Variable Hiatus in Persian is Affected by Suffix Length

Koorosh Ariyaee and Peter Jurgec
University of Toronto

1 Introduction

Sequences of vowels are often avoided. In fact, many languages ban hiatus outright (Casali 1998, 2011). Languages differ in how they resolve hiatus. For instance, Yoruba deletes the first vowel, Chichewa merges the vowels, while Malay epenthesizes a consonant between the vowels. But do these well-described patterns present a full range of cross-linguistic variation?

In this paper, we examine hiatus in Persian, which displays three rare properties. First, Persian most commonly avoids hiatus by deleting the second vowel, which is considerably less frequent than deletion of the first vowel across languages (Casali 1997). Second, while Persian generally avoids hiatus, it tolerates it when the suffix contains a single vowel, as to maintain paradigmatic contrast. Finally, hiatus in Persian is variable: more often than not, multiple realizations are attested within and across speakers.

Our work contributes to a growing body of research that has focused on variable sound patterns, which are subject to various restrictions just as much as exceptionless generalizations are. For instance, Ernestus & Baayen (2003) show that Dutch speakers' productions of nonce words reflect distributional characteristics of the Dutch lexicon, with velars in root-final position eliciting relatively more voicing than labials and coronals. Becker et al. (2011) demonstrate that variable laryngeal alternations in Turkish depend on word length and consonant place. Jurgec & Schertz (2020) show that variable velar palatalization in Slovenian depends on the suffix and is restricted by a consonant co-occurrence restriction on postalveolars. Persian hiatus constitutes another example.

This paper is organized as follows. Section 2 presents the language background and reviews the literature on hiatus in Persian. Section 3 presents the elicitation-based production experiment which allows us to gauge the extent of variation. Section 4 discusses the perception experiment in which we focus on three main variants: hiatus, elision of the second vowel, and epenthesis. We show that while elision is the most common variant, hiatus is generally retained with suffixes consisting of a single vowel.

2 Background

The variety we examine in this paper is Spoken Persian, which is also known as Tehrani Persian, Standard Spoken Persian, and vernacular Persian. The variety of Spoken Persian we take as the object of analysis is the prestigious variety used in the media and in formal and informal communication (Modaressi 1978, Kalbasi 2001, Ariyaee to appear).

Spoken Persian has six vowels [i e æ ɑ o u] (Majidi & Ternes 1999). When these vowels appear at the root-suffix boundary, the realizations vary. In (1) we compare two roots. The consonant-final root 'office' shows no variation: every suffix starts with a vowel, which is retained. The vowel-final root 'here', however, shows variation with the suffixes. The suffix vowel (or second vowel, V_2) can be deleted, [ʔ] can be epethesized or hiatus can be preserved. However, not all possibilities are equally attested: if the suffix consists of a single vowel, V_2 elision is strongly dispreferred ($^{???/*}$inʤɑ-ǿ), and V_1 elision is always ungrammatical (*inʤǿ-æm). The epenthetic variants are dispreferred by the speakers.

---

(1)   'COPULA-1P.SG'   'COPULA-3P.PL'   'COPULA-3P.SG'

    dæftær   dæftæræm   dæftæræn   dæftære   'office'

    inʤɑ   inʤɑæm   inʤɑæn   inʤɑe   'here'   (hiatus)

            ˀinʤɑˀæm   ˀinʤɑˀæn   ˀinʤɑˀe   (epenthesis)

            inʤɑm   inʤɑn   ???/*inʤɑ   (V_2 elision)

            *inʤæm   *inʤæn   *inʤe   (V_1 elision)

The Persian hiatus pattern is cross-linguistically rare. First, Casali (1997) reports only two other languages (Basque and Kagate) that have V_2 elision but never V_1 elision. Second, hiatus avoidance is typically exceptionless or its distribution is predictable based on morphological factors (see Garrido 2013 for such a case in Spanish). Third, hiatus in Persian shows a great deal of inter- and intraspeaker variability.

The variability is mirrored in the existing literature on Persian hiatus: while Sadeghi (1986), Shaghaghi (2000), Dehghan & Kord Zafaranlu Kambuziya (2012) focus on epenthesis, Jam (2015) studies elision, and Estaji et al. (2010), Yazarlou (2014), Ghatreh et al. (2020) suggest hiatus is retained. Beyond these basic facts, little is known about the nature of variation in Persian hiatus.

