Glossa a journal of general linguistics

Ahyad, Honaida and Michael Becker. 2020. Vowel unpredictability in Hijazi Arabic monosyllabic verbs. *Glossa: a journal of general linguistics* 5(1): 32. 1–18. DOI: https://doi.org/10.5334/gjgl.814

RESEARCH

Vowel unpredictability in Hijazi Arabic monosyllabic verbs

Honaida Ahyad[1] and Michael Becker[2]

[1] Stony Brook University, Stony Brook, NY, US

[2] University of Massachusetts Amherst, Amherst, MA, US

Corresponding author: Michael Becker (michael@linguist.umass.edu)

We study the distribution of vowels in the monosyllabic verbs of Urban Hijazi Arabic, showing that speakers use the presence of a root emphatic consonant to partially predict the quality of stem vowels. The effect of the emphatic is observed in the lexicon, and is productively extended to nonce verbs, showing that speakers generalize over lexical representations that include both vowels and consonants; the purely consonantal representations that are commonly assumed for Arabic are insufficient to capture speakers' knowledge of consonant-vowel interactions. We propose a probabilistic analysis that learns lexical trends from surface forms and extends them productively to nonce words.

Keywords: templatic morphology; Arabic dialects; understudied dialects; lexical statistics; productivity; learnability

1 Introduction

This paper analyzes the distribution of vowels in the monosyllabic verbs of Urban Hijazi Arabic, exemplified in Table 1. The vowel in a monosyllabic verb is predictably low in the perfective, while imperfective vowels vary by lexical item. In this paper, we show that the choice of imperfective vowel is partially predictable from the consonantal context, and argue in favor of lexical representations that allow simultaneous access to both vowels and consonants, with a grammar that productively and stochastically derives imperfectives from perfectives for existing verbs and novel verbs.

Urban Hijazi Arabic is spoken by a few million people in the cities of the Hijaz area of western Saudi Arabia (Makkah, Jeddah, Taif, and Madinah) and is easily mutually intelligible with neighboring urban dialects such as Egyptian and Levantine. The vowels of this dialect are short [a i u] and long [aa ee ii oo uu].

There are two types of monosyllabic verbs in the language: short vowel verbs (traditionally known as doubled, bi-consonantal, or geminated), and long vowel verbs (traditionally

Table 1: The two types of monosyllabic verbs in Urban Hijazi Arabic.

type	vowels	PERF.3SG.M	IMPERF.3SG.M	PERF.1SG	
short vowel	a ~ i	tamm	ji-timm	tammeet	'agree'
(doubled)	a ~ u	baχχ	ji-buχχ	baχχeet	'spray'
long vowel	aa ~ aa	baan	ji-baan	bint	'appear'
(hollow)	aa ~ ii	saab	ji-siib	sibt	'leave'
	aa ~ uu	ʃaaf	ji-ʃuuf	ʃuft	'see'

called hollow verbs). As shown in Table 1, short vowel verbs always have [a] in the perfective forms (e.g., [tamm] 'he agreed', [tammeet] 'I agreed'), while imperfective forms of the verb have either [u] or [i], with the choice depending on the lexical item (e.g. [ji-timm] 'he agrees'). Long vowel verbs are a little more varied. In the third person perfective forms, the vowel is always [aa] (e.g. [baan] 'he appeared', [baan-at] 'she appeard', [baan-u] 'they appeared'), while in first and second person perfective forms, the vowel is [i] or [u], with the choice depending on the lexical item (e.g. [bin-t] 'I appeared', [bin-na] 'we appeared', [bin-tu] 'you.PL appeared'). In the imperfective, the vowel may be [aa], [ii], or [uu] (e.g. [ji-baan] 'he appears'). In this paper, we focus on the third person singular masculine forms, comparing perfective and imperfective vowels. We also limit ourselves to the active forms; the Modern Standard Arabic passive is not generally used in Hijazi.

We will show that the distribution of vowels in the lexicon is sensitive to the presence of an emphatic (pharyngealized) consonant in the root. Urban Hijazi Arabic contrasts the plain [t d s z] with the emphatic [tˁ dˁ sˁ zˁ], e.g. [taar] 'he was furious' vs. [tˁaar] 'he flew'. In long vowel verbs, roots that have an emphatic consonant are more likely to have the front vowel [ii] in the imperfective, whereas roots without an emphatic are more likely to have [uu]. In short vowel verbs, the trend is the opposite: a root with an emphatic consonant is more likely to have an imperfective back vowel [u] compared to a root without an emphatic. We will show that this connection between root emphatics and imperfective vowels is not merely a fact about the lexicon, but it is also productively extended to nonce verbs, and thus forms a part of the native speaker's grammar.

Our results add a new type of empirical support for the view that Arabic verbs are stored in the lexicon with their imperfective vowels, and thus necessitate lexical representations that combine vowels and consonants, as argued for by Gafos (2003); Berent et al. (2007); Bat-El (2017) and others for other reasons. The results cannot disprove the existence of a more abstract level of representation that separates consonants from vowels, as proposed e.g. by Ussishkin et al. (2015), but they show that purely consonantal representations are insufficient to explain the productive connection between consonants and vowels.

The paper is organized as follows: first, §2 surveys a lexical database that we compiled, showing a significant interaction between the presence of an emphatic, vowel length, and vowel quality. A stochastic grammatical model that encodes the observed trends is offered in §3. Then in §4, we present a nonce word study that shows the productivity of the lexical trends, and §5 concludes.

2 The monosyllabic verbal lexicon

This section surveys the monosyllabic verbs of Urban Hijazi Arabic, showing that the presence of an emphatic consonant has a significant effect on the choice of imperfective vowel. We start in §2.1 with a quantitative look at a lexical database that we created, showing the effect of emphatic consonants on the selection of the vowel in the imperfective. We discuss the phonetic naturalness of the effect. In §2.2 we survey the emphatic effect in other dialects. In §2.3 we develop the argument that lexical representations in Arabic require vowels and consonants to be simultaneously accessible, and §2.4 concludes.

2.1 Quantitative lexicon study

To find all of the monosyllabic verbs in Urban Hijazi, the first author, who is a native speaker, examined all of the two-consonant combinations that are possible in the language, and marked the ones that are attested with either a short vowel or a long vowel. The list was then verified by two additional native speakers for accuracy.

A total of 238 monosyllabic verbs were identified, 133 short vowel verbs and 105 long vowel verbs, as summarized in Table 2; a full list is given in the supplementary files. All of the verbs have [a] or [aa] in their third person perfective forms. In the imperfective,

short vowel verbs usually have [u] (82%), and the rest have [i]. In the long vowel verbs, [ii] and [uu] are about equally frequent, with only four verbs selecting [aa] in the imperfective.

The choice of vowel in the imperfective is partially predictable from the surrounding consonants, in particular the emphatics [tˤ dˤ sˤ zˤ]. Table 3 shows the correlation between the presence of an emphatic in the stem and the selection of an imperfective vowel: in short vowel verbs, [u] is selected more strongly with an emphatic (97% vs. 77%), while the opposite is true in long vowel verbs (32% vs. 56%).

