

# Rule-based vs. similarity-based generalization: An experimental study on the Turkish Aorist

---

Elise Michon  
SYSTRAN  
elise.michon@systrangroup.com

Mine Nakipoğlu  
Boğaziçi University  
nakipogl@boun.edu.tr

## 1. Introduction

Acquisition of inflectional morphology is known to be less challenging for children when there is one-to-one mapping between a form and its function. When, however, a form has various functions and allomorphs, as in the case of Turkish aorist –a Tense/Aspect/Modality marker that can mark present tense/habitual aspect/epistemic modality– where the distribution of the allomorphs exhibits irregularity, acquisition gets demanding. Nakipoğlu and Ketrez (2006, and under review), testing 143 children through an elicited production task uncovers the developmental path the Turkish-children follow in learning the aorist. Building on the acquisition results, to dig deeper into the cognition puzzle the aorist presents, the present study looks into adults' production of the aorist when presented with non-existent roots, to unpack whether it suggests a rule-based or an analogy-based generalization.

The most famous case of such an irregular verbal pattern is the English past-tense, where regulars 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 undergo various phonological processes, such as change of the stem vowel (*sing-sang*), change of the rhyme (*bring-brought*), or no change at all (*hit-hit*). The irregularity in English past tense has provided a platform for the formulation of two processing models that differ from each other in respect of how they handle regulars and irregulars. The Single Mechanism Model advocates a single, analogy-based mechanism that is sensitive to phonology and the statistical properties of the stored exemplars in handling both regulars and irregulars. The Dual-Mechanism Model, however, dissociates between regulars and irregulars and argues that regulars should be a product of rule-based generalization and analogy should only play a role in the production of irregulars. The present study investigates how Turkish adults produce the aorist when presented with nonce roots in an experimental setting to see to what extent rules, similarity and type/token frequency are at work in the processing of an irregular verbal pattern in their native language. The research questions the present study is interested in addressing are as follows:

- (i) When conjugating a nonce root in a verbal paradigm that exhibits irregularity, do participants in a verb-elicitation task manipulate an abstraction in the form of a rule; or do they analogize across stored exemplars?
- (ii) If participants rely on similarity with stored exemplars, do they base their answers on rhyme (nucleus-coda) as a measure of similarity, or are there other similarity measures such as sharing the coda consonant, the nucleus, or having overlapping trigrams, quadrigrams that influence the outcome?
- (iii) Do type (number of different verbs) and/or token (surface-frequency) counts of stored exemplars influence participants' answers?

In Section 2, we describe the Turkish aorist, first in structural terms then in quantitative terms, providing the type/token distribution of the aorist forms in Turkish. Section 3 discusses the available processing models that have been offered to account for irregularly behaving inflectional morphemes. Section 4 presents the procedure of the psycholinguistic 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 Section 5. Finally, in Section 6 we will discuss what our findings reveal about the adult representation of the Turkish aorist, and more generally about irregular patterns in the mind.

## 2. Description of the Turkish aorist

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. The 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).<sup>1</sup>

- |     |               |         |                 |               |
|-----|---------------|---------|-----------------|---------------|
| (1) | <i>oku-</i>   | ‘read’  | <i>oku-r</i>    | ‘s/he reads’  |
| (2) | <i>kapat-</i> | ‘close’ | <i>kapat-ır</i> | ‘s/he closes’ |
|     | <i>pişir-</i> | ‘cook’  | <i>pişir-ir</i> | ‘s/he cooks’  |
|     | <i>otur-</i>  | ‘sit’   | <i>otur-ur</i>  | ‘s/he sits’   |
|     | <i>düşün-</i> | ‘think’ | <i>düşün-ür</i> | ‘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), there are 13 verbs that take the affix *-Ir*, as in (4).

