not make use of features. However we can evaluate the raw confidence of the potential rules using the first consonant, vowel, last consonant, rhyme and C_C pattern of the verb root, and look at if they tell us something. Albright and Hayes (2003) use the type frequency counts, but we show results for both type and token frequencies as

both were available. For instance let us look at two very similar items that obtained very different -Ir use rates: *rur*- (58%) and *rör*- (7%):

- The rules concerning the first consonant and the pattern C_C could not have appeared in the model since there is no corresponding monosyllabic verb root in Turkish, so they are left blank.
- The rule concerning the last consonant is the same for both items, predicting -Ar use if type frequency is used, -Ir use if token frequency is used.
- As for the rules concerning the vowel and the rhyme, they seem to give the inverse result: *rör*- has more rules predicting irregularization and with a higher confidence score but it has a -Ir use of only 7%, than *rur*- that has a -Ir use of 58%.

rur	**Rules**	**Type**			**Token**		
(58%)		**Hits/ scope**	**Raw conf**	**Prediction**	**Hits/scope**	**Raw conf**	**Prediction**
First C	r__ > r__Ir						
Vowel	_u_ > _u_ur	3/17	0,18	rurar	26660/80793	0,33	rurar
Last C	__r > __rIr	5/21	0,24	rurar	86324/131077	0,66	rurur
Rhyme	_ur > _urur	2/4	0,5	rurur / rurar	11398/15447	0,74	rurur
C_C	r_r > r_rIr						

rör	**Rules**	**Type**			**Token**		
(7%)		**Hits/ scope**	**Raw conf**	**Prediction**	**Hits/scope**	**Raw conf**	**Prediction**
First C	r__ > r__Ir						
Vowel	_ö_ > _ö_ür	2/17	0,11	rörer	29521/12542	0,7	rörür
Last C	__r > __rIr	5/21	0,24	rörer	86324/131077	0,66	rörür
Rhyme	_ör > _örür	1/2	0,5	rörer / rörür	24154/24388	0,99	rörür
C_C	r_r > r_rIr						

Successful examples exist though, such as *nul-* (31%) and *nül-* (0%):

- Again, the rules concerning the first consonant and the pattern C_C could not have appeared in the model since there is no corresponding monosyllabic verb root in Turkish, so they are left blank.
- The rule concerning the last consonant is also the same for both items, predicting -Ar use if type frequency is used, -Ir use if token frequency.
- However here, the rule concerning the vowel predicts the use of -Ar for both items, whether type or token frequency is used.
- The rule concerning the rhyme supports a high -Ir use for *nul-* (31%) and a low -Ir use for *nül-* (0%), even if type frequency is used as in the paper.

nul (31%)	**Rules**	**Type**			**Token**		
		Hits/ scope	**Raw conf**	**Prediction**	**Hits/scope**	**Raw. conf**	**Prediction**
First C	n__ > n__Ir						
Vowel	_u_ > _u_ur	3/17	0,18	nular	26660/80793	0,33	nular
Last C	__l > __lIr	7/22	0,32	nular	370767/379294	0,98	nulur
Rhyme	_ul > _ulur	1/1	1	nulur	15262/15262	1	nulur
C_C	n_l > n_lIr						

nül (0%)	**Rules**	**Type**			**Token**		
		Hits/ scope	**Raw conf**	**Prediction**	**Hits/scope**	**Raw conf**	**Prediction**
First C	n__ > n__Ir						
Vowel	_ü_ > _ü_ür	0/20	0	nüler	0/22396	0	nüler
Last C	__l > __lIr	7/22	0,32	nüler	370767/379294	0,98	nülür
Rhyme	_ül > __ülür	0/1	0	nüler	0/735	0	nüler
C_C	n_l > n_lIr						

From the *nul- / nül-* example we understand this raw defining and ordering of rules could capture some of the results, but the *rur- / rör-* example teaches us that these rules are not precise enough to account for experimental data.