To assess the strength of these trends in the lexicon and to predict their application to novel items, we fitted a logistic regression model using the *glm* function in R (R Core Team 2016), with the choice of [u]/[uu] vs. [i]/[ii] as the dependent variable, omitting the four items that have [aa] in the imperfective. We started with a model that contained the binary predictor *short vs. long vowel*. Adding a binary predictor *plain vs. emphatic* and the interaction with vowel length significantly improved the model. Following the recommendation of Gelman & Hill (2007: §4.2, §5.5), both predictors were centered. The predictor *short vs. long vowel* was entered with the values –.43 for short vowels and +.57 for long vowels, and *plain vs. emphatic* was entered with –.24 for plain consonants and +.76 for emphatic consonants. This model is reported in Table 4. It shows that overall, long vowels are correlated with significantly less [u/uu] in the imperfective than short vowels. The presence of an emphatic has no significant effect overall. Most importantly, the interaction with vowel length is significant: emphatics significantly decrease the probability of [u/uu] with long vowels and significantly increase the probability of [u/uu] with short vowels. The statistical model predicts that vowel length and the presence of an emphatic will together bias the selection of an imperfective vowel in novel items.

Table 2: Monosyllabic verbs in Urban Hijazi Arabic (*n* = 238); [u]/[uu] is the most frequent imperfective vowel.

verb type	vowels	*n*	%	PERF	IMPERF	gloss
short vowel	a ~ i	24	18%	tamm	ji-timm	‘agree’
(doubled)	a ~ u	109	82%	baχχ	ji-buχχ	‘spray’
long vowel	aa ~ aa	4	4%	baan	ji-baan	‘appear’
(hollow)	aa ~ ii	48	46%	saab	ji-siib	‘leave’
	aa ~ uu	53	50%	ʃaaf	ji-ʃuuf	‘see’

Table 3: Urban Hizaji Arabic monosyllabic verbs: emphatics correlate with more [u] in short vowel verbs and less [uu] in long vowel verbs.

verb type		vowels	*n*		PERF	IMPERF	gloss
short	plain	a ~ i	23		ʕall	ji-ʕill	‘sicken’
		a ~ u	78	77% [u]	baχχ	ji-buχχ	‘spray’
	emphatic	a ~ i	1		dˤall	ji-dˤill	‘get lost’
		a ~ u	31	97% [u]	tˤaχχ	ji-tˤuχχ	‘shoot’
long	plain	aa ~ aa	4		naam	ji-naam	‘sleep’
		aa ~ ii	31		ʕaad	ji-ʕiid	‘repeat’
		aa ~ uu	45	56% [uu]	ʃaaf	ji-ʃuuf	‘see’
	emphatic	aa ~ aa	0		—	—	—
		aa ~ ii	17		tˤaar	ji-tˤiir	‘fly’
		aa ~ uu	8	32% [uu]	sˤaam	ji-sˤuum	‘fast’

Table 4: The choice of imperfective vowel is sensitive to vowel length and the presence of an emphatic in the stem. Negative estimates (β) indicate a skew away from [u/uu].

	β	SE(β)	z	p(>|z|)
(Intercept)	1.04	.20	5.35	<.001
plain vs. emphatic	.77	.63	1.23	>.1
short vs. long vowel	−1.66	.37	−4.51	<.001
emphatic:long	−3.34	1.15	−2.90	<.005

A connection between the presence of an emphatic consonant and vowel backness is not surprising: Emphatics and back vowels share a lowered F2, and vowels are generally phonetically retracted in the presence of an emphatic consonant (see, e.g. Zawaydeh 1999 on Jordanian Arabic, and a recent review in Alammar 2017). For example, in the minimal pair [ji-dill] 'guide' and [ji-dˤill] 'mislead', both stem vowels are categorically front, but the vowel [i] is produced with lower F2 (further back) next to the emphatic. Similarly, in the minimal pair [ji-subb] 'swear' and [ji-sˤubb] 'pour', both stem vowels are categorically back, but the [u] is pronounced with lower F2 (further back) next to the emphatic. That is to say, while emphatics cause allophonic backing of both front and back vowels, the front-back contrast is maintained in Urban Hijazi. We thus observe two different effects of emphatics in this language: all vowels are allophonically backed in the presence of an emphatic in the language as a whole, and additionally, short imperfective vowels are more often back rather than front in the presence of an emphatic. As for the long vowels, they are similarly allophonically backed when adjacent to an emphatic overall in the language, but imperfective vowels are more often front in the presence of an emphatic — perhaps a case of dissimilation in a morphologically restricted environment.

For the sake of completeness, we mention that the effect of emphatics is also observed more generally in the verbal system of the language. Urban Hijazi, like other dialects of Arabic, strongly limits the shape of verbs; verbal stems are either monosyllabic or disyllabic but not longer.[1] In addition to the monosyllabic verbs surveyed above, our complete Urban Hijazi list of measure I verbs also contains 603 disyllabic verbs, all with short vowels, of which 166 contain an emphatic. Emphatics increase the incidence of imperfective [u] in disyllabic verbs: the imperfective vowel is [u] in 31% of disyllables without an emphatic (e.g. [nabat ~ ji-nbut] 'grow') and 50% of disyllables with an emphatic (e.g. [rabatˤ ~ ji-rbutˤ] 'tie'). Imperfective short vowels are therefore uniformly affected by emphatics in Hijazi, showing a preference for [u] in both monosyllabic and disyllabic verbs.

One main difference between monosyllabic and disyllabic verbs is that disyllabic verbs allow short [a] in the imperfective, in addition to short [i] and [u]. We attribute this difference to syllable structure: imperfective [a] is possible before the simple coda of a disyllabic verb, e.g. [ʃamal ~ ji-ʃmal] 'include', but only the short high vowels, which are phonetically the shortest in the language, are allowed before the geminate coda of a monosyllabic verb, e.g. *[mall ~ ji-mall].

2.2 Comparison with other dialects

Consonants have been reported to have an effect on short vowels in other dialects of Arabic, in particular gutturals (uvulars, pharyngeals, and glottals) which have an affinity for [a], and emphatics (pharyngealized or uvularized consonants) which have an affinity for [u].

[1] Further, verbs must fit into one of many templates (also called "measures", "forms", or "binyanim") that determine their syllable structure, vowels, and affixes. There is more phonological variability in "measure I" verbs, which includes all of the monosyllables.

Looking at Modern Standard Arabic, McCarthy (1994) surveys the perfective and imperfective short vowel verbs in Wehr's (1976) dictionary, and finds a strong bidirectional connection between gutturals and imperfective [a]: 94% of the verbs that take an imperfective [a] have a guttural, and 77% of the verbs with a guttural take an imperfective [a]. McCarthy does not find an effect of emphatics in his data, and suggests that the choice between imperfective [i] and [u] is entirely unpredictable.