- |     |             |          |               |                |
|-----|-------------|----------|---------------|----------------|
| (3) | <i>bak-</i> | ‘look’   | <i>bak-ar</i> | ‘s/he looks’   |
|     | <i>seç-</i> | ‘choose’ | <i>seç-er</i> | ‘s/he chooses’ |
| (4) | <i>al-</i>  | ‘take’   | <i>al-ır</i>  | ‘s/he takes’   |
|     | <i>bil-</i> | ‘know’   | <i>bil-ir</i> | ‘s/he knows’   |
|     | <i>bul-</i> | ‘find’   | <i>bul-ur</i> | ‘s/he finds’   |
|     | <i>dur-</i> | ‘stop’   | <i>dur-ur</i> | ‘s/he stops’   |
|     | <i>gel-</i> | ‘come’   | <i>gel-ir</i> | ‘s/he comes’   |
|     | <i>gör-</i> | ‘see’    | <i>gör-ür</i> | ‘s/he sees’    |
|     | <i>kal-</i> | ‘stay’   | <i>kal-ır</i> | ‘s/he stays’   |
|     | <i>ol-</i>  | ‘be’     | <i>ol-ur</i>  | ‘s/he is’      |
|     | <i>öl-</i>  | ‘die’    | <i>öl-ür</i>  | ‘s/he dies’    |
|     | <i>san-</i> | ‘think’  | <i>san-ır</i> | ‘s/he thinks’  |
|     | <i>var-</i> | ‘arrive’ | <i>var-ır</i> | ‘s/he arrives’ |
|     | <i>ver-</i> | ‘give’   | <i>ver-ir</i> | ‘s/he gives’   |
|     | <i>vur-</i> | ‘hit’    | <i>vur-ur</i> | ‘s/he hits’    |

<sup>1</sup> 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.

These so-called irregularly behaving verbs all end in the sonorants /l/, /r/, or /n/. Characterization of these verbs on the basis of sonority is necessary but not sufficient, since not all sonorant-ending monosyllabic verbs fall under the *-Ir* category. Consider the examples in (5a) and (5b) to see how identical phonological rhymes can be used with *-Ar*, and *-Ir*, yielding irregularity.

- (5) a. *öl-* → *öl-ür* ‘dies’ vs. *böl-* → *böl-er* ‘divides’  
 b. *gel-* → *gel-ir* ‘comes’ vs. *del-* → *del-er* ‘pierces’

In what follows, we present a quantitative picture of the aorist through a search of the BOUN Corpus, a web based corpus that contains 1.3 million word-forms (Sak et al. 2008). To get a solid grasp of the distribution of *-Ar* and *-Ir* types and tokens in the language, in the tables below we examine monosyllabic verbs and multisyllabic verbs in Turkish. Table 1 presents the breakdown of the monosyllabic verbs, and shows that of the 229 monosyllabic roots in Turkish, 58 are sonorant ending. Furthermore of these sonorant-ending roots, 13 are *-Ir* type and 45 are *-Ar* type. Overall, we see that *-Ar* is higher in type and token with monosyllabic roots.

**Table 1:** Monosyllabic roots: Type/Token Frequency of *-Ar* and *-Ir*

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

When we turn to the frequency count of the consonant-ending multisyllabic verbs, we see that *-Ir* is prevalent both in type and token, as illustrated in Table 2 below.

**Table 2:** Multisyllabic roots: Type/Token Frequency of *-Ar* and *-Ir*

Multisyllabic roots	<i>-Ar</i>		<i>-Ir</i>		Total	
	Type	Token	Type	Token	Type	Token
<b>Sonorant</b>	1	4	7976	2121451	7977	2121455
<b>Non-sonorant</b>	88 <sup>2</sup>	42953	1393	427607	1481	470560
<b>Total</b>	89	42957	9369	2549058	9458	2592015

Table 3 below lays out the overall *-Ar* and *-Ir* forms in the language. Collapsing all the counts, we observe that in Turkish, *-Ir* predominates over *-Ar* both in type and in token.

