

# Persian elides the second vowel

Koorosh Ariyae & Peter Jurgec

University of Toronto

ACL/CLA ★ May 31, 2020

# Highlights

- In this talk, we look at variable hiatus in Spoken Persian. Most commonly, hiatus is resolved by elision of the second vowel.
- Our production experiment reveals that variation is restricted:
  - elision of first vowel, which is cross-linguistically common (Casali 1997) is never attested
  - elision of the second vowel is rare with monosegmental suffixes
- The perception experiment confirms that elision of the second vowel is predominant in polysegmental suffixes, but rare with monosegmental suffixes.
- The preference of hiatus over epenthesis remains constant regardless of suffix length.
- This study contributes to the discussion of variable phonological processes.

# Hiatus

- Vowel hiatus is a sequence of adjacent vowels (Casali 1997, 1998, 2011):
  - koana 'space' (Hawaiian)
- Many languages restrict hiatus or outright ban it:
  - V<sub>1</sub> elision: /bu ata/ → [bata] 'pour ground pepper' (Yoruba)
  - V<sub>2</sub> elision: /bamb o-awa/ → [bamb owa] 'this man' (Chichewa)
  - epenthesis: /di-ubah/ → [diʔubah] 'to change (pass)' (Malay)

# Spoken Persian

- In this paper, we examine hiatus in Spoken Persian.
- Hiatus in Spoken Persian appears to be variable. When two underlying vowels appear at the root-suffix boundary /V-V/, the surface realizations vary between:
  - VV hiatus
  - V~~V~~ V<sub>2</sub> elision
  - V?V epenthesis
- This variation does not seem to be random, but is instead related to the length of the suffix:

	'our'	'his/her'	'my'	'the'	
dæftær	dæftær-emun	dæftær-ef	dæftær-æm	dæftær-e	'office'
baba	baba-mun	baba-ʃ	baba-m	??*/baba	'dad'
	baba-ʔemun	baba-ʔef	baba-ʔæm	baba-ʔe	
	baba-emun	baba-ef	baba-æm	baba-e	
	*bab-emun	*bab-ef	*bab-æm	*bab-e	

# Existing accounts

	'our'	'his/her'	'my'	'the'	
dæftær	dæftær-emun	dæftær-ef	dæftær-æm	dæftær-e	'office'
baba	baba-mun	baba-ʃ	baba-m	??*/baba	'dad'
	baba-ʔemun	baba-ʔef	baba-ʔæm	baba-ʔe	
	baba-emun	baba-ef	baba-æm	baba-e	
	*bab-emun	*bab-ef	*bab-æm	*bab-e	

- The variability is mirrored in the existing literature on Persian hiatus:
  - Sadeghi (1986), Shaghghi (2000), Dehghan & Kord (2012) address epenthesis
  - Jam (2015) studies elision
  - Estaji et al. (2010), Yazarlou (2014) suggest hiatus is retained

# Typological outlook

	'our'	'his/her'	'my'	'the'	
dæftær	dæftær-emun	dæftær-ef	dæftær-æm	dæftær-e	'office'
baba	baba-mun	baba-ʃ	baba-m	??*/* baba	'dad'
	baba-ʔemun	baba-ʔef	baba-ʔæm	baba-ʔe	
	baba-emun	baba-ef	baba-æm	baba-e	
	*bab-emun	*bab-ef	*bab-æm	*bab-e	

- Moreover, the Persian hiatus pattern is cross-linguistically rare:
  - Casali (1997):  $V_1$  elision is much more common than  $V_2$  elision; only 2 other known languages have  $V_2$  while not also having  $V_1$  elision.
  - Garrido (2013): Very few reported languages exhibit variation in which hiatus is allowed, but also variably resolved in multiple ways.

# This study

We conduct two experiments to investigate how hiatus varies in Spoken Persian:

## ① production

- small elicitation-based experiment
- the principal aim is gauge the variation within and across speakers

## ② perception

- larger controlled experiment
- designed to specifically investigate the relationship between the three principal variants and their dependence on suffix length

# Key issues

- There is no corpus of Spoken Persian which would allow us to check the distribution or variation in hiatus patterns.
- We conducted a small elicitation-based production experiment which would allow us to gain insight into several key issues:
  - ✓ What variants are possible and what is the relationship between them?
  - ✓ Does Persian allow both types of elision ( $V_1$  and  $V_2$ )?
    - How does variation depend on vowel quality?
  - ✓ Beyond vowel combinations, do all suffixes behave uniformly?
    - Do speakers differ in their distributions?