To compare the monosyllables of Urban Hijazi with the monosyllables of Modern Standard Arabic, we collected all 263 of the monosyllabic verbs from the Modern Standard Arabic section of Wikitionary: 136 short vowel verbs (doubled, geminated) and 127 long vowel verbs (hollow), as shown in Table 5. The short vowel verbs show a preference for imperfective [u], and there is slightly more [u] with root emphatics (66% vs. 69%), just as in Hijazi, but the effect is much smaller. The long vowel verbs take less [uu] with emphatics (60% vs. 50%), as in Hijazi, but again the effect is smaller than it is in Hijazi (cf. Table 3). Using the same statistical test as in Table 4 above, the effect of the emphatic is not significant in Modern Standard Arabic with either verb type.

Compared with Urban Hijazi, Modern Standard imperfective vowels are less impacted by consonants, and therefore less predictable and more contrastive. Modern Standard also allows low imperfective vowels more freely, adding another dimension of contrast, or lack of predictability: imperfective short [a] is allowed, unlike in Hijazi. However, imperfective short [a] is quite rare in monosyllabic verbs, and even in the presence of gutturals it does not go beyond 17% — compare with McCarthy's finding of 77% [a] with gutturals for all verbs (most of which are disyllabic). We conclude that a constraint against imperfective short [a] in monosyllabic verbs is active in both dialects, but its effect is gradient in Modern Standard and categorical in Urban Hijazi.

In Palestinian Arabic, imperfective short vowels are highly predictable: Herzallah (1990) reports that short [u] is only possible in verbs that have an emphatic. In this dialect, the imperfective vowel is predictably [a] in the presence of a guttural (e.g. [sa?al ~ ji-s?al] 'ask'), it is mostly [u] and occasionally [i] in verbs that have an emphatic, and predictably [i] in verbs that have neither guttural nor emphatic. Herzallah does not discuss any differences between monosyllabic and disyllabic verbs, but her examples of imperfective [a] are limited to disyllabic verbs. We turned to two linguists, one Palestinian and one Lebanese, and both broadly confirmed Herzallah's generalizations, but they were unable to identify any monosyllables with imperfective short [a]. This suggests that imperfective short [a] is disallowed in monosyllables, just like in Urban Hijazi. We suspect that short imperfective [a] is disallowed in monosyllables in Egyptian and other dialects as well.

Going beyond the imperfective, Blanc (1964) reports that in Iraqi Arabic, perfective verbs tend to surface with short [u] in the presence of a root emphatic, e.g. [tˤubaχ] 'he cooked' vs. short [e] without the emphatic, e.g. [ketab] 'he wrote'.

Table 5: Modern Standard Arabic monosyllabic verbs (*n* = 263): weak effect of emphatic consonants.

verb type		imperfective vowel			
		a	i	u	
short (doubled)	plain	13	34	73	66% [u]
	emphatic	3	10	18	69% [u]
		aa	ii	uu	
long (hollow)	plain	12	33	63	60% [uu]
	emphatic	1	11	11	50% [uu]

In several different dialects of Arabic, then, short vowels show the affinity of emphatics for [u]. Additionally, gutturals prefer [a], but this preference is mostly limited to disyllabic verbs. In monosyllabic verbs, short [a] is either rare or absent, even in the presence of a guttural. As for long vowels, they show a preference for the front [ii] in the presence of an emphatic, weakly and non-significantly in Modern Standard, but more strongly and significantly in Urban Hijazi. We have no information about long imperfective vowels in other dialects.

Across dialects, imperfective vowels exhibit a scale of predictability or lexicality: imperfective vowels are least predictable from consonantal context in Modern Standard, more predictable in Urban Hijazi, and most predictable in Palestinian.

Generalizations about the distribution of short vowels can be gradient in one dialect but categorical in another: for example, imperfective short [a] in monosyllables is completely prohibited in Hijazi, but in Modern Standard it is allowed yet dispreferred relative to imperfective short [a] in disyllables. In a cross-dialectal analysis, the same constraint would apply in both dialects, gradiently in one and categorically in the other. Similarly, the connection between short imperfective [u] and emphatics is weak in Modern Standard, stronger and significant in Hijazi, and categorical in Palestinian, where short [u] requires the presence of an emphatic. Again, the same constraint applies to different degrees. In all three dialects, however, some lexical listing is required: for example, some Palestinian short vowel verbs with an emphatic take [i] rather than [u], and this information must be learned and listed for these individual lexical items.

2.3 Word-based representations in the Arabic lexicon

Much of the work on Arabic and other Semitic languages separates vowels from consonants in underlying representations, assuming that roots are purely consonantal; this assumption underlies much of the traditional work in generative linguistics (e.g. Brame 1970; McCarthy 1979; a.o.) and in more recent experimental work (Frost et al. 2000; Boudelaa & Marslen-Wilson 2001; 2011; Ussishkin et al. 2015).

However, certain aspects of Semitic morphophonology require the presence of underlying vowels in lexical representations, as discussed by Gafos (2003), who shows that several systematic aspects of Modern Standard Arabic morphophonology follow if vowels and consonants coexist in underlying representations, such as the ban on initial geminates (*mmVd); similar arguments for a word- rather than root-based verbal morphology are in Benmamoun (1999); Teeple (2007). Further support for lexical representations that combine vowels and consonants comes from Berent et al. (2007), who show that speakers are sensitive to the type frequencies of vowels that combine with repeated consonants in Hebrew, e.g. [meded] is less frequent than [midud]; these type frequencies are accessible only if vowels and consonants coexist in lexical representations.

Our results offer an additional argument in favor of lexical representations that combine vowels and consonants, since speakers choose imperfective vowels in accordance with the frequency of their cooccurrence with consonants in the lexicon. As Berent et al. (2007) explain, it is logically possible that lexical items have additional representations that are purely consonantal, and Gafos (2003: §6.4) suggests that perhaps these consonantal representations are accessed in certain kinds of tasks.

In the Urban Hijazi lexicon, emphatic consonants can partially predict the quality of imperfective vowels, and speakers project the lexical trends onto novel words. We suggest that the productivity of the lexical trends is due to a probabilistic grammar that is learned from the lexicon, as proposed by Zuraw (2000); Ernestus & Baayen (2003); Hayes & Londe (2006), among others. Evidence for the mediation of a grammar in learning gradient lexical trends comes from effects of phonetic naturalness in the learning of trends:

lexical trends are learned best when they can be represented by phonetically motivated constraints, e.g. voicing of consonants depending on their place of articulation (Ernestus & Baayen 2003). Unnatural connections, e.g. between consonant voicing and the backness of an adjacent vowel, are not learned (Becker et al. 2011; 2012) or learned weakly (Hayes et al. 2009; Hayes & White 2013).

An advantage of Zuraw's (2000) USELISTED model and later similar models (e.g. Zuraw 2010) is that a single constraint, e.g. a constraint that prefers [u] with an emphatic, LICENSE(back), can be used in multiple dialects either gradiently or categorically, in our case gradiently in Hijazi and categorically in Palestinian. However, these models require the analyst to choose the constraints, and are therefore unable to discover the relevant generalizations on their own. To learn with less supervision, we use the MGL (Albright & Hayes 2002; 2003; 2006, see §3), since it has the power to discover the connections between vowel quality and consonant quality in the lexicon without the analyst's supervision, and to use them to derive novel words probabilistically. A complete model that combines the strength of the MGL with a flexible architecture that represents both gradient and categorical effects in the same grammar is still in the future (although see Wilson 2017 for constraint induction in the analysis of alternations).