**Table 3:** Total *-Ar* and *-Ir* counts in Turkish

Mono/Multi/Vowel-ending verbs	<i>-Ar</i>		<i>-Ir</i>		Total	
	Type	Token	Type	Token	Type	Token
<b>Total</b>	998	1537983	9433	3399037	10530	4937020

In learning the aorist, what appears to be quite challenging for Turkish children is the behavior of the sonorant-ending monosyllabic verbs, as children produce few errors with multisyllabic and non-sonorant-ending monosyllabic verbs. With sonorant-ending monosyllabics, however, they tend to overregularize the irregularly behaving roots in higher proportions early on and irregularize the regularly behaving roots for extended periods of

<sup>2</sup> By rule, multisyllabic verbs all take *-Ir*, but some compound verbs constructed with the monosyllabic root *et-* ‘make’, for example, take *-Ar* as in *haket-/hakeder* ‘deserves’ or *affet-/affeder* ‘forgives’, etc.

acquisition. Building on these acquisition results, the current study turns to the question of how Turkish adults represent the affix when given sonorant-ending monosyllabic nonce roots that mimic the existing sonorant-ending roots. We hypothesize that if in the production of aorist for sonorant-ending nonce roots, adults opt for *-Ar*, this would suggest that an abstraction is already in place and is acting over the roots; if, however, adults analogize over existing exemplars, *-Ir* may be the choice of the participants due to the high type and token frequency of the form. Thus we hypothesize that both rules and similarity would matter in handling a morphological pattern in one's native language and that frequency, in particular type frequency, may enhance the effect of similarity.

### 3. Theoretical background

The processing of irregularly behaving patterns has been mostly investigated through the example of the English past tense. Under the Single-mechanism model, every verb, whether regular or irregular, is 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 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 this account, 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. Several implementations and refinements of this usage-based model (Plunkett & Marchman 1993; Hare et al. 1995; Nakisa et al. 2001, among others) have been built on the seminal two-layer artificial neural network proposed by Rumelhart and McClelland (1986). Such models, fed with pairs of the type "verb-past tense", adjust their weights to learn this mapping and tend to reflect general learning and processing mechanisms.

On the other hand, the Dual-mechanism model has advanced the idea that two mechanisms are needed to account for regulars and irregulars: a rule-based mechanism for regulars and an analogy-based mechanism for irregulars. 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 retrieved from the lexicon, otherwise it is computed directly by the rule, with no need to store the form. This view has been developed by Pinker and Prince (1988), Marcus et al. (1992), Pinker (2001), among others. The strength of the implementations of the Dual-mechanism model 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 (McClelland & Patterson 2002; Pinker & Ullman 2002). At the heart of the discussion lies the observation that usage-based 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 use of abstraction that is specific to the linguistic domain.

With this background in mind, in what follows we lay out the procedure of the experiment testing adults' production of the aorist with nonce roots.

## 4. Methodology

The experiment was presented in the form of an Inflected-Verb-Elicitation-Task programmed with the software Open Sesame. Seated in front of a computer in the lab, participants were instructed to carefully listen to the sentences they would hear as they contained made-up verbs. As the irregularity lies in the monosyllabic sonorant-ending verbs, in constructing the nonce roots we have restricted ourselves with 168 monosyllabic nonce-roots that end either in /l/ or /r/. The 168 items were comprised of 16 rhyme templates (Consonant-Vowel-Sonorant, CVS) each occurring in various numbers due to the extent of overlap they exhibit with the existing verbs in Turkish. To make the purpose of the experiment opaque to the participants, 27 multisyllabic fillers were added. After 6 training items, test items were displayed in a subject-by-subject randomized order.

A nonce root was introduced in a context sentence that was presented auditorily as in (6a). It was then presented in a carrier sentence in written form (6b) and participants were instructed to complete the sentence orally. For example, the nonce monosyllabic root *pül-* when conjugated in the aorist could either be produced as *pül-er*, with the regular *-Ar*, or as *pül-ür* with the irregular *-Ir*.

- (6) a. *Ece bu akşam pül-ecek.* ‘Ece will pül tonight.’  
 b. *Ece akşamları **pül**\_\_\_\_\_.* ‘Ece every night pül\_\_\_\_\_.’

90 monolingual Turkish-speaking students were tested in two conditions where the stimuli were distributed randomly. Age was concentrated around 20 and gender was roughly balanced across conditions (52 women, 37 men, 1 other answer). All participants had a good command of English, and 71 reported having knowledge of at least one other foreign language.