In this study we investigate the main variants of Persian hiatus altogether and discuss the factors that motivate such variation. To tackle these issues, we conduct two experiments. In the elicitation-based production experiment we will gauge the variation across and within speakers. The results reveal a few core variants and dependence of one of them on suffix length. Then, the perception experiment will allow us to focus on the effect of suffix length. As we we will see, the variation in Persian hiatus is systematic and highly dependent on the length of the suffix.

3 Production

One way to extract information about variable phonological phenomena is to look at corpus data (Zuraw 2006; Kager & Pater 2012; Becker & Jurgec 2020, among many others). To the best of our knowledge, no corpus of Spoken Persian exists. To gain insight into the general properties of hiatus and its resolution in spoken Persian (such as consistency and variation within and across speakers, the effect of vowel quality and lexical effects), we thus conducted a small elicitation-based production experiment.

3.1 *Methods*

3.1.1 *Stimuli*   In this study, we focus on hiatus at the root-suffix boundary. Our stimuli for the elicitation task were polymorphemic words consisting of a vowel-final root and vowel-initial suffix. Table 1 shows the variables and their levels in this task. Our primary variable is Suffix Length, as evidenced in (1). Suffixes are categorized into two groups: polysegmental and monosegmental. With our second variable, Lexical Category, we tackle the question whether the patterns observed in native words are productively extended to loanwords (see Yip 1993; Itô & Mester 1995; Jurgec 2010; Kang 2011) and nonce words (Berko 1958 et seq.). Here native words include established Arabic loans.

	Variable	*Levels*	*Example item*	
a.	Suffix Length	Polysegmental	inʤɑ-<u>æm</u>	'here-COPULA-1P.SG'
		Monosegmental	bɑbɑ-<u>e</u>	'the dad'
b.	Lexical Category	Native	<u>xune</u>-e	'the home'
		Loans	<u>instɑ</u>-e	'the Instagram'
		Nonce	<u>xirʤɑ</u>-e	'the (nonce)'

Table 1: Elicitation-based production experiment variables and their levels.

Due to lexical gaps and phonological restrictions, the variables are not perfectly balanced. To start with, we used all 17 productive suffixes in Spoken Persian, but there are more polysegmental suffixes than monosegmental ones, as shown in Table 2. These suffixes begin with all of the vowels of the Persian vowel inventory, except for [u].

Polysegmental (10)		*Monosegmental* (7)	
-im	'copula-1P.PL'	-i	'copula-2P.SG'
-in	'copula-2P.PL'	-i	'adjective'
-et	'possessive-2P.SG'	-i	'noun'
-eʃ	'possessive-3P.SG'	-e	'copula-3P.SG'
-emun	'possessive-1P.PL'	-e	'definite'
-etun	'possessive-2P.PL'	-ɑ	'plural'
-eʃun	'possessive-3P.PL'	-o	'conjunction'
-æm	'copula-1P.SG'		
-æm	'possessive-1P.SG'		
-æn	'copula-3P.PL'		

Table 2: The exhaustive list of suffixes in Spoken Persian.

We included a total of 195 vowel-final roots per participant. The roots ended in every vowel of Spoken Persian, except for [æ], which is not allowed root-finally. This means that not all vowel combinations are possible in Spoken Persian: there were no *æ*-initial and *u*-final hiatus combinations.

Each root belonged to one of the three lexical categories. The native roots include nativized Arabic loans (75 roots in total). The reason for considering Arabic loans as part of the native category is that these borrowings are completely adapted to the Persian phonology, since Persian has a long history of borrowing from Arabic (Morony 1986). Arabic loans also differ from other loanwords (60 roots in the experiment), which were borrowed into Persian more recently from English, Russian, and French (Zomorodian 1994; Navabzadeh Shafi'i 2014; Ariyaee 2019). The nonce words (60 roots) used in the elicitations were consistent with the Persian phonotactics. They were either CVCV, CVCCV or CVCVCV, with final stress. All nonce words were checked by a native speaker to make sure they sound natural and are not attested words.