Our proposal for Urban Hijazi is that imperfectives are derived from the surface forms of perfectives, e.g. the long [uu] of imperfective [ji-ʃuuf] 'see' is derived from the long [aa] of [ʃaaf] (see §3). Brame (1970), and later Rosenthall (2006), provide an analysis of long vowel verbs (hollow verbs) in Modern Standard Arabic, deriving imperfective [uu] and [aa] from underlying /w/ and imperfective [ii] from underlying /j/ and (see also Chekayri & Scheer 2005). The advantage of this approach is that it connects some monosyllabic verbs to morphologically related verbs (e.g., causatives) in which the glide appears on the surface. For example, it connects the [ii] of [maal ~ ji-miil] 'bend' to the [jj] of [majjal] 'cause to bend', and similarly it connects the [uu] of [χaan ~ ji-χuun] 'cheat' to the [ww] of [χawwan] 'distrust'; the vowels of the imperfective are related to the glides of the causative through a shared underlying representation. Additionally, this analysis correctly predicts the perfective vowel when consonant-initial suffixes are added, e.g. [mil-t] 'I bent', [χun-t] 'I cheated'. The glide-based analysis is further supported by the absence of singleton glides between low vowels in Modern Standard Arabic (*awa, *aja).

This glide-based analysis of long vowel verbs cannot be straightforwardly extended to Urban Hijazi, Palestinian, or Egyptian, since these dialects have surface singleton [w] between low vowels in several verbs, e.g. [ħawal ~ ji-ħwil] 'be cross-eyed', [dawaʃ ~ ji-dwiʃ] 'irritate' (Hinds & Badawi 1987); our Hijazi lexicon has a total of eight verbs with [w] between low vowels. Therefore, underlying glides cannot uniformly give rise to surface long vowels in these dialects. Our MGL analysis takes the perfective long vowel as an input and predicts the imperfective long vowel stochastically based on neighboring consonants. Admittedly, this MGL analysis cannot account for the appearance of glides in causatives, as it only analyses two morphological categories at a time, but the MGL can derive causatives from imperfectives in a separate analysis, e.g. [ji-χuun → χawwan] 'betray', [ji-miil → majjal] 'bend'.

In all of the dialects that we surveyed, the imperfective vowel is not completely predictable, and therefore some lexical listing of imperfective vowels would be needed in any theory. We go further to claim that Arabic verbs are stored in the lexicon using representations that combine vowels and consonants, or representations that allow simultaneous access to vowels and consonants, since such representations are needed for a statistical learner to identify the partial dependence of vowels on consonantal quality.

2.4 Local summary

We presented a lexical database of 238 monosyllabic verbs, showing a significant effect of an emphatic consonant on the choice of imperfective vowel. The imperfective vowels are only partially predictable, and therefore imperfective vowels must be stored for existing lexical items, but we will see that their distribution is extended productively to novel words.

The effect of the emphatic in Urban Hijazi was compared to similar effects in other dialects, and in particular to a categorical effect of emphatics in Palestinian. We propose that imperfectives surface forms are derived from perfective surface forms that include both vowels and consonants. Competing proposals that derive long vowel verbs from underlying glides are undermined by the grammaticality of surface singleton glides in Urban Hijazi and other neighboring dialects.

3 Predictive models of alternations

For our analysis of the interaction between vowels and consonants in Urban Hijazi verbs, we aim to provide a model that learns the mapping from perfective to imperfective, while also learning the effect of the consonants on vowels from the existing verbs of the language, and then extending its knowledge by creating imperfective forms of novel perfective verbs. This task, central to the goal of generative linguistics, can currently only be carried out by one learner: The Minimal Generalization Learner (MGL, Albright & Hayes 2002; 2003; 2006). The learner is freely available from Adam Albright's website (http://www.mit.edu/~albright/mgl/).

Other computational models of alternations are limited in some ways: the Sublexical Learner (Gouskova & Becker 2013; Allen & Becker 2015) creates derivatives productively, but relies on the analyst to identify the relevant environments for any given alternation. Wilson's (2017) learner identifies relevant environments, but does not generate novel derivatives. Hayes & Wilson (2008) is a phonotactic learner, meaning it assigns probabilities to individual words, and thus it is not suitable for studying paradigms, such as the perfective-imperfective pairs studied here. Outside of generative linguistics, there is a plethora of analogical models with excellent performance, but no human-interpretable internal components; these black boxes do not have any internal components such as rules or constraints that can be reused by a human analyst.

Albright & Hayes's MGL is built to learn relations between two morphological categories, e.g. the perfective and imperfective. It accepts a list of two-word paradigms, in their unanalyzed surface forms, and it identifies the changes between the two forms and the local environments that condition the change. First, it creates a rule for each paradigm by removing identical material from the edges and isolating the change; for example, given the paradigm [sann ~ sunn] 'whet' (without the imperfective prefix), it removes the identical [s] segments from the left edges of the perfective and imperfective, and removes the identical [nn] from the right edges, isolating the change [a → u]. Similarly, the paradigm [mann ~ munn] 'guilt' is analyzed as having the same [a → u] change, as seen in Table 6a–b. The MGL made such rules for each of the 238 paradigms, isolating the four vowel changes

Table 6: Rule generation with the MGL (Albright & Hayes 2002; 2003; 2006).

	change			environment			
a.	[a]	→	[u]	/	s	__ nn	sann ~ sunn
b.	[a]	→	[u]	/	m	__ nn	mann ~ munn
c.	[a]	→	[u]	/	[+cons]	__ nn	generalized rule

[a → u], [aa → uu], [a → i], [aa → ii], and "no change" for the case of imperfective long [aa], e.g. [naam ~ naam] 'sleep'. Note that the imperfective prefix was removed in order to allow the MGL to isolate the stem vowel changes; including the prefix would have been counter-productive, e.g. in [sann ~ ji-sunn] the shared edge material would have been just the final [nn], isolating [sa → jisu], and similarly in [mann ~ ji-munn] the isolated change would have been [ma → jimu], failing to identify the two paradigms as undergoing the same change. The exclusion of the prefix is a simple work-around for the MGL's limited ability to isolate changes from complete surface forms.

After a rule is created for each paradigm, the MGL takes pairs of paradigms that share the same change, such as [a → u], and creates more general rules by identifying shared features, such as the rule in Table 6c. Here, [s] and [m] only share the feature [+consonantal]. This more general rule can now apply to all verbs that match its structural description, correctly in the case of [ʃann ~ ji-ʃunn] 'attack', but incorrectly in the case of [ʔann ~ ji-ʔinn] 'moan'. The MGL assigns each rule a confidence score that is calculated based on the number of paradigms it derives correctly and those it derives incorrectly. This process of generalization and assignment of confidence scores continues with further pairwise comparisons of rules, progressively creating rules with increasing generality.