## 5. Results

This section first lays out the general results and touches upon whether what we have observed is correlated with application of a rule. We then investigate the effect of similarity and frequency in forms that have given rise to *-Ir* use, for different measures of similarity.

The results we have obtained show that overall, *-Ar* has proven to be the dominant form, as the participants have opted for the use of *-Ar* with a rate of 86% on the 6391 elicited aorist-attached nonce-roots. More participants produced the *-Ar* form than the *-Ir* form for every nonce-root tested except for *rur-* where *rur-ur* was produced by 21 participants and *rur-ar* by 15 participants. Furthermore, 6 nonce roots, i.e., *nül-*, *pül-*, *şöl-*, *mol-*, *cil-*, *şil-* are observed to have always been produced with the *-Ar* suffix.

Participants have used *-Ir* in only 14% of their answers. This clearly shows that the role of analogy has been scarce. In what follows, we will look into the characteristics of the stems that have given rise to more *-Ir* use.<sup>3</sup> We first examine to what extent Rhyme, the measure of similarity chosen a priori in English past tense studies, determines the *-Ir* use in Turkish aorist. There are 16 possible Rhyme templates for nonce roots (see Table 4) and the differences with respect to *-Ir* use among these 16 possible rhymes proves to be significant (X-squared = 246.0527, df = 15, p-value < 0.001). As is apparent in Table 4, there is great variability in the templates tested, in respect of to what extent they triggered *-Ir* use.

<sup>3</sup> To account for the variation in the *-Ir* use, Chi-squared tests of independence were run on R 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. Section 2), focusing on the frequencies of monosyllabic verbs in the aorist forms.

Furthermore type and token frequency of stored exemplars do not appear to correlate with participants' answers. More precisely, one would expect more *-Ir* use when there are more rhyming *-Ir* tokens and types. For example, in the case of the rhyme template *-ul*, there are no *-Ar* types/tokens, hence the only existing verb in the language is *-Ir* type (i.e., *bul-ur* 'finds'), yet *-Ir* use is only 16%. The templates *-ür*, *-ir*, *-or*, *-il* and *-ül*, however, have no *-Ir* types, hence tokens, but they appear to have triggered *-Ir* use from 20% to 5%. The rhyme template *-ur* -the template that has given rise to the most *-Ir* use (31%)- is in the neighborhood of two *-Ir* types i.e., *dur-ur* 'stops' and *vur-ur* 'hits' and two *-Ar* types, *kur-ar* 'sets' and *bur-ar* 'twists'. So typewise there is no winner, but tokenwise *-Ir* appears to get the upperhand with a total of 11398 tokens, and it could be that the 31% *-Ir* use with *-ur* is correlated with that.

**Table 4.** Rhymes of Nonce Stems and Irregularization Rates according to Type and Token Frequencies of Regular and Irregular Monosyllabics in BOUN corpus

In Type ▼ In Token ►	Irregular > Regular	Irregular < Regular
<b>Irregular &gt; Regular</b>	ul (16%)	
<b>Irregular = Regular</b>	ur (31%), ör (19%), öl (9%)	
<b>Irregular &lt; Regular</b>	el (22%), al (16%), er (13%), ol (9%), ar (7%), il (5%)	ür (20%), ir (20%), ir (15%), or (12%), ıl (12%), ül (5%)

We conclude that, contrary to what has been postulated for English past tense, rhyme does not consistently explain the recourse to the so-called 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* affix. Inspired by the study of Hahn and Bailey (2005) on similarity, we investigate other measures of similarity at the phonemic level and test the effect of the following four measures on the *-Ir* use:

- (i) C\_C Pattern of the nonce root (First and Last consonant)
- (ii) First consonant of the nonce root
- (iii) Vowel of the nonce root
- (iv) Final consonant of the nonce root

(i) We first investigate the effect of sharing the consonantal pattern (onset-coda) with the stored exemplars. We have found that *-Ir* use differed significantly among categories of nonce verbs sharing the same first and last consonants, even after a continuity correction was applied since more than 20% of the counts in the categories were less than 5 (X-squared = 221.902, df = 38, p-value < 0.001). We have observed that the more the nonce root has existing irregulars in number (type) with the same C\_C (if any), the more it is irregularized, i.e. used with *-Ir*. For example, the C\_C patterns /v\_r/, /b\_l/, /g\_r/, /g\_l/, /k\_l/, /d\_r/) in analogy with the irregular verbs *ver-* 'give', *bul-* 'find', *gör-* 'see', *gel-* 'come', *kal-* 'stay' and *dur-* 'stop' have been used with the *-Ir* affix significantly more than others. The least irregularized nonce roots among those sharing the same C\_C with existing verbs (e.g., /d\_l/, /s\_r/, /ç\_l/, /s\_l/) are found to be sharing it with several regulars but no irregulars.

(ii) We have further examined whether the initial consonant of the nonce root *per se* has any role in triggering the *-Ir* use. Even though there is variability in items starting with the same consonant, some first consonants have triggered significantly more *-Ir* use than others: the difference in *-Ir* use between the categories of nonce verbs defined by their first consonant is significant (X-squared = 108.3059, df = 19, p-value < 0.001). For example, for the nonce verbs the initial consonant of which is /v/ (3 existing irregulars), we have encountered an *-Ir*

use of 25%, however, the initial consonant /t/ (no irregular) only triggered an *-Ir* use of 5%. Just like the C\_C pattern, *-Ir* use for nonce verbs seems to correlate with the type number of existing irregular verbs starting with the same consonant. Thus, high rates of *-Ir* use cluster around existing irregulars (*vur*, *ver*, *var*, *al*, *bul*, etc.), low rates around existing regulars (*sar*, *sol*, *kar*, *çal*, etc.).

(iii) Among the eight vowels that constitute the nucleus of the CVC we have observed a significant difference with respect to the *-Ir* use (X-squared = 93.387, df = 7, p-value < 0.001). While the high, back, rounded /u/ in the root rendered the most *-Ir* use (23%), the high, front, unrounded /i/ yielded the least (10%). 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 (X-squared = 15.4717, df = 1, p-value = 0.001). 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 root containing the back vowel *zur-* was irregularized as *zur-ur* at a proportion of 32%, while its front version the nonce root *zür-* was irregularized as *zür-ür* at only 17% of the uses. We have, however, observed no significant difference between rounded (14%) and unrounded (15%) vowels (X-squared = 1.0884, df = 1, p-value = 0.4765), nor between high (14%) and non-high (15%) vowels with respect to *-Ir* vs. *-Ar* use (X-squared = 1.3908, df = 1, p-value = 0.4765). We have further observed that the more existing irregular verbs there are with the same vowel, the more irregularized a nonce root sharing the same vowel is. For instance, the vowel /u/ in a nonce root has yielded the highest *-Ir* use and /u/ also happens to be the vowel with the highest number of existing verbs that behave irregularly, i.e., *dur-/durur* ‘stops’, *vur-/vurur* ‘hits’, *bul-/bulur*, ‘finds’, while the vowel with the lowest *-Ir* use, /i/, has only one existing irregular, i.e., *bil-/bilir* ‘knows’. Here token frequency appears not to play a role at all, since there are 26,660 tokens of irregular verbs containing the vowel /u/, as opposed to 36,313 tokens of irregular verbs containing the vowel /i/.