3.1.2 *Procedure*   The data elicitation procedure included two parts: the familiarization task and the main task. The familiarization task included four stages. At the first stage, participants were presented with a consonant-final root in Persian orthography. At the second stage, they heard a derived word, consisting of the same consonant-final root with a vowel-initial suffix. Then, the participant was presented with a new consonant-final root in Persian orthography. Finally, the participant was asked to derive the word using the suffix they heard before.

In the main task, the participants were similarly presented with a vowel-final root, and were then asked to derive it with a particular suffix. Once the participants pronounced the derived word, they were asked if there is another way to pronounce it. When the pronunciation was unclear, the participants were asked to repeat the word. The stimuli were organized into blocks by suffix to make the task easier for the participants.

3.1.3 *Participants*   We recruited 6 participants, of which 3 were female and 3 were male, with the mean age of 30. One of the participants participated twice (with the two sessions being apart more than a month) so to that we could better gauge the variation within speaker. Thus, this speaker had twice as much data as the other participants.

3.1.4 *Analysis*   The productions were then manually categorized into three categories: V_2 elision, hiatus, and epenthesis. This categorization was conducted by a native speaker both impressionistically and using an acoustic analysis. It was challenging to distinguish between hiatus and [ʔ] epenthesis. While there were many cases of a clear [ʔ], there were also 17 instances of glottalization and creaky voice. All these cases were transcribed as having a [ʔ] and included in the results we report below. Another challenging distinction involved glides. Persian allows [w] and [j] intervocalically to resolve hiatus in specific contexts (Sadeghi 1986; Shaghaghi 2000; Dehghan & Kord Zafaranlu Kambuziya 2012). The glide [j] can be epenthesized intervocalically after an [i], while [w] can be epenthesized next to a round vowel; the [w] contexts also variably allow [ʔ] epenthesis (Sadeghi 1986; Windfuhr & Perry 2009). Unless there was evidence of a glottal stop, the glides were transcribed in all expected environments, since we could not find a reliable method to distinguish glide epenthesis from hiatus acoustically. Across all speakers, there were 180 such tokens. Beyond these cases, there were also a few instances of allomorph selection for one particular suffix, which

we leave out from what follows. There were no other realizations.

3.2 *Results*

3.2.1 *Suffix Length*   Our primary variable of interest is Suffix Length. Figure 1 shows the share of three realizations across all participants by Suffix Length. The area represents the relative share among all tokens. We can see that V_2 elision (henceforth, simply elision) is frequent with polysegemtanl suffixes, but rare with monosegmental suffixes. Epenthesis is more common with monosegmental suffixes, but it is the rarest of the three variants. Finally, hiatus is the most frequent realization with monosegmental suffixes.
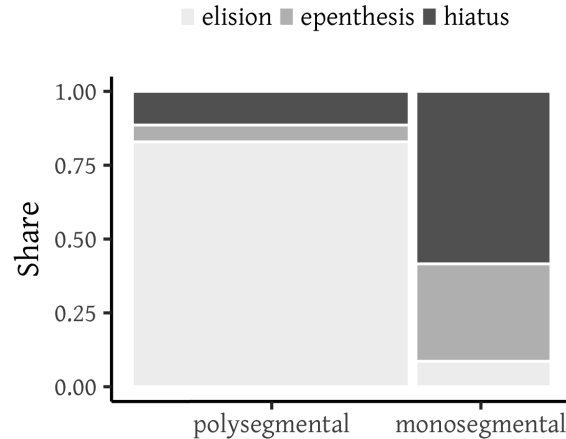

Figure 1: Share of the three realizations across all speakers by Suffix Length.