Using the 238 paradigms in the lexicon, the MGL created 45,434 rules. We visually inspected some of these, and noticed that emphatics appeared in many of the most reliable rules. The classes of liquids [l, r] and nasals [m, n] appeared often as well, but to a lesser extent. The confidence of each rule depends on the number of forms it derives correctly relative to the number of forms it potentially applies to, e.g. the rule of [a → u] in the environment of an emphatic consonant has high confidence, because of the 32 verbs with short [a] in the environment of an emphatic, 31 verbs actually take [u], yielding a confidence score of 31/32, or 97%. In contrast, the more general rule of [a → u] in the environment of any consonant(s) has lower confidence, because of the 132 verbs with short [a], only 109 actually take [u], yielding a confidence score of 109/132, or 83%. The MGL further adjusts these confidence scores to favor rules with wider coverage (the reader is referred to Albright & Hayes 2002 for details). Given a novel perfective, the MGL will derive imperfectives for it using the rules with the highest confidence that match its structural description, and therefore a perfective with an emphatic will most likely be derived using one of the [a → u] rules, while a perfective without an emphatic is somewhat more likely to be derived with one of the [a → i] rules.

To generate MGL predictions for nonce monosyllabic verbs, we used all the possible combinations of CVC syllables with [a] or [aa] as the nucleus, e.g. [ʔaʔ], [ʔaaʔ], [ʔab], etc. With 26 consonants in this dialect, the total number of potential nonce verbs was 26*26*2 = 1352. The MGL generated imperfectives for all of these, using the most reliable rules that match any given novel perfective. For example, given the novel [hann], the MGL generated [hunn] based on a rule of [a → u] in the environment [+cont]__[–cont], which applies correctly to 35 of the 38 real verbs that match this environment. It also generated [hinn] based on a rule of [a → i] in the environment [dorsal]__[n], which applies correctly to all 3 of the real verbs that match this environment. Despite being exceptionless, the rule [a → i]/[dorsal]__[n] has lower confidence due to the small number of words it can apply to, and therefore the derivative [hunn] has higher confidence than the derivative [hinn]. The entire simulation, with its input files and output files, is available at becker.phonologist.org/hijazi.

From this list of 1352 potential nonce verbs, we excluded the real words and then selected 30 nonce verbs with a short vowel and 30 nonce verbs with a long vowel. Within each group, 15 verbs were selected from those that have a strong preference for a front

vowel (i.e. high confidence assigned to imperfective [i/ii] and low to imperfective [u/uu]) and 15 with a strong preference for a back vowel. This method also correlated with the presence of emphatics: for short vowel verbs, there were emphatics present in the high [u] confidence group, while for the long vowel verbs, there were emphatics present in the high [ii] confidence group. The list of selected verbs with their confidence scores (as well as the experimental results) are provided in the supplementary files.

The MGL can also be used to map in the other direction, from imperfective to perfective. In the case of monosyllabic verbs, the result is entirely predictable: all imperfective stem vowels become low in the perfective. The MGL derives this result by identifying the necessary changes, e.g. [u → a], [uu → aa], [i → a], etc. Since only one of these changes can apply to any given imperfective form, the MGL correctly lowers all stem vowels when generating perfective forms.

In one central way, the MGL does not mimic human behavior: it does not memorize existing lexical items. It creates general rules from the lexical items given to it, but does not use item-specific knowledge in its derivations. Thus, for example, given the perfective [sann] 'whet', it generates both the correct imperfective [ji-sunn] and the incorrect [ji-sinn], and assigns some confidence to each form. A native speaker of Urban Hijazi would identify [sann] as a real word, and would only accept the attested [ji-sunn], rejecting the possible but unattested [ji-sinn]. Albright & Hayes do not suggest a remedy for this shortcoming, essentially neglecting to implement the Elsewhere Principle (Kiparsky 1973). The MGL could be improved by adding a mechanism such as USELISTED (Zuraw 2000), which blocks productive application of lexical trends when an existing derivative is listed. In our materials, we manually removed the real words from the list of potential nonce words.

To summarize, the MGL is unique in generative linguistics in its ability to learn mappings between two morphological categories and the environments for these mappings, and then productively extend this learned knowledge to novel words. Compared to handcrafted generative analyses, the MGL has two limitations: first, its ability to isolate changes is rather limited, but we were able to work around this limitation by removing the imperfective prefix from the lexicon. Second, it only computes two morphological categories at a time, in our case the third person singular masculine perfective and imperfective. The MGL's strength is in its ability to mimic human learning, paying attention to subtle details in the the data, and productively generating novel phonological representations.

4 Nonce monosyllables: Vowel-emphatic interaction

We present a nonce word task (Berko 1958), asking native speakers of Urban Hijazi Arabic to judge novel imperfective verbs. The results show that speakers productively extend the correlation between the presence of an emphatic and the quality of the imperfective vowel.

4.1 Participants

Participants were recruited through social media (Facebook, Instagram, and WhatsApp) from the Hijaz area of Saudi Arabia (the cities of Makkah, Jeddah, Taif, and Madinah). They volunteered their time and effort. The experiment was conducted online using Experigen (Becker & Levine 2015).

Speakers in the Hijaz generally speak one of two different dialects, known as Urban and Bedouin. Since our lexicon study is based on the Urban dialect, we sought to limit ourselves to Urban participants. To this end, participants were asked to fill a demographic form at the end of the experiment, asking about their gender, year of birth, where they were born, and six questions about their dialect. In each of these six questions, a sentence was followed by two options that contained a morphological, phonological, or syntactic feature that distinguished the Urban dialect from the Bedouin dialect. We included in our study participants who chose the Urban option for at least five of the six questions, who

completed the experiment, and who indicated that they are at least 18 years old, discarding the rest. This left us with data from 104 participants.

Of the total 104 participants, 86 self-identified as female and 18 as male. All identified as being from the Hijaz. The average self-reported age was 32, range 18–50.

4.2 Materials

To prepare the verbs for auditory presentation, each of the 60 nonce verb paradigms from §3 were recorded, where each paradigm consisted of one monosyllabic perfective base and two possible imperfective forms, e.g. [naad ~ ji-niid, ji-nuud]. The existing verb [ʃaaf] 'see' was included as an example item.

The verb forms were recorded by the first author three times in random order in a quiet room. Using Praat (Boersma & Weenink 2015), the best token was selected for each form, and then converted into mp3 format. The audio files were not manipulated in any way.

4.3 Procedure

The experiment was administered online using Experigen (Becker & Levine 2015). Participants were free to use any browser of their choice. To keep the experiment reasonably short, the server made a random selection of a total of 28 items for each participant out of the total 60 items, balanced for vowel length and predicted vowel quality (7 of the 15 items that were available per condition). We judged that 7 items per condition are sufficient for measuring by-participant effects.

Before the beginning of the experiment, the participants were asked to put their headphones on and listen to a recording of the word [ʃaaf] 'see' and then press a button indicating whether they heard [ʃaaf], [ʃuuf], or [ʃiif]. In a following screen, written instructions in standard Arabic explained that nonce verbs will be presented for judgment in the Hijazi dialect, and participants were asked to judge them based on their vernacular.

Items were presented in individual screens as schematized in Figure 1, with an auditorily presented perfective, followed by two auditorily presented imperfectives in random order. The selection buttons appeared only after the sound files were played in order. Participants were free to click the sound buttons more than once.