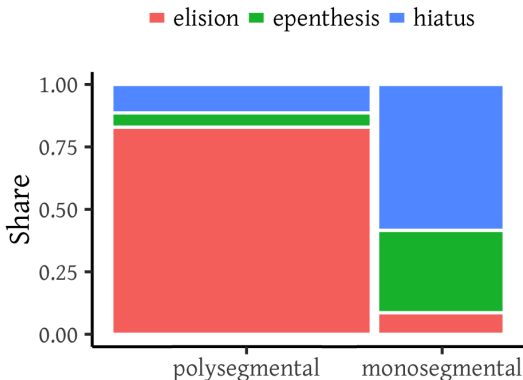


# Methods

- Stimuli: 108 roots, 17 V-initial suffixes:
  - $V_1$ : {i, e, a, o, u}
  - $V_2$ : {i, e, æ, a, o}
  - Suffix length: -V, -VC, -VCVC
  - Root stratum: native, loanwords, nonce words
- Word-formation production experiment:
  - Familiarization stage: researcher provided C-final root + V-initial suffix
  - Main task: participant derived V-final roots + (the same) V-initial suffix
- 7 participants completed the experiment (mean age = 30)

## V<sub>2</sub> elision, epenthesis and hiatus depend on suffix length

- V<sub>2</sub> elision, epenthesis, and hiatus are most frequent variants.
- Participants provided about 8% variable realizations.
- The productions depend on suffix length (1,202 tokens):



# Other findings

- $V_2$  elision is the most frequent realization, but  $V_1$  elision is unattested.
- ... but  $V_2$  elision is extremely rare with monosegmental suffixes.
- $V_1$  determines choice of the epenthetic segment:
  - i : epenthetic j
  - u : epenthetic w
  - e a o : epenthetic ?

# Discussion

- Persian is rare:
  - hiatus is variable
  - $V_2$  elision is common, while  $V_1$  elision is unattested
- The production experiment did not tightly control for various variables:
  - a broad range of roots and suffixes was considered
  - lexical gaps: no *u*-initial suffixes, no polysegmental *a*-initial suffixes
- While the presence of [ʔ] was clear, this was not so for [j] and [w].
- Because all results were phonetically examined, the productions were limited to a small number of participants.
- To better control for these variables, we conducted a perception experiment.

# Why perception?

We ask whether the generalizations observed in the production experiment are extended to nonce words:

- Is  $V_2$  elision productive, given its cross-linguistic rarity?
- Would suffix length influence the variation?
- Do speakers consistently distinguish between hiatus and epenthesis?
- What happens if participants can judge grammaticality of multiple realizations of the same root–suffix combination?

# Methods

- Stimuli:
  - 30 V-final nonce roots
  - 3 monosegmental (-V) suffixes, 3 polysegmental suffixes
- Procedure:
  - Each participant judged acceptability of 30 nonce paradigms.
  - Each paradigm consisted of a bare and derived root.
  - Each of the paradigms appeared under three conditions (elision, epenthesis, hiatus; randomized), for a total of 90 items per participant.
- 54 participants (mean age = 29) completed the experiment.

# Procedure

## Sample experimental item

1. I bought 3 \_\_\_\_\_ for dinner yesterday.
2. I ate 2 of the \_\_\_\_\_ today.



I accept = Y  
I don't accept = N

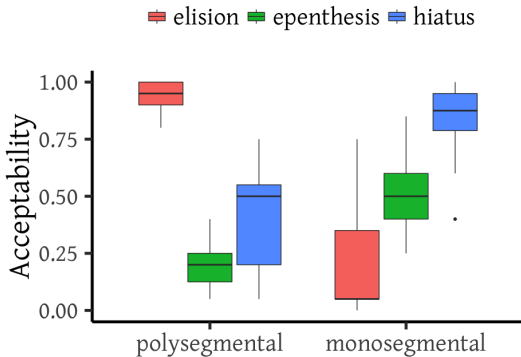
۱. دیروز ۳ تا \_\_\_\_\_ برای شام خریدم.
۲. امروز ۲ تا از \_\_\_\_\_ رو خوردم.



Y = می‌پسندم  
N = نمی‌پسندم

# Results by variant and suffix length

- $V_2$  elision is the most acceptable variant with longer suffixes.
- Hiatus is more acceptable than epenthesis across conditions.





# Inferential statistics

We fit the acceptability in a mixed-effects logistic regression model.

- Fixed effects:
  - Variant (Helmert coded: elision vs. other, hiatus vs. epenthesis)
  - SuffixLength (simple coded: monosegmental vs. polysegmental)
- We also included:
  - interactions between Variant and Suffix-Length
  - random by-participant and by-item intercepts
  - random by-participant slopes

# Model results

	$\beta$	SE( $\beta$ )	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	*

## Findings:

- 1 Elision has the highest acceptability rates, followed by hiatus.
- 2 Elision is less acceptable with monosegmental suffixes.
- 3 Hiatus is more acceptable with monosegmental suffixes.

# Discussion

- The results of the production and perception experiments are consistent.
- Even with greater control for condition, elision is the most common variant.
- Elision is affected by suffix length.
- Hiatus is preferred over epenthesis at a rate about 2 : 1.

# Maximum Entropy grammar

- To model variation, we fed the results of the perception experiment to a Maximum Entropy learner (Goldwater & Johnson, 2003; Hayes & Wilson, 2008).
- MaxEnt is a constraint-based grammar:
  - constraints are weighted, not categorical (Harmonic Grammar)
  - outputs are probabilistic
- We considered four key constraints:
  - REALIZEMORPHEME (Kurusu 2001)  
Morphemes must have output realizations.
  - \*HIATUS (Casali 1998)  
Assign a violation mark for every pair adjacent vowels.
  - DEP  
No epenthesis.
  - MAX  
No deletion.

# MaxEnt grammar of Persian hiatus

- Elision is preferred over hiatus and epenthesis:

/hutʃa-emun/	REALMORPH $w = 2.2$	DEP $w = 1.5$	*HIATUS $w = 0.9$	MAX $w = 0.0$	$H$	$p$
a. hutʃamun				-1	-0.0	.61
b. hutʃaʔemun		-1			-1.5	.14
c. hutʃaemun			-1		-0.9	.25

- ... but hiatus is the most common variant with V-suffixes:

/hutʃa-e/	REALMORPH $w = 2.2$	DEP $w = 1.5$	*HIATUS $w = 0.9$	MAX $w = 0.0$	$H$	$p$
a. hutʃa	-1			-1	-2.2	.14
b. hutʃaʔe		-1			-1.5	.31
c. hutʃae			-1		-0.9	.55

# Conclusions

We found that hiatus in Persian is variable:

- Variation in Persian is not random, but systematic.
- $V_2$  elision is the most common resolution with polysegmental suffixes, but rare with monosegmental suffixes.
- Hiatus and epenthesis are both possible, but the former is more frequent than the latter, regardless of suffix length.

# Conclusions

Bigger picture:

- This is the first study showing the productivity of  $V_2$  elision experimentally.
- This is the first experimental study showing the variable hiatus resolutions typically found across languages are observed variably in a single language.
- The results can be modelled using probabilistic grammars such as MaxEnt.

# Thanks to . . .

- our participants
- This work is supported by the University of Toronto and SSHRC Institutional Grant and Graduate Research Grant.



Social Sciences and Humanities  
Research Council of Canada

Conseil de recherches en  
sciences humaines du Canada

Canada



# Persian elides the second vowel

Koorosh Ariyae & Peter Jurgec

University of Toronto

ACL/CLA ★ May 31, 2020

# References I

- Casali, Roderic F. (1997). Vowel elision in hiatus contexts: Which vowel goes? *Language* **73**. 493–533.
- Casali, Roderic F. (1998). *Resolving hiatus*. New York: Garland.
- Casali, Roderic F. (2011). Hiatus resolution. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren D. Rice (eds.) *The Blackwell Companion to Phonology*, Malden, MA: Blackwell. 1434–1460.
- Dehghan, Masoud & Aliyeh e Zafaranlu Kambuziya Kord (2012). A short analysis of insertion in Persian. *Theory and Practice in Language Studies* **2**. 27–50.
- Estaji, Azam, Mojtaba Namvar Fargi & Sarira Keramati Yazdi (2010). An acoustic analysis of the glottal stop consonant and the investigation of hiatus in adjacent syllables in fast speech in Persian. *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**. 339–350.

## References II

- Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Workshop on Variation within Optimality Theory*, Stockholm: Stockholm University.
- Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* **39**. 379–440.
- Jam, Bashir (2015). Hiatus resolution strategies in Persian. *The Journal of Linguistics and Dialects of Khorasan* **12**. 79–100.
- Kurusu, Kazutaka (2001). *The Phonology of Morpheme Realization*. Ph.D. dissertation, University of California, Santa Cruz, Santa Cruz. Available on Rutgers Optimality Archive, ROA 410, <http://roa.rutgers.edu>.
- Sadeghi, Ali Ashraf (1986). Vowel adjacency and the issue of epenthetic consonants. *Linguistics* **6**. 3–22.
- Shaghaghi, Vida (2000). An investigation of the vowel adjacency in Persian. *Language and Literature* **9**. 1–14.
- Yazarlou, Samaneh (2014). Glottal stop in hiatus: An acoustic investigation in Persian. *Procedia Social and Behavioral Sciences* **136**. 12–20.