Participants provided variable realizations in 8% of cases, which we turn to next. Figure 2 shows that the second variant is most commonly hiatus when the first variant is elision. Only 8 tokens in total have elision and epenthesis as the second variant.
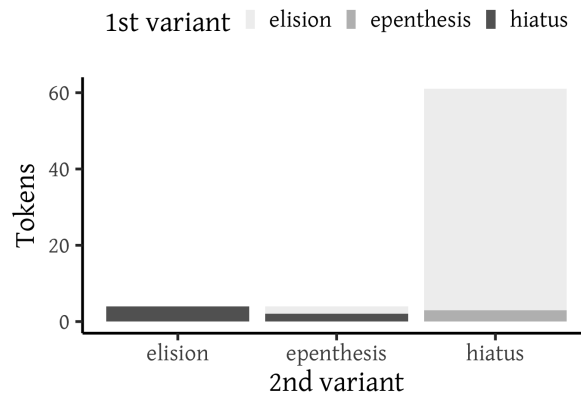

Figure 2: Variable realizations across all speakers.

3.2.2 *Lexical Category*   There were no substantive difference across different lexical categories, which suggest that the generalizations in the native words are productively extended to loanwords and nonce words. In lieu of a graphic representation, we present the inferential statistics below.

3.2.3 *Statistical analysis*   To test whether hiatus patterns across conditions differ significantly, the data was fit into a mixed-effects logistic regression model the *glmer* function from the *lme4* package (Bates et al.

2015) in R (R Core Team 2013). The response variable was binary coded: elision versus other realizations (hiatus or epenthesis). Both predictor variables were simple coded with polysegmental (for Suffix Lenth) and native (for Lexical Category) as reference levels. We first ran the model with the interactions between these two variables, but the model did not converge; hence, the interactions were not included in the model we report here. Finally, we also included random intercepts for item and participant, and random slopes for each fixed factor allowed to vary by participant. This model is reported in Table 3.

	β	SE(β)	z	p	
(Intercept)	0.48	0.57	0.84	.401	
loans vs. native	0.47	1.19	0.40	.691	
nonce vs. native	0.71	1.21	0.58	.559	
monosegmental vs. polysegmental	24.99	1.53	16.29	$<.001$	***

Table 3: Production-data regression model: elision is significantly less common with monosegmental suffixes than hiatus or epenthesis.

We can see that the Lexical Category was not a significant predictor, which means that generalizations observed in native words did not differ from the ones in loanwords and nonce words. Suffix Length, however, was significant: there was less elision with monosegmental suffixes when compared to polysegmental suffixes.

3.3 *Interim summary*   In this section we showed that hiatus is variable in Persian. Among the variants, V_2 elision is common, while V_1 elision is unattested. We also found that the length of the suffix has an effect on the realization of the variants.

Since the elicitation-based production experiment was exploratory, the variation across different conditions was not tightly controlled for. For instance, while we considered all suffixes and a broad range of roots, the vowels appearing in hiatus contexts were not controlled for. There were lexical gaps and the insertion of the two glides was difficult to determine beyond doubt. Finally, our production data is limited to only six participants. We designed the perception experiment presented in the following section with these concerns in mind.

4   Perception

In the perception experiment, we further tackle the effect of suffix length on hiatus variation. Unlike the production experiment, the participants will provide acceptability judgments for all three main variants (hiatus, elision, epenthesis), with all the variables tightly controlled for.

4.1 *Methods*

4.1.1 *Stimuli*   We compiled a set of 30 CVC(C)V nonce roots, half of which were *a*-final and the other half were *e*-final. These vowels were chosen because they are licit word-final vowels in Persian and do not condition glide epenthesis, which proved problematic in the production experiment. Our main factor of interest was suffix length. We chose 3 polysegmental and 3 monosegmental suffixes, shown in Table 4.

Polysegmental		*Monosegmental*	
-emun	'possessive 1P.PL'	-i	'copula 2P.SG'
-eʃ	'possessive 3P.SG'	-e	'definite marker'
-æm	'copula 1P.SG'	-o	'conjunction'

Table 4: The suffixes used in the perception experiment.

The combination of the suffixes with the nonce roots resulted in polymorphemic words. Bare roots as well as suffixed words were pronounced by a native speaker. The suffixed words were recorded under three conditions: with elision, hiatus, and [ʔ] epenthesis.

4.1.2 *Procedure*   The perception experiment was conducted online. The participants were asked to judge acceptability of the paradigms consisting of a bare nonce root and its suffixed form. The participants saw two frame sentences and played the corresponding paradigm (Figure 3). After they had heard the recordings, they were asked to judge the paradigms as either acceptable or unacceptable. At each trial, the suffixed word was pronounced under one of the three conditions (elision, epenthesis or hiatus), which were randomized across trials. The participants could play the recordings as many times as they wanted.

> 1. I bought 3 _____ for dinner yesterday.
> 2. I ate 2 of the _____ today.
>
> ▶ ● —— 0:00 / 0:01 ◀ঃ) ——●
>
> I accept = Y
> I don't accept = N

> ۱. دیروز ۳ تا _____ برای شام خریدم.
> ۲. امروز ۲ تا از _____ رو خوردم.
>
> ▶ ● —— 0:00 / 0:01 ◀ঃ) ——●
>
> Y = می‌پسندم
> N = نمی‌پسندم

Figure 3: An example stimulus (right) and English translation (left) in the perception task.

To make the task easier for the participants, the participants heard all paradigms with the same suffix in one block. Each root appeared only once, and each suffix appeared 5 times. Each pair of words appeared under three randomized conditions (elision, epenthesis, hiatus), for a total of 90 items per participant.

4.1.3 *Participants*   In total, the survey was completed by 54 participants (28 female and 26 male, with a mean age of 29). They all reported to be native speakers of Spoken Persian.

4.2 *Results*   As in the production experiment, our primary variable of interest was Suffix Length. Figure 4 shows the mean acceptability rates by the length of the suffix, separated for the three different conditions. One data point in this graph is the mean acceptability rate across all speakers for one paradigm (bare root and its suffixed form). When comparing acceptability rates across different conditions, we see that the paradigms with elision in polysegmental suffixes have the highest values. Conversely, elision is the least acceptable with monosegmental suffixes. Finally, epenthesis is always less acceptable than hiatus.
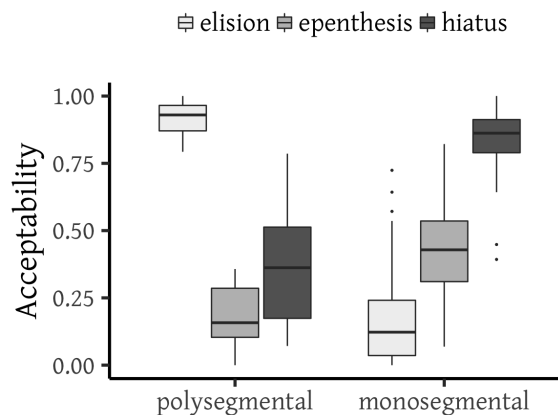

Figure 4: Perception experiment results: mean acceptability rates of hiatus variants across all speakers by suffix length.

The results demonstrate that hiatus resolution is sensitive to suffix length, with elision being the most acceptable with polysegmental suffixes but the least acceptable with monosegmental suffixes.

Figure 5 presents the results by suffix. While the polysegmental suffixes appear to be very similar, the

suffix *-e* stands out among the monosegmental suffixes, as it has a higher than expected acceptability of elision. This opens the question how much individual suffixes can vary (beyond the effect of their length), which should be addressed in subsequent work.
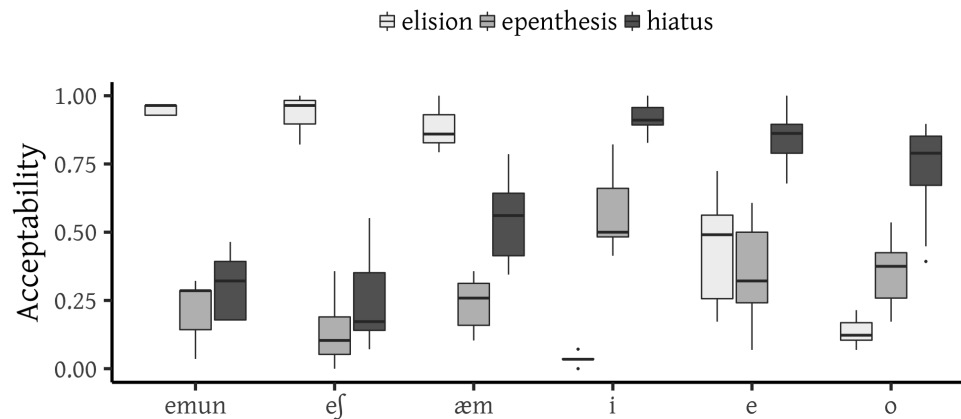

Figure 5: Perception experiment results: mean acceptability rates of hiatus variants across all speakers by suffix.

4.2.1 *Statistical analysis* We fitted acceptability into another mixed-effects logistic regression model, using the same program and package as in the production experiment. The first set of predictors in our model was the variant, which was Helmert coded as two binary predictors: *elision versus other*, with values of +2/3 for elision and −1/3 for hiatus or epenthesis, and *hiatus versus epenthesis*, with values of +1/2 for hiatus and −1/2 for epenthesis, and 0 for elision. We also considered the length of the suffix (polysegmental, monosegmental), with polysegmental as the reference level. We included the interactions between the predictor variables as well as random intercepts for participant and item, with slopes for variants and suffix length allowed to vary by participant. This accounted for variability across participants and items. We report the results of the model in Table 5.

	β	SE(β)	z	p	
(Intercept)	−0.09	0.13	−0.64	.520	
elision vs. other	0.80	0.27	2.95	.003	**
hiatus vs. epenthesis	2.05	0.29	7.01	<.001	***
monosegmental vs. polysegmental	−0.16	0.19	−0.80	.422	
elision : monosegmental	−7.41	0.41	−18.24	<.001	***
hiatus : monosegmental	0.94	0.41	2.28	.023	*

Table 5: Perception-data regression model (54 participants).

The model shows that elision is statistically significantly more acceptable than hiatus and epenthesis combined, and further that hiatus is more acceptable than epenthesis. Suffix length has no overall effect on acceptability. The key result is the interactions. In particular, elision is less acceptable with monosegmental suffixes, which confirms the trend seen in Figure 4. Conversely, hiatus is also more acceptable with monosegemental suffixes.

The statistical analysis confirms what we have seen so far: elision is the preferred hiatus resolution strategy in Persian, except with monosegemental suffixes, where hiatus is the most acceptable.

4.3 *Constraint-based analysis* Both experiments confirm that hiatus resolution strategies in Spoken Persian are dependent on suffix length. In particular, elision is the default strategy, except with monosegmental suffixes. This makes sense, since deleting the suffix vowel would delete the entire suffix, neutralizing

the paradigmatic contrast. In this section, we model the results of the perception experiment in a Maximum Entropy (MaxEnt) grammar.

MaxEnt is an OT grammar with weighed constraints and probabilistic outputs. The results of the perception experiment were fed to a MaxEnt learner (Goldwater & Johnson 2003; Hayes & Wilson 2008). We included two faithfulness constraints—DEP and MAX (McCarthy & Prince 1995)—as well as two markedness constraints—*HIATUS and REALIZEMORPHEME. The former is violated by a sequence of two vowels and as such drives epenthesis or elision (Casali 1998). We use REALIZEMORPHEME (2) to model the asymmetry between polysegmental and monosegmental suffixes; in particular, elision of the entire suffix violates REALIZEMORPHEME (Kurisu 2001).

(2)   REALIZEMORPHEME
        Morphemes must have output realizations.

REALIZEMORPHEME was assigned the highest weight, followed by DEP and *HIATUS. With polysegmental suffixes, the elision candidate (3-a) is preferred over the other two, while with monosegmental suffixes, hiatus (4-c) is the most common. Regardless of the suffix length, the probability of hiatus (c) is estimated at rates 1.8-times higher than the epenthesis (b). This follows directly from the violation profile of these candidates which are identical in (3) and (4). This prediction closely matches the perception and production data, suggesting that the proposed constraints are adequate.

(3)   Polysegmental suffix

/hutʃɑ-emun/	REALIZEMORPHEME $w = 2.2$	DEP $w = 1.5$	*HIATUS $w = 0.9$	MAX $w = 0.0$	$\mathcal{H}$	p
a. hutʃɑmun				-1	-0.0	.61
b. hutʃɑʔemun		-1			-1.5	.14
c. hutʃɑemun			-1		-0.9	.25

(4)   Monosegmental suffix

/hutʃɑ-e/	REALIZEMORPHEME $w = 2.2$	DEP $w = 1.5$	*HIATUS $w = 0.9$	MAX $w = 0.0$	$\mathcal{H}$	p
a. hutʃɑ	-1			-1	-2.2	.14
b. hutʃɑʔe		-1			-1.5	.31
c. hutʃɑe			-1		-0.9	.55

4.4 *Interim summary*   The perception experiment had greater control for condition and confirmed the results of the elicitations. The results reveal that hiatus variants are dependent on suffix length. In particular, V_2 elision was the most acceptable variant with polysegmental suffixes, but the least accetable variant with monosegmental suffixes. This asymmetry can be captured in probabilistic constraint-based grammars, such as MaxEnt.

5   Conclusions

In this paper, we report novel data on variable hiatus in Persian. This variation mainly depends on the length of the suffix: V_2 elision is preferred with polysegmental suffixes, but dispreferred with monosegmental suffixes. This asymmetry is grounded in the fact that deleting the entire suffix would neutralize the paradigmatic contrast. We also found that Persian allows hiatus particularly with monosegmental suffixes. To the best of our knowledge, this is the first experimental study of hiatus to date.

The reported data contribute to the existing literature on hiatus. First, Persian shows that the hiatus resolutions typically found across languages can be observed variably in a single language. Second, hiatus is marked in Persian, but can surface under specific situations, such as with monosegmental suffixes. Third, Persian displays a cross-linguistically rare type of elision of the second vowel, which is productively extended from real words to loanwords and nonce words.

References

Ariyaee, Koorosh (2019). Loanword adaption in persian: A core-periphery model approach. *Toronto Working Papers in Linguistics: Proceedings of the Montréal-Ottawa-Toronto Workshop in Phonology/Phonetics (MOT)* 41:1.

Ariyaee, Koorosh (to appear). The need for indexed markedness constraints: evidence from spoken Persian. *Proceedings of the 2019 annual conference of the Canadian Linguistic Association*.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1, 1–48.

Becker, Michael & Peter Jurgec (2020). Positional faithfulness drives laxness alternations in Slovenian. *Phonology* 37:3, 335–366.

Becker, Michael, Nihan Ketrez & Andrew Nevins (2011). The surfeit of the stimulous: Analytic biases filter lexical statistics in Turkish devoicing neutralization. *Language* 87, 84–125.

Berko, Jean (1958). The child's learning of English morphology. *Word* 14, 150–177.

Casali, Roderic F. (1997). Vowel elision in hiatus contexts: Which vowel goes? *Language* 73, 493–533.

Casali, Roderic F. (1998). *Resolving hiatus*. Garland, New York, NY.

Casali, Roderic F. (2011). Hiatus resolution. van Oostendorp, Marc, Colin J. Ewen, Elizabeth Hume & Keren D. Rice (eds.), *The Blackwell Companion to Phonology*, Blackwell, Malden, MA, 1434–1460.

Dehghan, Masoud & Aliyeh Kord Zafaranlu Kambuziya (2012). A short analysis of insertion in Persian. *Theory and Practice in Language Studies* 2:1, 27–50.

Ernestus, Mirjam & Harald Baayen (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79:1, 5–38.

Estaji, Azam, Mojtaba Namvar Fargi & Sarira Keramati Yazdi (2010). Tahlile akostike hamkhane ensedadiye chaknaee va barresiye emkane vojude do vakeye payapey dar do hejaye motevali dar goftare sari' va peyvaste dar zabane Farsi (An acoustic analysis of the glottal stop consonant and the investigation of hiatus in adjacent syllables in fast speech in Persian). *Jostarhaye Zabani (Language Related Research)* 4, 27–50.

Garrido, Marisol (2013). Hiatus resolution in spanish: Motivating forces, constraining factors, and research methods. *Linguistics and Language Compass* 7:6, 339–350.

Ghatreh, Fariba, Maral Asiaee & Saeed Rahandaz (2020). Negahi sarfi-avaee be elteghaye vake'ee dar Farsiye goftari (A morphophonetic approach to hiatus in Spoken Persian language). *Zabanpazhuhi (Journal of Language Research)* 12:35, 297–318.

Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. *Proceedings of the Workshop on Variation within Optimality Theory*, Stockholm University, Stockholm.

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

Itô, Junko & Armin Mester (1995). Japanese phonology. Goldsmith, John A. (ed.), *The Handbook of Phonological Theory*, Blackwell, Cambridge, MA, 817–838.

Jam, Bashir (2015). Rahkarhaye bartaraf kardane elteghaye vakeha dar zabane Farsi (Hiatus resolution strategies in Persian). *Zabanshenasi va Gooyeshhaye Khorasan (Linguistics & Khorasan Dialects)* 7:12, 79–100.

Jurgec, Peter (2010). Disjunctive lexical stratification. *Linguistic Inquiry* 41:1, 149–161.

Jurgec, Peter & Jessamyn Schertz (2020). Postalveolar co-occurrence restrictions in Slovenian. *Natural Language & Linguistic Theory* 38:2, 499–537.

Kager, René & Joe Pater (2012). Phonotactics as phonology: Knowledge of a complex restriction in Dutch. *Phonology* 29, 81–111.

Kalbasi, Iran (2001). Farsiye goftari va neveshtari (Spoken and written Persian). *Farhang (Culture)* 37–38, 49–68.

Kang, Yoonjung (2011). Loanword phonology. van Oostendorp, Marc, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, Blackwell, Malden, MA, 2258–2282.

Kurisu, Kazutaka (2001). *The Phonology of Morpheme Realization*. Ph.D. thesis, University of California, Santa Cruz, Santa Cruz. Available on Rutgers Optimality Archive, ROA 410, http://roa.rutgers.edu.

Majidi, Mohammad Reza & Elmar Ternes (1999). Persian (Farsi). Arvaniti, Amalia (ed.), *Handbook of the International Phonetic Association*, Cambridge University Press, Cambridge, 124–125.

McCarthy, John J. & Alan Prince (1995). Faithfulness and reduplicative identity. Beckman, Jill N., Laura Walsh & Suzanne Urbanczyk (eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, GLSA, University of Massachusetts, Amherst, 249–384. Available on Rutgers Optimality Archive, ROA 60, http://roa.rutgers.edu.

Modaressi, Yahya (1978). *A sociolinguistic investigation of modern Persian*. Ph.D. thesis, University of Kansas.

Morony, Michael (1986). Arabic II. The Arab conquest of Iran. *Encyclopædia Iranica* 1, 203–10.

Navabzadeh Shafi'i, Sepideh (2014). Barresiye taghirate ma'naee va karbordiye vam-vazhehaye zabane Faranse dar Farsi (An investigation of changes in meaning and application of French loanwords in Persian). *Elme Zaban (Language Science)* 2:3, 107–128.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Sadeghi, Ali Ashraf (1986). Elteghaye mosavetha va mas'aleye samethaye miyanji (Vowel adjacency and the issue of epenthetic consonants). *Zabanshenasi (Linguistics)* 6, 3–22.

Shaghaghi, Vida (2000). Barresiye elteghaye mosavetha dar zabane Farsi; takvazhe miyanji ya vaje miyanji? (An investigation of the vowel adjacency in Persian: epenthetic morpheme or epenthetic phoneme?). *Matn Pazhuhiye Adabi (Literary Text Research)* 9, 1–14.

Windfuhr, Gernot L. & John R. Perry (2009). Persian and Tajik. Windfuhr, Gernot (ed.), *The Iranian Languages*, Routledge, New York, NY, 416–544.

Yazarlou, Samaneh (2014). Glottal stop in hiatus: An acoustic investigation in Persian. *Procedia Social and Behavioral Sciences* 136, 12–20.

Yip, Moira (1993). Cantonese loanword phonology and Optimality Theory. *Journal of East Asian linguistics* 2, 261–291.

Zomorodian, Reza (1994). *Farhange vazhehaye dakhile orupaee dar Farsi be hamrahe risheye har vazhe (Dictionary of European words borrowed in Persian with the root of each word)*. Astan Qods Razavi, Khorasan, Iran.

Zuraw, Kie (2006). Using the web as a phonological corpus: a case study from Tagalog. *EACL-2006: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics/Proceedings of the 2nd International Workshop on Web As Corpus*, Trento, 59–66.