After all the items were presented, participants were asked to provide demographic information as explained above. On average, participants took 9 minutes to complete the task (range 4–27, median 9).

4.4 Results

Overall, participants chose back vowels in the imperfective at similar rates for both short and long vowels, 58% and 56% respectively. The presence of an emphatic consonant in the base had a significant effect on the choice of vowel, as seen in the right panel of Figure 2: in

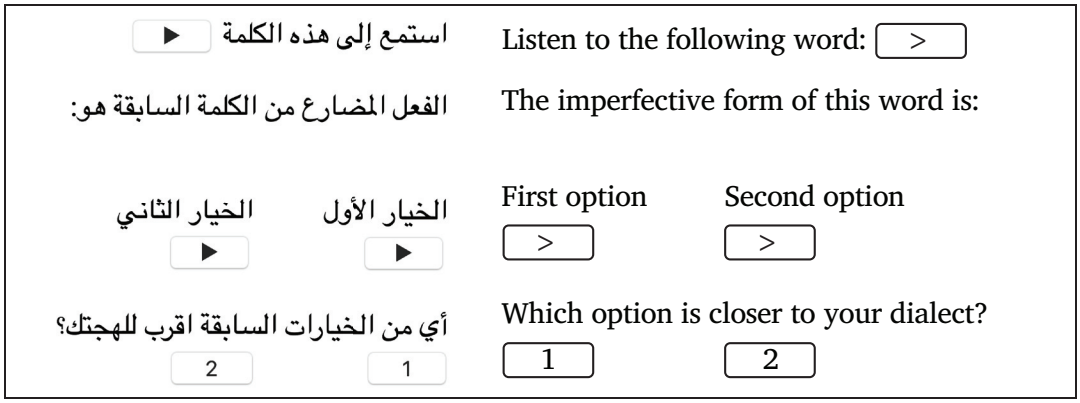

Figure 1: Sample trial (original on the left, translation on the right).

short vowel verbs, emphatics correlated with choosing more back vowels, whereas in long vowel verbs, emphatics correlated with choosing fewer back vowels, mirroring the lexical trends that we identified in §2.1. The raw experimental results are available at becker.phonologist.org/hijazi/.

The results were assessed with a mixed effects logistic regression model using the *glmer* function from the *lme4* package (Bates et al. 2015). The variables were the same ones used in §2.1; again the dependent variable was the selection of a back vowel [u, uu] vs. [i, ii] in the imperfective. Following the recommendation of Gelman & Hill (2007: §4.2, §5.5), the predictors were centered, i.e. their values were chosen such that their average would be zero. Two binary predictors were used: *short vs. long vowel*, with a value of –.5 for short vowels and +.5 for long vowels, and *plain vs. emphatic*, with a value of –.34 for verbs without an emphatic and +.66 for verbs with an emphatic, as well as the interaction of the two. A fully crossed model was fitted first, with random intercepts for item and participant and random slopes for *short vs. long vowel* and *plain vs. emphatic* and the interaction given participant. This model did not converge, so the random slope for the interaction was removed. The resulting model is reported in Table 7. Correlations of the fixed effects are all less than .21, i.e. the model is reasonably free of collinearity (see e.g. Baayen 2008: §6.2.2).

While *short vs. long vowel* and *plain vs. emphatic* had no significant main effects, their interaction was highly significant: emphatics significantly decreased the choice of imperfective back vowels in long vowel verbs and significantly increased the choice of imperfective back vowels in short vowel verbs.

The result of the experiment mirrors the lexicon in two ways: both in the lexicon and in the experiment, there is a significant interaction of *emphatic* consonants with vowel

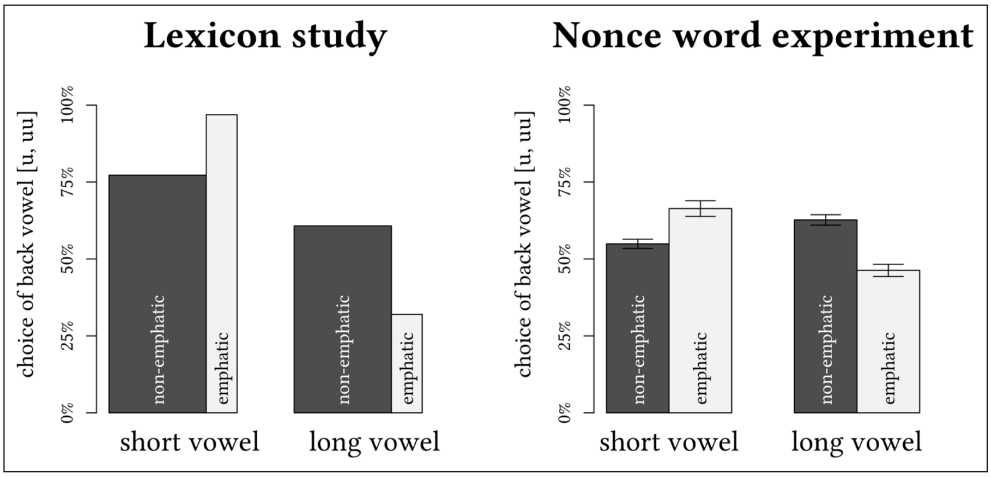

Figure 2: Choice of [u, uu] by vowel length and presence of emphatic in the lexicon and the nonce word study.

Table 7: Regression model for the experiment, 104 participants. Negative estimates indicate a skew away from [u/uu].

	β	SE(β)	*z*	*p*(>|*z*|)
(Intercept)	.41	.12	3.28	
plain vs. emphatic	–.08	.22	–.34	>.1
short vs. long vowel	–.08	.21	–.40	>.1
emphatic:long vowel	–1.40	.44	–3.15	<.005

length, but no main effect of *emphatic*. In one way, however, the lexicon and the experiment diverge: in the lexicon, there is a main effect of vowel length, such that long vowel verbs tend to choose front vowels in the imperfective. In the experiment, vowel selection is less extreme: long vowel verbs do not select front vowels as strongly as they do in the lexicon, and short vowel verbs do not select back vowels as strongly as they do in the lexicon. We attribute this difference to the observation that experimental items generally elicit less extreme reactions than established lexical items, as discussed e.g. by Zuraw (2000), and that overall preferences in this type of experiment are sensitive to the experimental methods rather than to overall lexical baselines, as discussed by Albright & Hayes (2003). Additionally, the attenuated result might also be due to the limited phonetic motivation for connecting vowel length and vowel backness (cf. Becker et al. 2011; Hayes & White 2013).

4.5 Correlation with the MGL predictions

The nonce verbs for the experiment were selected based on the MGL predictions, which in turn correlated with the presence of an emphatic consonant. The model in §4.4 confirmed the significance of the emphatic effect; here we assess the strength of the MGL predictions.

According to the MGL, the most productive rules are the general [a → u] and [aa → uu] that apply regardless of consonantal quality, reflecting the high type frequency of [u/uu] in the lexicon (54% of all monosyllabic verbs, see also Table 2). Further rules refer to features of neighboring consonants, in particular emphatics, as well as liquids and nasals.