(iv) Considering the final consonant of the root, we have observed that /r/-ending nonce stems have yielded *-Ir* use (19%) significantly more than /l/-ending nonce stems (11%) (X-squared = 79.4109, df = 1, p-value < 0.001). Contrary to the other measures of similarity, surprisingly neither the type nor the token frequency of the monosyllabic verbs in the language appear to correlate with the participants' answers, since /l/-ending irregular verbs are more frequent in type and token than /r/-ending irregulars. However, we have observed that the frequency of trigrams /VrV/ and /VlV/ may impact participants' answers. The sequences /Vru, Vrü, Vrı, Vri/ corresponding to the 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 consonant /l/, we see that the trigrams /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 supports a higher use of *-Ir* after /r/ than after /l/. Frequencies of corresponding quadrigrams also confirm our expectations of finding more *-Ir* forms after /r/ than after /l/: the quadrigrams /Vrur, Vrür, Vrır, Vrir/ are more frequent than /Vrar, Vrer/ in type and token while the quadrigrams /Vlar, Vler/ are more frequent than /Vlur, Vlür, Vlır, Vlir/. For example, /örür/ is more frequent than /örer/, confirming the use of *zör-ür* over the regular *zör-er* for the nonce root *zör-* which has been produced with *-Ir* by the participants with a rate of 31%. The quadrigram /olar/, however, is more frequent than /olur/,

this time supporting the use of regular *zol-ar* over the irregular *zol-ur* which is irregularized at 3% by the participants. We now turn to the discussion of the presented results.

## 6. General Discussion

The results of the present study show that at the adult stage, when tested on nonce roots that resemble sonorant-ending monosyllabic verbs in Turkish, an already attained symbolic abstraction in the form of *-Ar*, takes the center stage and rules the behavior of the participants. Complemented with the acquisition data, the results of the current study help us draw a more complete picture of the representation of aorist in native speakers' minds. Nakipoğlu and Ketz (2006) and (under review) demonstrate that at 3 years of age, children learning Turkish make more overregularization errors – they use *-Ar* for *-Ir* taking monosyllabic verbs (*\*öl-er* instead of *öl-ür*). From 4;6 years onward, children make more irregularization errors, i.e., they use *-Ir* for *-Ar* taking verbs (*\*bin-ir* instead of *bin-er*), and continue erring in the use of aorist well into the school age years. Acquisition findings convincingly show that an abstraction in the form of a rule is at work from 4;6 years of age for sonorant-ending irregularly behaving monosyllabic verbs as the rate of children's overregularization errors drops from 39% to 20% around four and a half years of age. This significant leveling off in the error rate appears to correlate with the high type frequency of the *-Ar* suffix with monosyllabic verbs in Turkish. Thus an abstraction is already in the making in child Turkish from 4;6 years-of-age. To converge on the adult representation, in other words, to master the irregular and regular forms, at some point during acquisition it seems that the child brain has to start developing insensitivity to the phonological properties of the root, i.e., the sonority of the final consonant. We conjecture that it is primarily the type frequency of *-Ar* within the domain of monosyllabic verbs that renders this possible. At the adult stage, participants' overwhelming *-Ar* use with sonorant-ending nonce-roots, in our view strongly confirms the unfolding of the abstraction and the default status of *-Ar*.

Pinker (2001) has suggested that the default form must be the result of symbolic composition, not storage in memory, and must 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*). Because participants are found to resort to the use of the *-Ar* form for nonce verbs with such high rate (86%), 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. 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. It should be noted that recent years have seen an evolution of the mathematical frameworks (Bayesian Probability Theory, Machine Learning), allowing for new accounts of language processing and acquisition, that are probabilistic in nature (Chater & Manning 2006; Tenenbaum et al. 2011). So adults could 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 nonce verbs.

Our work also takes place among recent studies investigating irregular patterns other than the English past tense, checking the validity of the Single Mechanism and Dual Mechanism models against acquisition and processing data. The literature on the issue is far from converging to consensus: on one hand, Italian verbs (Orsolini & Marslen-Wilson 1997), French verbs (Meunier & Marslen-Wilson 2004), Dutch plurals (Keuleers et al. 2007) and Serbian nouns that encode number, gender and case (Mirković et al. 2011), support the Single



Mechanism model and exhibit similarity-based generalizations. On the other hand, Spanish verbs (Clahsen et al. 2002), Portuguese verbs, in particular 1<sup>st</sup> person conjugation (Veríssimo & Clahsen 2014) and Hungarian plurals and past tense (Nemeth et al. 2015) advocate the Dual Mechanism model as regulars and irregulars appear to be processed differently and regulars to generalize beyond similarity and exhibit a rule-driven generalization.

In the overall Turkish adult productions, the role of analogy has been scarce. 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. Sharing the (i) C\_C Pattern, (ii) First consonant, (iii) Vowel and (iv) Last Consonant appear to play a more active role in relating nonce stems with existing features of the Turkish language compared to the Rhyme variable. In particular, *-Ir* use for nonce verbs seem to be correlated with the type frequency of existing irregulars sharing the same C\_C pattern, the same first consonant C\_\_ or the same vowel \_V\_: for a given nonce stem, if any such irregulars exist in the language, the higher the number of types, the higher the *-Ir* use. Finally, the Last Consonant of the nonce roots seems to have a decisive influence on the outcome: the *-Ir* forms appear significantly more with nonce verbs ending in /r/ (19%) than with nonce verbs 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 /VrV/ and /VIV/ trigrams and /VrVr/ and /VIVr/ quadrigrams in Turkish. That the *-Ir* use did not exceed 14%, however, is a clear indication that a rule-driven mechanism must be blocking further analogizing across exemplars. Thus the results of the present study indicate that adults' processing of the Turkish aorist, as it pertains to the irregularity observed in the domain of monosyllabic verbs, is sensitive to type frequency and is best explained by a rule-driven generalization where analogy clearly plays a minor role.

## Acknowledgments

The presentation of this work at MMM12 was made possible through the support of SYSTRAN to Elise Michon and Boğaziçi University Research Fund (Grant No: 14041) to Mine Nakipoğlu.

## References

- Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in cognitive sciences*, 10(7), 335-344.
- Clahsen, H., Avelledo, F., Roca, I. (2002). The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language*, 29: 591-622.
- Hahn, U., & Bailey, T. M. (2005). What makes words sound similar?. *Cognition*, 97(3), 227-267.
- Hare, M., Elman, J. L., & Daugherty, K. G. (1995). Default generalisation in connectionist networks. *Language and Cognitive Processes*, 10(6), 601-630.
- Keuleers, E., Sandra, D., Daemans, W. Gillis, S., Durieux, G. and Martens, E. (2007). Dutch plural inflection: the exception that proves the analogy. *Cognitive Psychology*, 54(4):283-318.
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the society for research in child development*, i-178.
- 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.
- Meunier, F., & Marslen-Wilson, W. (2004). Regularity and irregularity in French verbal inflection. *Language and Cognitive Processes*, 2004, 19 (4), 561-580.
- Mirković, J., Seidenberg, M. S., Joanisse, M. F. (2011). Rules vs. Statistics: Insights from a Highly Inflected Language. *Cognitive Science* 35 (4): 638-681.
- Nakipoğlu, M., & Ketz, N. (2006). Children's overregularizations and irregularizations of the Turkish aorist. In *BUCLD 30: Proceedings of the Boston University Conference on Language Development* (Vol2, 399-410).

- Nakipoğlu, M. & Ketrez, N. (under review). The irregularity that proves the rule: Insights from an agglutinative language.
- Nakisa, R. C., Plunkett, K. & Hahn, U. (2001). A cross-linguistic comparison of single and dual-route models of inflectional morphology. *Cognitive Models of Language Acquisition*.
- Nemeth, D., Janacek, K., Turi, Z., Lukacs, A., Peckham, D., Szanka, S., Gazso, D., Lovassy, N., & Ullman, M. T. (2015). The production of nominal and verbal inflection in an agglutinative language: Evidence from Hungarian. *PLoS one*, 10(3), e0119003.
- Orsolini, M., & Marslen-Wilson, W.D. (1997). Universals in morphological Representation: Evidence from Italian. *Language and Cognitive Processes*, 12 (1), 1-47.
- 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.
- Pinker, S. (2001). Four decades of rules and associations, or whatever happened to the past tense debate. *Language, the brain, and cognitive development: Papers in honor of Jacques Mehler*, 157-179.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in cognitive sciences*, 6(11), 456-463.
- Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48(1), 21-69.
- 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 *Advances in natural language processing* (pp. 417-427). Springer Berlin Heidelberg.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022), 1279-1285.
- 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.