REFERENCES

Aksu-Koç, A., & Slobin, D. I. (1985). The acquisition of Turkish. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, Volume 1: The data* (pp. 839-878). Hillsdale, NJ: Lawrence Erlbaum Associates.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, *90*(2), 119-161. doi:10.1016/s0010-0277(03)00146-x

Baayen, R. H. (2001). *Word frequency distributions*. Dordrecht, Netherlands: Springer Science & Business Media.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*(2-3), 150-177. doi:10.1080/00437956.1958.11659661

Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language*, 265-289. doi:10.2307/414099

Bybee, J. L., & Moder, C. L. (1983). Morphological classes as natural categories. *Language*, 251-270. doi:10.2307/413574

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425-455. doi:10.1080/01690969508407111

Bybee, J. L., & Hopper, P. J. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam, Netherlands: John Benjamins Publishing.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335-344. doi:10.1016/j.tics.2006.05.006

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition*, *65*(2), 197-230. doi:10.1016/S0010-0277(97)00044-9

Hahn, U., & Bailey, T. M. (2005). What makes words sound similar? *Cognition*, *97*(3), 227-267. doi:10.1016/j.cognition.2004.09.006

Hare, M., Elman, J. L., & Daugherty, K. G. (1995). Default generalisation in connectionist networks. *Language and Cognitive Processes*, *10*(6), 601-630. doi:10.1080/01690969508407115

Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say ‘holded’ more than boys. *Developmental Science*, *9*(1), 21-32. doi:10.1111/j.1467-7687.2005.00459.x

Kanık, M. (2015). The Turkish aorist and progressive: Present tense, future tense, or what? *Journal of Language and Linguistic Studies*, *11*(1), 103-115. Retrieved from http://www.jlls.org

Marcus, G. F., Pinker, S., Ullman, M. T., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57*(4) 1-178. doi:10.1111/j.1540-5834.1992.tb00313.x

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology, 37*(3), 243-282. doi:10.1006/cogp.1998.0694

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences, 6*(11), 465-472. doi:10.1016/S1364-6613(02)01993-9

Nakipoğlu, M., & Ketrez, N. (2006). 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* (pp. 399- 410). Somerville, MA: Cascadilla Press.

Nakipoğlu, M., & Üntak, A. (2008). A complete verb lexicon of Turkish based on morphemic analysis. *Turkic Languages, 12*, 221-280. Retrieved from http://www.harrassowitz-verlag.de

Nakisa, R. C., Plunkett, K., & Hahn, U. (2000). Single and dual-route models of inflectional morphology. In P. Broeder & J. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches* (pp. 201-222). Oxford, England: Oxford University Press.

Nemeth, D., Janacsek, K., Turi, Z., Lukacs, A., Peckham, D., Szanka, S., ...Ullman, M. T. (2015). The production of nominal and verbal inflection in an agglutinative language: Evidence from Hungarian. *PloS One, 10*(3). doi:10.1371/journal.pone.0119003

Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, 34*(4), 393-418. doi:10.1016/0022-2496(90)90020-a

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*(1-2), 73-193. doi:10.1016/0010-0277(88)90032-7

Pinker, S. (2001). Four decades of rules and associations, or whatever happened to the past tense debate. In E. Dupoux, *Language, the brain, and cognitive development: Papers in honor of Jacques Mehler* (pp. 157-179). Cambridge, MA: MIT Press.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*(11), 456-463. doi:10.1016/S1364-6613(02)01990-3

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*(1), 21-69. doi:10.1016/0010-0277(93)90057-3

Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*(1), 1-56. doi:10.1080/01690969308406948

Racz, P., Beckner, C., Hay, J. B., & Pierrehumbert J. B. (2014). *Rules, Analogy and Social Factors codetermine past-tense formation patterns in English*. Paper presented at the Joint 1-Day Workshop between SIGMORPHON and SIGFSM, Association for Computational Linguistics, Baltimore MD, June 2014. Retrieved from http://faculty.wcas.northwestern.edu/~jbp/publications/Racz_SIGMORPHON.pdf

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, *45*(1), 45-94. doi:10.1016/S0010-0285(02)00001-4

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

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* (pp. 417-427). Berlin Heidelberg, Germany: Springer.

Say, T., & Clahsen, H. (2002). Words, rules and stems in the Italian mental lexicon. In S. G. Nooteboom, F. Weerman & F. N. K. Wijnen (Eds.), *Storage and computation in the language faculty* (pp. 93-129). Dordrecht, Netherlands: Springer Science & Business Media.

Seidenberg, M. S., & Plaut, D. C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive Science*, *38*(6), 1190-1228. doi:10.1111/cogs.12147

Tekin, T. (1995). Relics of Altaic stem-final vowels in Turkic. In B. Kellner-Heinkele & M. Stachowski (Eds.), *Laut- und Wortgeschichte der Türksprachen* (pp. 173-188). Berlin, Germany: Harrassowitz Verlag.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. S*cience, 331*(6022), 1279-1285. doi:10.1126/science.1192788

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. doi:10.1016/j.jml.2014.06.001