For short vowel verbs, the MGL predictions always strongly favor [u], either without an emphatic (about 93%) or with an emphatic (97%). For long vowel verbs, the prediction for plain verbs is 86% [uu], and the presence of an emphatic reduces the prediction to 53%, as seen in Figure 3. Since the MGL predictions are not normally distributed, the correlation was assessed with Spearman non-parametric test; the correlation is highly significant ($p < .005$) but moderate ($\rho = .40$).

The MGL is unique in its ability to predict vowel quality in novel imperfective verbs based on the trends in the lexicon. Categorical generative analyses of Arabic verbs (e.g., Brame 1970) aim to derive possible imperfective forms from various properties of underlying

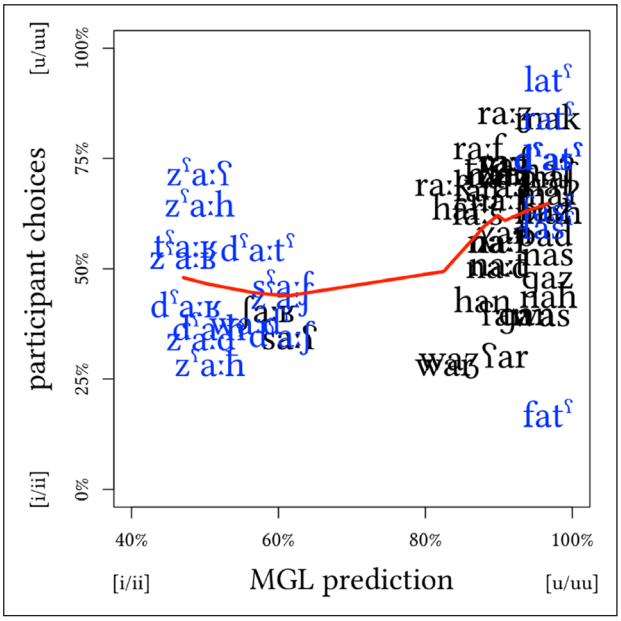

Figure 3: Participant choices as a function of the MGL predictions with a red LOWESS line (locally weighted average line). Plain stems are in black, stems with emphatics are in blue.

representations, such as the selection of a glide, but do not predict whether a given novel verb is more likely to take a front or a back vowel. For example, given the nonce perfectives [raaf] and [zˤaaʕ], these analyses predict that the possible imperfective vowels are [aa], [ii] and [uu], but they do not predict differences in acceptability between the three vowels, either independently or in reference to the emphatic consonant. The results of the current study, however, show that such detailed predictions form a part of the native speakers' knowledge.

4.6 Summary

This section presented the results of a large-scale nonce word task, showing that native speakers of Hijazi Arabic judge nonce verbs using detailed knowledge about the existing verbs in their lexicon. In short vowel verbs, participants chose significantly more [u] in the presence of an emphatic, and in long vowel verbs they chose significantly less [uu] in the presence of an emphatic. The predictions of the MGL model were significantly positively correlated with the choices of the native speakers.

5 Conclusions

We presented a lexicon study and a nonce word experiment on the monosyllabic verbs in Urban Hijazi Arabic, showing that the presence of an emphatic consonant in the verbal stem is a significant predictor of the stem vowel in the imperfective. This connection between stem consonants and stem vowels is strong in the lexicon, and it is shown to be productive in a large-scale nonce word task with 104 participants. The same connection between emphatics and short [u] is also observed in the disyllabic verbs of the language. As for the effect of emphatics more broadly in the language, e.g. in the nominal system, no relevant data is currently available. We also compared monosyllabic verbs in Urban Hijazi to monosyllabic verbs in two other varieties of Arabic, Modern Standard and Palestinian, and showed that emphatics prefer short [u] in all three, with the effect being strongest in Palestinian and weakest in Modern Standard.

Our analysis uses the surface-based Minimal Generalization Learner (MGL Albright & Hayes 2002; 2003; 2006). The Learner was given pairs of perfective-imperfective stems, based on which it created general rules that probablistically predict the imperfective vowel in reference to its consonantal environment, and in particular the presence of emphatic consonants. Other natural classes, such as liquids and nasals, had a weaker effect, which we omitted here for brevity. A strong connection between emphasis (pharyngealization) and vowel backness is phonetically plausible, as both are cued by F2, and the selection of imperfective back vowels in the presence of an emphatic can be construed as assimilation. Becker et al. (2011; 2012); Hayes et al. (2009); Hayes & White (2013) suggest that speakers are particularly likely to notice and extended phonetically motivated trends in the lexicon.

Less phonetically motivated, however, is the mediation of vowel length, with imperfective long vowels more likely to be front in the presence of an emphatic, even though these vowels are allophonically backed by the emphatic. This effect might be construed as morphologically-restricted dissimilation. The MGL analysis we present learns statistically reliable vowel-consonant interactions in the lexicon regardless of their phonetic plausibility or naturalness.

The main strength of the MGL analysis is its ability to detect trends in the lexicon with minimal supervision. However, it is limited to modeling a pair of morphological categories at a time, e.g. the mapping from perfective to imperfective. On its own, then, the analysis of this mapping does not capture some broader generalizations, e.g. the connection to the mapping from the imperfective to related causatives. A fuller model of Arabic morphophonology will require an architecture that captures these and other

generalizations, both gradient and categorical, such as the ban on initial identical stem consonants (*mmVd). We hope that such a model will emerge in the future.

The productive extension of the vowel-consonant interactions inside the stem suggests that speakers have access to lexical representations that include both vowels and consonants, much like the lexical representations that are standardly assumed outside of Semitic. Learning vowel-consonant connections presupposes simultaneous access to vowels and consonants, and therefore lexical representation in Arabic must include both vowels and consonants. Furthermore, the dialect comparison in §2.2 shows that the strength of the emphatic effect varies across dialects, and therefore the emphatic effect must be learned by generalizing over the lexicons of individual language varieties. Phonetic naturalness provides the direction of the effect but not its magnitude.

Our results provide a new argument in favor of lexical representations that combine vowels and consonants, and thus add support to the proposals in Gafos (2003); Berent et al. (2007); Bat-El (2017), and others. As noted by these authors, the evidence points towards lexical representations that combine vowels and consonants, but the evidence cannot disprove the notion of a purely consonantal root, as defended in Frost et al. (2000); Ussishkin et al. (2015) and many others. Such purely consonantal representations, however, would have to coexist with the full representations that we assume. As Bat-El (2017) notes, linguists assume that roots are separable from affixes in all languages, and further that consonants and vowels can be accessed separately in all languages — therefore root consonants can be accessed separately in all languages.

Abbreviations

PERF = perfective, IMPERF = imperfective, SG = singular, PL = plural, M = masculine

Acknowledgements

For their thoughtful comments and suggestions, we thank Mahasen Abu Mansour, Sam Alxatib, Christina Bethin, Ellen Broselow, Stuart Davis, Hovsep Dolatian, Maria Gouskova, Larry Hyman, Bob Hoberman, Ghada Khattab, John McCarthy, and the audience of 43rd meeting of the Berkeley Linguistic Society. We also thank Adamantios Gafos and the anonymous reviewers for their constructive criticism. Finally, we thank the anonymous participants in our studies who volunteered their time and effort and the community members who helped with recruitment.

Competing Interests

The authors have no competing interests to declare.

References

Alammar, Ammar. 2017. *Emphasis in Zilfaawi Arabic*: Stony Brook University dissertation.

Albright, Adam & Bruce Hayes. 2002. Modeling English past tense intuitions with minimal generalization. In Michael Maxwell (ed.), *Proceedings of the sixth meeting of the acl special interest group in computational phonology*, 58–69. Philadelphia, PA: ACL. DOI: https://doi.org/10.3115/1118647.1118654

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90. 119–161. DOI: https://doi.org/10.1016/S0010-0277(03)00146-X

Albright, Adam & Bruce Hayes. 2006. Modeling productivity with the gradual learning algorithm: The problem of accidentally exceptionless generalizations. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradience in grammar*, 185–204. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199274796.003.0010

Allen, Blake & Michael Becker. 2015. Learning alternations from surface forms with sublexical phonology. Lingbuzz/002503.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511801686

Bat-El, Outi. 2017. Word-based items-and processes (WoBIP): Evidence from Hebrew morphology. In Claire Bowern, Larry Horn & Raffaella Zanuttini (eds.), *On looking into words (and beyond)*, 115–135. Berlin: Language Sciences Press.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. http://CRAN.R-project.org/package=lme4. R package version 0.999999-4. DOI: https://doi.org/10.18637/jss.v067.i01

Becker, Michael, Andrew Nevins & Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88(2). 231–268. DOI: https://doi.org/10.1353/lan.2012.0049

Becker, Michael & Jonathan Levine. 2015. Experigen – an online experiment platform. Available at https://github.com/tlozoot/experigen.

Becker, Michael, Nihan Ketrez & Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1). 84–125. DOI: https://doi.org/10.1353/lan.2011.0016

Benmamoun, Elabbas. 1999. Arabic morphology: The central role of the imperfective. *Lingua* 108. 175–201. DOI: https://doi.org/10.1016/S0024-3841(98)00045-X

Berent, Iris, Vered Vaknin & Gary Marcus. 2007. Roots, stems, and the universality of lexical representations: Evidence from Hebrew. *Cognition* 104(2). 254–286. DOI: https://doi.org/10.1016/j.cognition.2006.06.002

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177. DOI: https://doi.org/10.1080/00437956.1958.11659661

Blanc, Haim. 1964. *Communal dialects in Baghdad*. Cambridge, MA: Harvard University Press.

Boersma, Paul & David Weenink. 2015. Praat: Doing phonetics by computer. [Computer program]. http://www.praat.org/.

Boudelaa, Sami & William D. Marslen-Wilson. 2001. Morphological units in the Arabic mental lexicon. *Cognition* 81. 65–92. DOI: https://doi.org/10.1016/S0010-0277(01)00119-6

Boudelaa, Sami & William D. Marslen-Wilson. 2011. Productivity and priming: morphemic decomposition in Arabic. *Language and Cognitive Processes* 26(4–6). 624–652. DOI: https://doi.org/10.1080/01690965.2010.521022

Brame, Michael. 1970. *Arabic phonology: Implications for phonological theory and general Semitic*: MIT dissertation.

Chekayri, Abdellah & Tobias Scheer. 2005. Biliteral approach to weak verbs in Arabic. *Paper presented at the 19th Arabic Linguistics Symposium*. DOI: https://doi.org/10.1163/9789047405085_006

Ernestus, Miriam & R. Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79(1). 5–38. DOI: https://doi.org/10.1353/lan.2003.0076

Frost, Ram, Avital Deutsch & Kenneth I. Forster. 2000. Decomposing morphologically complex words in a nonlinear morphology. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(3). 751–765. DOI: https://doi.org/10.1037/0278-7393.26.3.751

Gafos, Adamantios. 2003. Greenberg's asymmetry in Arabic: A consequence of stems in paradigms. *Language* 79(2). 317–355. DOI: https://doi.org/10.1353/lan.2003.0116

Gelman, Andrew & Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511790942

Gouskova, Maria & Michael Becker. 2013. Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory* 31(3). 735–765. DOI: https://doi.org/10.1007/s11049-013-9197-5

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3). 379–440. DOI: https://doi.org/10.1162/ling.2008.39.3.379

Hayes, Bruce & James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44. 45–75. DOI: https://doi.org/10.1162/LING_a_00119

Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85(4). 822–863. DOI: https://doi.org/10.1353/lan.0.0169

Hayes, Bruce & Zsuzsa Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104. DOI: https://doi.org/10.1017/S0952675706000765

Herzallah, Rukayyah. 1990. *Aspects of Palestinian Arabic phonology: A nonlinear approach*: Cornell University dissertation.

Hinds, Martin & El-Said Badawi. 1987. *A dictionary of Egyptian Arabic*. French & European Publications.

Kiparsky, Paul. 1973. 'Elsewhere' in phonology. In Stephen Anderson & Paul Kiparsky (eds.), *A festschrift for Morris Halle*, 93–106.

McCarthy, John J. 1979. *Formal problems in Semitic phonology and morphology*: MIT dissertation.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.), *Papers in Laboratory Phonology III: Phonological structure and phonetic form*, 191–233. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511659461.012

R Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria. (http://www.r-project.org).

Rosenthall, Samuel. 2006. Glide distribution in clasical Arabic verb stems. *Linguistic Inquiry* 37. 405–440. DOI: https://doi.org/10.1162/ling.2006.37.3.405

Teeple, David. 2007. Intra-paradigmatic contrast in Arabic verbal morphology. In *UC Santa Cruz: Working Papers* (Phonology at Santa Cruz), Retrieved from https://escholarship.org/uc/item/9342n2v6.

Ussishkin, Adam, Colin Reimer Dawson, Andrew Wedel & Kevin Schluter. 2015. Auditory masked priming in Maltese spoken word recognition. *Language, Cognition and Neuroscience* 30(9). 1096–1115. DOI: https://doi.org/10.1080/23273798.2015.1005635

Wehr, Hans. 1976. *A dictionary of modern written Arabic*. Ithaca, New York NY: Spoken Language Services 3rd edn.

Wilson, Colin. 2017. Learning constraints for morphophonological classification. Talk given at AMP 2017, https://github.com/colincwilson/maxentclassphon.

Zawaydeh, Bushra. 1999. *The phonetics and phonology of gutturals in arabic*: Indiana University dissertation.

Zuraw, Kie. 2000. *Patterned exceptions in phonology*: UCLA dissertation.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28(2). 417–472. DOI: https://doi.org/10.1007/s11049-010-9095-z

How to cite this article: Ahyad, Honaida and Michael Becker. 2020. Vowel unpredictability in Hijazi Arabic monosyllabic verbs. *Glossa: a journal of general linguistics* 5(1): 32. 1–18. DOI: https://doi.org/10.5334/gjgl.814

Submitted: 21 September 2018 **Accepted:** 03 May 2019 **Published:** 16 March 2020

]u[*Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS