

Zabaan: The Descendance of Classical Persian in Urdu and Iranian Persian

Abstract

The Indo-Aryan language Urdu, spoken largely in South Asia, has been widely known as a “mixed” language possessing a Hindi grammar and highly Persianized lexicon due to contact with Persian-speaking conquerors. What are the systematic phonological differences between current Urdu and Persian productions of a shared word? This paper evaluates how modern Persian and Urdu vary on their productions of low vowels in words that derive from Classical Persian and historically contained the /a/ and /a:/ vowels. This is a replication study of Kuczkiewicz-Fraś, which suggests that due to different sound changes and nativization processes, Classical Persian’s /a/ is now /æ/ in modern Persian and /ə/ in Urdu, while the Classical Persian /a:/ is now /ɒ:/ in modern Persian, and /ɑ:/ in Urdu (153). Additionally, an impressionistic analysis of how the dorsal consonants [q], [ɣ], [x], [k] and [g] are used between the two languages is also reviewed. The data implies the collected F₁ and F₂ group averages for each vowel have a similar spatial relationship to the canonical position of their theorized current vowel, and significantly different vowel spaces among them. However, as these significant differences wash away with the inclusion of a random intercept variable, this study did not find unambiguous evidence that the Classical Persian vowel /a/ went to /æ/ in Persian and /ə/ in Urdu, not that the Classical Persian /a:/ went to /ɒ:/ in Persian and /ɑ:/ in Urdu as suggested by previous research. Within nearly all Persian speakers, [ɣ] and [q] are merged to /q/. Contrariwise, most Urdu speakers retained Classical Persian’s /ɣ/, but the explicit Classical Persian distinction between /ɣ/ and /q/ has been lost as the latter segment is rare or moved to the velar position. One speaker regularly substituted [k^h] for /x/, [g] for /ɣ/ and [k] for /q/. These replacements were all in accordance with the Rule 3, 6, and 7 of Kuczkiewicz-Fraś’s observations. These findings support existing research positing that the dorsal consonants /x/, /ɣ/, or /q/ are rarer and have prestige value associated with their use in Urdu.

0. Acknowledgments

I'd like to thank my participants for helping me with this project, as without them it would never have seen the light of day. I'm deeply appreciative of the insights they provided that went beyond the requirements of the task, and added intriguing depth to the dataset.

The basis of this project morphed with time, as I discovered my original concept had already been studied quite exhaustively by Dr. Agnieszka Kuczkiewicz-Fraś, a professor specializing in Hindi-Urdu linguistics at Jagiellonian University in Krakow, Poland. The LGCS department purchased her dictionary for my research, but she also graciously shared the full PDF of both her dictionary and her book on Perso-Arabic loanwords in Hindustani through email on request when the dictionary was delayed in shipping (and the book was not even on any American market). I'm immensely obliged by the show of generosity on her part.

I'm also thankful for the efforts of my advisors, namely Dr. Michael Diercks and Dr. Franny Brogan, for guiding me through a project that felt well beyond my competence and reassuring me when I faltered. Finally, my partner, Jorj, and close friends handled the duty of providing support, for which I am grateful.

As I write this essay, the US has an uneasy relationship with Iran following the assassination of Iranian General Qassem Soleimani on January 3rd, 2020. This is, of course, just the most recent event in a conflictuous rapport, but one that has again alarmed the population of Iranian-Americans in the US, whose citizenship status is put in a precarious position. Iranian students studying abroad in the US are being "returned" (i.e. deported) to Iran; Iranian-Americans with variable citizenship statuses have been detained and questioned at the Canadian border which the US Customs and Border Patrol justifies as due to the "current threat environment" necessitating CBP to act with "an enhanced posture at its ports of entry to

safeguard our national security and protect the American people...” It doesn’t take a critical discourse analyst to surmise that protecting the American people from Iranian-Americans—who had popped over to Canada for an Iranian pop concert—signals that Iranian-Americans, regardless of citizenship status, are not wholly and undisputedly American thanks to the pesky hyphen. My project was only possible due to the sizable Iranian diaspora in Southern California, the members of which are impacted by this legislatively-enforced tension.

Simultaneously, in India, the Citizenship Amendment Act (CAA) and National Register of Citizens (NRC) are two outputs of the right-wing BJP party, working together to slowly and systematically remove Muslims from the subcontinent. The CAA grants citizenship to all fleeing religious persecution in neighboring countries save for Muslims (thus denying them legal entry into India on the grounds of persecution), while the NRC represents an attempt to document every residents’ legal presence in India. For the latter, it is not enough to have a passport, national ID card, or any other sort of conventional “proof”; rather, each individual must prove their ancestry through historical documents to confirm the legality of citizenship. Many in India do not have these documents- my own grandparents do not have birth certificates, as was the norm in that time, and to retrieve the relevant documents from the government can take well over several months. Already, Muslims who have lived in India for generations are the first to be targeted on the basis of their lack of legal historical documentation. The BJP party is heavily endorsed by Hindutva subscribers, a type of Hindu nationalism claiming that India’s true history is Hindu; that Muslim conquerors were akin to British colonizers with respect to the damage inflicted through foreign rule; that Indian Muslims (the second largest religious group following Hindus) degrade the country’s prosperity and should be purged. When students of Jamia Millia Islamia—an Islamic, public university—in Delhi protested CAA peacefully, they were brutalized

by the police. The police further destroyed the campus and a local mosque. The phenomenon of Persian loanwords in Hindi-Urdu (among many other South Asian languages!) was dependent on the foreign Muslim rule that Hindutva so despises. As the news of what was happening at Jamia Millia Islamia was covered, interviewers demanded repeatedly to hear students' names whenever they expressed contempt at the police; if their name was decidedly Persian or Arabic in origin, the appropriate assumptions could be made about their allegiances, and their stories of unjust violence could be dismissed, or worse, deemed deserved. These are both but contemporaneous events that relate the overall history my project draws from and its ability to be run. I wanted to acknowledge this openly, as it is sometimes easy to isolate a linguistic phenomenon from its material relevance in the present day (and all the more so when the phenomenon is a historical one). I hope that this context is not lost on any readers of this paper.

1. Introduction

The Indo-Aryan language Urdu has been widely known as a “mixed” language, born out of military and trading encounters between an array of Muslim outsiders and Hindi-speaking natives across the subcontinent ranging from 711 AD to 1857 AD (Abbas 4). Currently the national language of Pakistan and widely spoken in North India (Khan 277), Urdu contains a significant number of Persian loanwords, although estimates as to the proportion of borrowing are speculative, inconsistent across both different sources and types of speakers. Systemic sound changes for these loans (known as the nativization process) that took place during or after borrowing have, thus far, only been preliminarily outlined.

Concerning a range of Persian loanwords in Urdu, what are the systematic phonological differences between current Urdu and Persian productions of the same word? This paper evaluates how modern Persian and Urdu vary on their productions of low vowels in words that derive from Classical Persian and historically contained the /a/ and /a:/ vowels. This is a replication study of Kuczkiewicz-Fraś which suggests that due to different sound changes and nativization processes, Classical Persian’s /a/ is now /æ/ in modern Persian and /ə/ in Urdu, while the Classical Persian /a:/ is now /ɒ:/ in modern Persian, and /ɑ:/ in Urdu (153). Additionally, an impressionistic analysis of how the dorsal consonants [q], [ɣ], [x], [k] and [g] are used between the two languages will also be touched on.

The division between Urdu and Hindi is contentious and often reduced to discussions of the former as a Persianized register, and the latter as a Sanskritised register (Everaert 14). Simultaneously, both languages are mutually intelligible at the colloquial level (known together as *Hindustani*), and slight differences in vocabulary etymology are usually the result of preference, not from an inability to produce or comprehend words from the alternative

etymology (18). Therefore, the usage of Persian loanwords in Urdu have a variety of motivating forces; some are fully integrated into the normal, everyday lexicon, while others index a Muslim religiosity, a national or political sense, a higher level of education (and relatedly a poetic or romantic quality) (Billah 104; Mir 317). Considering the conspicuity of such loanword usage, and the political divides among which this usage aligns, it initially seems curious that these patterns have not been extensively elaborated on.

However, there are several justifications for the conspicuous lack of research on the topic (Bashir 98). Loanword phonology primarily deals with real-time nativization, whereas many Persian loans in Urdu fully assimilated some centuries ago. Both languages have undergone sound change since this previous time of exchange, and any evaluation of the current phonological manifestations of the same word in Persian and Urdu will unintentionally take these respective sound changes as part of the loanword phonology. That is, it is currently impossible to ascertain what is truly the nativization process as the sounds and phonology of Mughal Persian (and preceding Persians) and the newly born Urdu are not presently accessible to researchers.

An extension of this problem deals with the wide range of contact between various Persianate kingdoms and South Asian subjects. Certain loans are older, and may have been subject to different rules, whereas others are far more recent and could have evolved differently under newer rules. This is itself related to the status of English loanwords and English code-switching with Hindi-Urdu, which introduced a new vowel, /æ/, to Hindi-Urdu phonology that was previously unavailable; hence, older Persian loans that had this vowel may have been replaced wholesale in older nativization processes, whereas modern Persian loans may have the vowel preserved. Additionally, the education, religion, class, and register of a speaker of Urdu

can change the nature of the nativization. Certain consonants that are native to Persian are variably pronounced by Urdu speakers, akin to how elite English-speaking American students may perfect and pronounce a French *R* to index a worldliness and class standing (Billah 105).

However, it is the complexity of this project that requires some basic research about phonological change in loaning, despite the many confounding factors. Until the phenomena is broadly understood, it cannot be refined in scope, detail, and accuracy. As such, my project focuses on a site of prominent difference between modern Urdu and Persian productions of the same word: the low vowels, deriving from the common base of Classical Persian's /a/ and /a:/. (Kuczkiewicz-Fraś 153). Furthermore, I will touch on the adaptable nature of shared dorsal consonants between the two languages (Kuczkiewicz-Fraś 173).

2. Literature Review

2.1 History

It is pertinent to study the historical basis of Persian language influence on South Asian languages to begin to understand the origin of borrowed loans. Muzaffar examined in detail the historical forces that shaped Mughal promotion of Persian as a courtly language. He states that Urdu (that is, a “Persianized Hindi”) as a language was not born out of Mughal contact, but first from the Delhi sultanate. This is a controversial stance, as some differentiate the Delhi sultanates’ language as a separate register dubbed Deccani—and how Deccani Urdu travelled up the north half of the subcontinent to further develop with later Mughal rule is left unmentioned.

The Mughal rule of India took place from 1530 AD to 1857 AD, led by Turko-Mongol conquerors from what is now Uzbekistan. The rulers, while not ethnically Persian, were Persianate and therefore normalized Persian traditions, customs, arts, clothing, foods, and language as the courtly standard (317). Prior to Mughal conquering, Persian had already been

established in India as the language of a Muslim elite (intelligentsia classes as well as merchants, traders, and religious leader), thereby setting a precedent for its continued promulgation. The article cites a cheeky poem regarding Indian fascination with Persian poetry: “All the Indian parrots will turn to crunching sugar with this Persian candy...” (318). The Mughal “decision” to continue with this historical precedent, then, is multi-faceted: first, there was the problem of Persia having already lent many helping hands to the Mughals when it came to their conquering in the sub-continent (319); it would reflect well on empire to respect this help by further “Persianizing” themselves despite being historically Turko-Mongol. The renowned Mughal Emperor Akbar took additional interest in fostering relationships with Iran through adoption of culture and language in North India, a tale that lines up with traditional Indian narratives on Akbar as worldly and cunningly diplomatic (320). Akbar’s courts called for Persian artists, many of whom came in droves in order to avoid the sectarian Safavid rule in Iran (321). This open invitation is regarded by the author as potentially derived from Akbar’s desire to repay the debt Mughals owed Persians for aforementioned support. However, Akbar ambitiously positioned himself in obvious competition with the Shah, complicating and straining the relationship once more (321).

The influx of highly trained/educated Persian native speakers in addition to Mughal fluencies in Persian led to Persian emerging as the primarily language of the court (324). Akbar’s subsequent lineage were remarkably poor in their Chagatai Turkic fluency (the native language of the Turko-Mongols) in comparison to Persian, although their adroitness in Hindustani was variable (324). Akbar declared Persian to be the language of administration at all levels, prompting Hindu communities to learn Persian in order to receive clerical jobs, particularly those of the clerical castes (325). Madrasas (typically Muslim sites of education, equivalent to

schoolhouses) that taught Persian were patronized by non-royal Hindus and Muslims alike, garnering a greater appreciation for Persian literature and arts by the equivalent of the middle-class (328). Muslims were subjected to higher expectations of Persian knowledge, such that Muslims literate in any language were expected to be able to read simple Persian (329).

Persian was cited as the ideal political language for South Asia due to its perceived uniformity. Persian purportedly had a standardized dialect, whereas every hundred miles the standard Indian language differed (much to the frustration of the nobles who cared). Various efforts to empower the local languages at an administrative level such as Hindustani were met with indignation (331). Justifications ranged from Persian as symbolically victorious and Muslim in ways Hindustani could never convey, or that Hindustani was simply too varied to employ. Hindustani use of Persian loanwords was similarly ridiculed, with reference to age-old prescriptivist beliefs—the nativization process for loanwords through spelling, pronunciation, or re-ordering were seen as bastardizing (341).

The Mughals eventually admitted that a working knowledge of Hindustani would benefit their leadership, and work was gathered and prepared to pass on this knowledge to the rulers (343). As they were largely literate in Persian only, the Hindustani language was written in Persian script. But only the “pure” form of Hindustani, that which was spoken at the exalted imperial camps (called *Urdū-yi mu‘allā* in Turkish) was to be learned, leading to the dubbing of this variety as *Urdu*. The author notes that in making these key decisions, the Mughals failed to recognize the largely-used local Hindustani but rather the Mughal-influenced variety of it, which rendered their task insufficient in accomplishing local ties (346). This *Urdu* was a Persianized register, supposedly for the means of Mughals raising it to their linguistic standards. The first scholar to prepare a Hindi-Persian dictionary in this same time period- *Arzu*- declined to note the

“Indianized” pronunciations of Persian words into the orthography, instead insisting on their originating forms in Classical Persian (347).

Billah provides more context for Persian exchange prior to Mughal administrative decisions. Multiple stages of the Persian language’s development have occurred, and contact with South Asian languages took place throughout these stages. These include such varieties as Avesta, Pahlavi, and modern Persian (98). These phases of Persianate influence were markedly less institutionalized than Mughal reforms, but relevant all the same for understanding the ebb and flow of Persian on the subcontinent. Beyond initial use as a literary and religious language as outlined earlier, Persian later gained ground as a courtly language as well within the Delhi Sultanate court (108). Within literary use, it had particular influence on the romantic traditions, hence its predominance in poetry. The Ghaznavid dynasty in Lahore (977-1186 AD) and the Ghorid dynasty in Delhi (879-1215 AD) furthered the courtly tradition, whereon Hindus began learning Persian to maintain social and official status, being the major language of administration, literature, and culture in these areas (103). Later dynasties, like the Khalji, Tughlaq, Lodi, and finally Mughal Empire (the range of these spanning from 1290 AD to 1857 AD) brought more Persian into the fold. These are many disparate ruling periods, and the flow of Persian “into” local vernaculars varied in rate and type. Because of the concentration of Persianate influences in the northern plains of the subcontinent, the South Asian languages that took on the most Persian loanwords are from this area, such as Bengali, Punjabi, and Gujarati. The shared register of *Hindustani* had a fair deal of borrowing, although formal or *Shudh* Hindi has lesser representation of Persian loans; conversely, formal Urdu takes on a great deal of Persian influence, more so than *Hindustani*. (104). Based on a cursory survey of common Persian loans, Billah notes that Persian loanwords tend to reside in the following semantic

categories: religion, distinctive Muslim practices, tailored clothing and garments, urban orientation, literacy, administration, honorific formations, body-organs, kitchens, fruits, flowers, and inspiration (107).

Khansir and Mozafari open the scope of Persian influence to languages other than Urdu, and concurrently provide fascinating linguistic ideology. Persian loanwords are populous among many languages throughout the subcontinent beyond Urdu. Indo-Aryan languages spoken in the Northern plains (that is, Hindi-Urdu, Bengali, Punjabi, Gujarati, Sindhi and Rajasthani) utilize Persian words for which they have no synonym themselves, suggesting that the words have become completely integrated into the language(s) (2360). The article cites Urdu as a language born out of Hindi and Persian, with some influence of Arabic (both directly borrowed or tangentially through Persian, which itself borrowed from Arabic) (2363). This is, however, strangely differentiated from the borrowings of Persian into Hindi, which the paper seems to view as a completely different language from Urdu even prior to Mughal interaction. This raises the question as to whether there are Persian loans in Hindi that do not exist in Urdu, a pattern that is opposite to what would be expected when trying to distinguish Hindi and Urdu formally.

British influences first pushed Persian as an institutionalized language (as a way to appease Muslims rulers, who were in power at the time of initial colonization) and then suppressed it (in favor of English usage with later colonization) (2364). Additionally, the paper mentions that Persian influence is tolerable to a certain extent for Indian languages, but may exceed what is “acceptable”, suggesting a kind of nationalism. Even within academia, there is still a sense that Persian influences on North Indian languages are encroaching, much like popular non-Muslim regard of the Mughal rulers themselves (2364).

Notably, these three articles as well as a variety of other sources vary widely in their label for the shared register for Hindi-Urdu. Some call it *Hindawi* (Alam 319), while others differentiate between *Hindustani* and Urdu, referring the former as a “lingua franca” and the latter as the “highly Persianized language of the elite”. The latter is synonymous with *Lashkari* (the Sanskritic word for camp, instead of the Turkish *Urdu*) (Billah 104). Still others call this register *Dehlavi* and *Rekhta* (meaning “scattered; mixed” in Persian) (Brown and Anderson, 2006). This is yet differentiated from *Deccani*, a South Indian language highly resembling Urdu in both grammar, vocabulary, and intelligibility, but solely apart by the number of loans from Marathi, Telegu, and Kannada (which are decidedly Dravidian in nature) (Prakāśam 156). The array of terms for what is essentially the same colloquial language (with natural geographic variance) is clearly political in nature, furthered by the notion that it is in the formal, academic, and/or elite registers that the etymology of words changes significantly enough to cease being mutually intelligible.

Simultaneously, Urdu is romanticized by non-speakers when in the realm of poetry (Mir 317). The popular Bollywood song *Chaiyya Chaiyya* from the movie *Dil Se...* recites the following: “*Voh yaar hai jo khushboo ki tarah, Voh jiski zabaan urdu ki tarah*”, broadly translating to “The companion who walks along like fragrance, whose melodious tongue is [as beautiful as] Urdu.” (321). Within the cited lyric, *yaar*, *khushboo*, *tarah*, and *zabaan* are almost all Persian borrowings, with a lack of clarity regarding cognates that naturally derive from being within the Indo-Iranian language family. These lines were written with the explicit purpose of being comprehensible to a Hindi-speaking audience in India, illustrating how integrated many Persian loans are in the Hindi-Urdu language.

2.2 Phonology

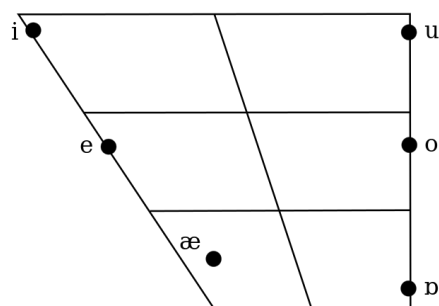
2.2.1 Inventories

As I am intending to explore the low vowel differences between modern Persian words and Urdu loanwords of the same Classical Persian (CP) origin, it is relevant to outline the languages' current vowel inventories. The same applies for the dorsal consonants. Because there are a number of different phonemic analyses for any given language, a variety were utilized to build as clear a picture as possible. Northern Urdu and Tehrani Farsi were centered as the dialects of study, as they both represent the most normative and widely-spoken registers of each language (Rees 1; Kuczkiewicz-Fraś 29).

Majidi & Ternes put forth the following analysis of Persian phonology (124), where bracketed sounds are lesser-occurring sounds that are also included in Kuczkiewicz-Fraś (132) and Rees (20). Sounds preceded by a tilde indicate those that occur in variation.

	Velar	Uvular
Plosive	k g	q [g]
Fricative	x	~χ

Persian consonants

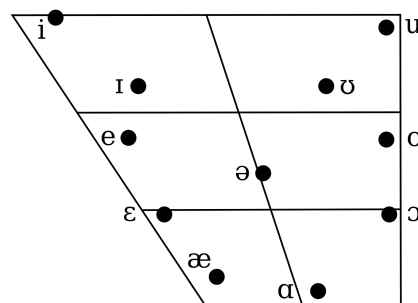


Persian vowels (© Majidi & Ternes 1999)

Kuczkiewicz-Fraś (135), Schmidt (308) and Ohala (102) suggested the following Northern Urdu phonemic inventory.

	Velar	Uvular
Plosive	k, g, k ^h g ^h	[q] [ɢ]
Fricative	[x]	[χ]

Urdu consonants



Urdu vowels (© Ohala 1994)

All Urdu vowels are also distinguished in the dimension of nasalization.

The differences between Tehrani Persian and normative Northern Urdu are summarized below:

	Persian only	Shared	Urdu only
Vowels	/ɒ, *æ/	/i, e, u, o/	/ɪ, ε, ə, ɑ, ʊ, ɔ/
Dorsal consonants	/~ ɣ/	/k, g, q/ [x, ɢ]	/k ^h , g ^h /, [χ]

*This sound is in some Urdu vowel inventories, but used exclusively for English loanwords (Ohala 102)

The asymmetries between the modern vowel inventories of Persian and Urdu set up an interesting contrast. Provided that Urdu currently has the comparatively larger vowel inventory, which Urdu vowels are now typically pronounced in the place of the exclusively Persian vowels /æ/ and /ɒ/? Is there a one-to-one correspondence between Persian and Urdu vowels, or do multiple types of Urdu vowels replace one vowel from Persian? At an initial glance, /æ/ may be regularly replaced by Urdu /ə/, /ɑ/, or /ε/ and Persian /ɒ/ will likewise either be regularly replaced by Urdu /ɑ:/ or /ɔ/, based solely on proximity alone.

Furthermore, only one Persian consonant is not present in the Urdu inventory, while there are many Urdu consonants absent from the Persian inventory. With consideration that loaning occurred under a different Persian phonology, and certain segments exclusive to Persian could

have been dropped, it also seems as though there should be very little consonantal shift when it comes to Urdu pronunciations of a Persian loanword. However, certain shared phonemes are considered lesser-occurring in Urdu and therefore are also likely unpronounced by many speakers. In this case, such rarer consonants as [x, q, ɣ] may be assimilated to the velar stop position, as Urdu does not possess any other velar or uvular sounds (save for the velar nasal).

2.2.2 Previous Comparative Work

It is critical to review not only the inventories but the historic nativization processes that illuminate patterned alterations between the languages. Kuczkiewicz-Fraś explicitly examines how Persian and Urdu manifestations of the same word differ (1). The author is primarily a scholar of Hindi-Urdu linguistics, and used her own corpus of Hindi-Urdu tokens to compare against historical pronunciations of the same word in Persian (data gathered from other existing works on Persian phonology). Analysis of the phonological changes of Persian to Urdu was multi-faceted, as the loaning process was a diachronic, rather than synchronic event (3). This was critical to the accuracy of the phonological analysis, as a synchronic account could posit that Urdu loaning reduced the Persian /æ/ to /ə/; however, as indicated by Kuczkiewicz-Fraś's investigation, many of the Persian loanwords in Hindi-Urdu were borrowed from CP, where /æ/ did not exist (124). It is far more likely, then, that Hindi-Urdu productions have preserved the vowels of CP, whereas Modern Tehrani Persian has undergone vowel shifts that altered their productions. This is in line with both Kuczkiewicz-Fraś's analysis as well as Friedrich Rosen, the once Foreign Minister of Germany (in October 1921) and later professor of Persian and Urdu at the Department of Oriental Languages at Friedrich-Wilhelms-Universität in Berlin:

The Persian of India may be looked upon as a petrification of the old classical language...
The Persian-speaking Indians, whose studies are mostly confined to the classics and

poetic exercises, have followed none of the developments of the modern language.
(Rosen 203)

Kuczkiewicz-Fraś expands on this, noting that in understanding the vowel system of Persian loanwords in Urdu, it is critical to understand the Classical Persian vocalic system: the short vowels (/i/, /u/, and /a/) and long vowels (/i:/, /e:/, /o:/, /u:/, and /a:/) (122). Modern Persian short vowels (/ɛ/, /o/ and /æ/) and long vowels (/i:/, /u:/, and /ɒ:/) derive from these earlier equivalents, although /i:/ and /e:/ converged to /i:/, while /o:/ and /u:/ converged to /u:/. Additionally, /a:/ transformed to /ɒ:/ (Rees 94).

Kuczkiewicz-Fraś additionally handily summarizes the Hindi-Urdu replacements for these older vowels, listing the original CP sound and the Urdu sound that is used in replacement or variation. The sum of these are listed below, with those not related to the vowels in question removed for brevity:

Summary 1. Vowels

1. /a/ → [ə] (154)
 - a. Or [ɪ], to preserve vocalization (156)
 - b. Or [ʊ] (158)
 - c. Or [e] in proximity to tautosyllabic /h/ or morphophonemically motivated (158)
 - d. or [ɑ:] word-finally, for morpho-phonemic reasons (153)
2. /a:/ → [ɑ:] (124)
 - a. Or [ə] (154)

Therefore, between these two paths of CP to either modern Persian or Urdu, there are several diverging changes. Take the word *kam*, *yaar*, and *zabaan* (meaning *less*, *friend*, and *tongue* in both languages). In combining these two theories of sound change, a potential mapping of the historic CP sound to modern sounds would resemble the following:

Persian		Classical Persian		Urdu
/kæm/	←	/kam/	→	/kəm/
/jɒ:r/	←	/ja:r/	→	/ja:r/
/zæbɒ:n/	←	/zabā:n/	→	/zəbā:n/

These paths are both, however, *theoretically* posited. They have yet to be supported through the collection of phonetic data that directly examines modern Persian in contrast to modern Urdu. Even with twenty of Kuczkiewicz-Fraś's observations omitted, the list of possible outcomes is dauntingly encompassing. Any collected data on Persian loans that currently have /æ/ could have five different outcomes, each of which is attested. Thus, the goal of the project is to see how Persian loanwords in Urdu containing either /a/ and/or /a:/ in the Classical Persian form and if it replicates Kuczkiewicz-Fraś's findings. More simply, do modern Persian and Urdu manifestations of the same word differ significantly in the dimension of low vowels?

In the consonantal realm, Hindi-Urdu phonology prior to Persian influence and Classical Persian (the Persian from which Urdu draws most of its loans) both had four places of articulation common with one another: labial, dental, palatal and velar. The fifth unshared place of articulation for Hindi-Urdu was retroflex consonants, whereas Classical Persian had uvular consonants. Urdu borrowed the uvular placement and now possesses /q/ and /ɣ/ ~ /g/, although transiently. Although these transient phonemes were included in her analysis, Kuczkiewicz-Fraś notes they may be absent in utterances save for the “most puristic of registers” and “most formal contexts.” (134) Billah corroborates this, spending a moment on the phonological qualities of Persian that have influenced South Asian languages. The phonemes /z/, /ʃ/, and /f/ are Persian borrowings, in lieu of the traditional /dʒ/, /s/ and /p^h/ in most Indo-Aryan languages. Additionally, the Persian phonemes /g/, /x/, /ɣ/ and /q/ are cited as being used sparingly in Urdu in exclusive and prestigious contexts (105), corroborating the same observation from Hakala (39). These dorsal consonants (/g/, /x/, /ɣ/ and /q/) are thus highly marked when used in Urdu and may occur in variation with more “native” phonemes, such as /k/ and /g/. In fact, prestige Urdu

speakers are curiously even more conservative with their productions of these sounds than the average Tehrani Persian speaker (Kuczkiewicz-Fraś 144).

Current analyses of Tehrani Persian agree that /q/ and /ɣ/ have merged into one phoneme with several allophones, but accounts disagree on what the original phoneme is and resultant allophones. Rees (128) finds that [g], [ɣ], and sometimes [χ] are allophones of /g/:

/g/ → [g] / # ___ C[+Voice]	example: [gejr] 'un-' example: [baɣ-bun'] 'gardener'
/g/ → [ɣ] / V___V	example: [fæ-ɣæt'] 'only'
/g/ → [χ] / ___ C[-Voice]	example: [væχt] 'time'

Oppositionally, Kuczkiewicz-Fraś relies on an account that the merger has led to the voiced velar fricative /q/ as a phoneme and the voiced uvular plosive [g], the voiced uvular fricative [ɣ], and the voiced velar fricative [ɣ] as allophones. Loaning of these sounds into Urdu occurred before the merger took place, so it is possible that Urdu maintains the distinction between [q] and [ɣ] and has the allophonic variant [g], as it has remained in Dari (i.e. Afghan Persian) as well (Rees 152). For the sake of this study, I will be relying on Kuczkiewicz-Fraś's suggestion that /q/ is the merged category for [q] and [ɣ].

Kuczkiewicz-Fraś again summarizes the relevant changes, listing the original Classical Persian dorsal sound and the predicted Urdu sound that is used in replacement or variation:

Summary 2. Consonants

3. /q/ → [k] in casual speech and before front close vowels /i/ and /ɪ/ (145)
 - a. /k/ → [x] (177)
4. /q/ → [x] before a voiceless stop (146)
5. /q/ → [ɣ] before a voiced stop (146)
6. /ɣ/ → [g] initially and intervocally, or [g] in casual speech (146)
7. /x/ → [k^h] in casual speech (149)

Once more, there seem to be multiple paths for a particular sound to shift in Urdu given the context—and even further, these paths are dependent on the formality of the register. In combining the paths suggested by Kuczkiewicz-Fraś’s theorized merger for Tehrani Persian and the various Urdu nativizations, there may be three different current forms stemming from a provided Classical Persian form. Take the words *gam*, *khud*, and *qanun* (meaning *sorrow*, *reflexive self*, and *law* in both languages):

Persian		Classical Persian		Informal Urdu/ Formal Urdu
/qæm/	←	/ɣam/	→	/gəm/ OR /gəm/
/xod/	←	/xud/	→	/kʰud/ OR /xud/
/qɒ:nu:n /	←	/qa:nu:n/	→	/ka:nun/ OR /qa:nun/

*Vowel changes outside those previously reviewed are taken from Kuczkiewicz-Fraś

As the type and use of the less common dorsal consonants in Urdu is highly socially motivated, and the study at hand is not sociolinguistic, I will only impressionistically sketch whether my study’s Urdu speakers maintain distinction between Classical Persian’s /ɣ/ and /q/ in their current productions. As a sanity check and point of comparison, I will also ascertain whether the category is generally merged in modern Persian.

Finally, Kuczkiewicz-Fraś is careful to note that although Urdu has historically constructed a system of nativization for Persian loans, phonetic realization is a quite different beast depending on register, style, region, and speaker. Arabic-originating loanwords that have been filtered through Persian and then Urdu phonology may be retroactively Arabicized in pronunciation to indicate a Muslim scholarly sense, even though this filter is applied on a phonetic representation that has already been twice-filtered itself (134). Since Persian (and Urdu) words are written in their native scripts without indication of the short vowels, there is yet disagreement on the pronunciation of certain rare and largely literary loans which are variably interpreted by speakers in prestige settings (62). As implied by the length and seeming contradictory processes of adaption listed above, many are alternations that can be used in

different circumstances and exist within the same speaker. An omitted rule from **1. Vowels**, /ɪ/ → [ə], is illustrated through the word “use.” This began in CP as /ɪntɪzɑːr/ (currently /ɛntɛzɒːr/ in modern Persian), and takes variable form as /ɪntəzɑːr/ ~ /ɪntɪzɑːr/ in normative Northern Urdu. However, I have heard both forms, and produce both forms as well, and have no conception of one form as ascribed to a certain register. Often, a sound between /ə/ and /ɪ/ is also produced for the second critical syllable. So, while there is knowledge that both are acceptable forms, there is less known about the significance of one or the other, or why one has not been made the standard form as of yet.

3. Methods

3.1 Materials

Kuczkiewicz-Fraś’s *Perso-Arabic Loanwords in Hindustani. Part I. Dictionary* (2008) was used to gather words attested in both languages. The dictionary is a compilation of other such existing dictionaries, and each entry is extensively researched for etymology, including pronunciations in the CP as opposed to current normative Urdu iterations. Because the size of the corpus was too large to test with native speakers, words were selected on the basis of matching an existing category known for Persian borrowing: religion, distinctive Muslim practices, tailored clothing and garments, urban orientation, literacy, administration, honorific formations, body-organs, kitchens, fruits, flowers, and inspiration (Billah 107). The selection from this work created the first list.

A second list was also constructed through the dictionary to assess accuracy/retention of [q], [ɣ], [x], [k] and [g]. Nine words beginning with a given sound in their CP form were selected, primarily based on my own ability to recognize the word. As an imbalanced bilingual, this was a quasi-litmus test for my intended participant pool of Urdu speakers—if even I

produced it, this usually indicated it was in the lexicon of balanced Urdu-English bilinguals. This was insufficient for Persian speakers, but no sources exist to my knowledge that could confirm that both Urdu and Persian speakers currently use the same given word at comparable rates.

This initial body of words to test consisted of the following: a first list comprised 140 words, with a range of words that were semantically in line with prototypical Persian loans, and a second list, comprised of 45 word-initial [q], [ɣ], [x], [k] and [g] (9 tokens per phoneme). Both lists were scrambled as to reduce phonemic priming.

Words were removed from the list as the study proceeded, due to methodological difficulties in elicitation, i.e. participants regularly failed to recognize the word, or there were too many alternatives that were more common in use before the target word was produced, therefore disproportionately extending the length of the interview. Words that were shared between List 1 and List 2 were removed from List 1 as to avoid overlap. As the scope of the project changed, only words that contained /ə/ and /ɑ/ in Urdu were kept for analysis.

This resulted in a List 1 with 39 words as opposed to 140, and a List 2 with 30 words. Word translations were continuously updated on participant's judgements. Both the final versions of List 1 and List 2 as well as eliminated types from each are provided in the Appendix.

3.2 Participants

Participants were native Urdu or Persian speakers. To control for regional variation, speakers of Iranian Persian (as opposed to Afghan or Tajik, etc.) and Pakistani Urdu were targeted for recruitment. In order to control for interference, imbalanced bilinguals were not eligible for study; participants needed to have spoken the language at a native-level up to puberty in a setting in which it was one of the dominant languages, i.e., primarily Iran or Pakistan. One Persian speaker grew up outside of Iran, but spoke exclusively Persian at home. Age was not

controlled for, as the loaning process had occurred long enough ago that most forms were static. As such, most Urdu participants were under 24 years old while most Persian speakers were over 30 years old. All speakers were literate in the respective language they were selected for. Six speakers per language were recruited through Facebook, Twitter, and email. Urdu speakers were 7C international students from Pakistan. Persian speakers varied more, as none of them were students but rather professors or residents in the Claremont area. Two Persian consultants were L.A residents found through Twitter. Save for the one participant who grew up outside of Iran, 4 or the 6 participants were from Tehran; 1 was from Isfahan, a city 211 miles south of Tehran.

3.3 Procedure

Participants were briefed on consent and were filled out a demographics survey collecting data about age, gender, nationality, how long they've spoken their respective language for, where they grew up, their religious background, and the nature of their high school schooling: if it was private or public, the languages classes were taught in, and how typical it was for students to go to college internationally. They were then taken to the LGCS sound booth (or in the case of the two LA residents, to an adequately quiet, private room), where they were seated, facing myself. A Zoom H5 Handy Recorder was placed on a table between myself and participant, about two and a half feet from the participant's mouth. The recorder was angled up at the participant by resting against a small block, and turned on with the input level set at seven. I had a laptop on my lap with a spreadsheet of both lists open on Excel.

Participants were instructed to provide translations for English words into the language for which they were at the study for. They were told that it was alright if they could not think of a

word immediately, as they could return to the word later on. They were free to indicate if they wished to skip a word and return to it later.

Starting from List 1, I listed English words one word at a time, pausing in-between for the translations provided by participants. If participants guessed the target word, I would nod and proceed to the next English item. If not, I would indicate with a rolling hand motion for the participant to continue guessing. If the participant indicated they could not think of another alternative, the English definition would be extended on, or provided in an example situation. For example, *fikr* (thought, concern) was first introduced as “thought, or concern” and then later extended upon (if needed) as “You may use it in the phrase, ‘Don’t worry’ in [respective language]. If participants could then not produce the target word, I would mark the word on a spreadsheet with both lists and proceed to the next item.

After the List 1’s first run, marked words were returned to for additional consideration, with the extended English definition and an example context (Urdu speakers) or literary form (Persian speakers) provided. If the word was then produced, the mark on the spreadsheet was adjusted to reflect that it was induced at the second interval. If participants were still unable to produce the target word, I would ask, “Do you know the word, [target word]?” I would carefully pronounce the target word. If participants answered affirmatively, they were both asked to repeat the word that I had just produced and what their definition of the word may be, or where they may use this word. Feedback was appended to the current list and utilized for future sessions, and the mark on the spreadsheet was adjusted to reflect that it was overtly and transparently induced through repetition. If the word was still not recognized, it was left with the original mark to reflect its lack of recognition. List 2 was recited in a similar manner in a separate recording immediately following List 1.

Sessions lasted anywhere from 30 minutes to an hour, including time for consent-briefing and the survey. Participants were compensated \$12.50 for their time.

4. Analysis

The target words from each interview were copy and pasted into a new WAV file, such that each participant had a WAV file dedicated solely to their target words for vowel analysis and another WAV file for their consonant target words. Vowels were then segmented in Praat and annotated for language, participant number, gender, word, syllable position, syllable stress, and following syllable stress. Additionally, the “underlying” Classical Persian vowel as well as the impressionistically interpreted surface vowel were annotated. A Praat script collected every instance of a segmented vowel and output a spreadsheet containing all the values.

Six of the words from the final List 1 included words that ended in [ɑ:] in Urdu and [e] in Persian due to morphemic motivations, as outlined in 2.2.2. Their second vowel was excluded from analysis due to phonetic differences between the languages that are not phonologically motivated. Three of the words in List 1 had two syllables in Urdu, and one syllable in Persian due to elision of the second vowel, and these latter sounds were also excluded from the analysis. 38 types/461 utterances yielded 641 tokens comprised of the following: 350 /a/ and 291 /ɑ:/ in the CP form; 154 /ɒ/, 159 /æ/, 174 /ə/, 137 /ɑ:/, and 17 /e/ in the impressionistic surface form. Each data point consisted of the first three formant values, and the total set of data was normalized through the Bark Difference Metric. The Bark Difference Metric is a vowel-intrinsic method that can normalize formant values within and between speakers without requiring the total vowel space, a method that is ideal for the scope of this project. It is also normally able to retain sociolinguistic variation in the data. Nonetheless, this method also relies on F₃ values than

more so than other methods, a measurement that can be quite faint in recordings and thus prone to misreading by Praat.

Formant values are run through the formula:

$$Z_i = 26.81/(1+1960/F_i) - 0.53$$

When i represents the formant's ordinal value. The normalized F_1 is calculated through Z_3-Z_1 , the normalized F_2 through Z_3-Z_2 , and the normalized F_3 through Z_2-Z_1 .

Generally—but not always—the higher the non-normalized F_1 , the lower the BARK normalized F_1 , and vice versa. Likewise, the same relationship applies F_2 . Conversely, the actual measured F_3 and the normalized F_3 have a positive relationship.

The Interquartile Rule was employed to exclude outliers from analysis through the following process: An Interquartile Range (IQR) is calculated by subtracting the first quartile of data (Q_1) from the third quartile of data (Q_3). The IQR is multiplied by 1.5, and the resulting product is added to Q_3 . All data points greater than this final value is deemed an outlier and removed from analysis. Similarly, the same product is subtracted from the first quartile. Any data points lesser than this final value is also deemed an outlier and removed from analysis. The Interquartile Rule was applied on the first three formant values by speaker language. This resulted in 537 tokens (removing 104 tokens from analysis): 255 /a/ and 282 /a:/ in the CP form; 131 /ɒ/, 130 /æ/, 147 /ə/, 124 /ɑ:/, and 5 /e/ in the impressionistic surface form. There were 582 observations for F_1 (287 for Urdu), 628 observations for F_2 (306 for Urdu), and 585 observations for F_3 (298 for Urdu).

5. Results

5.1 Vowels

Results have been interpreted with the “underlying” Classical Persian vowel in mind. Kuczkiewicz-Fraś proposes that typically, the Classical Persian vowel /a/ shifted to /ə/ in Urdu, and likewise the Classical Persian /a:/ shifted to /ɑ:/ in Urdu. She simultaneously cites that Classical Persian reconstruction research has suggested that this same Classical Persian /a/ went to /æ/ in modern Persian, and /a:/ went to /ɒ:/. Thus far, these are two separate theories that have not been examined together. Although Classical Persian reconstruction has relied on data from phonetic studies, Persian loanwords in Urdu have been examined phonologically but never phonetically. Subsequently, the theory of vowel shift for /a/ and /a:/ is largely based off impressionistic data. In this section, I examine the acoustic quality of the current vowels from Urdu and modern Persian that have derived from the “underlying” Classical Persian vowel. I will be answering the following questions:

1. Do modern Persian and Urdu manifestations of the same word differ significantly in the dimension of low vowels?
2. Do these current low vowels match the existing theories of vowel movement, visualized below?

Urdu Vowel	Classical Persian Vowel	Modern Persian Vowel
/ɑ:/	← /a:/ →	/ɒ:/
/ə/	← /a/ →	/æ/

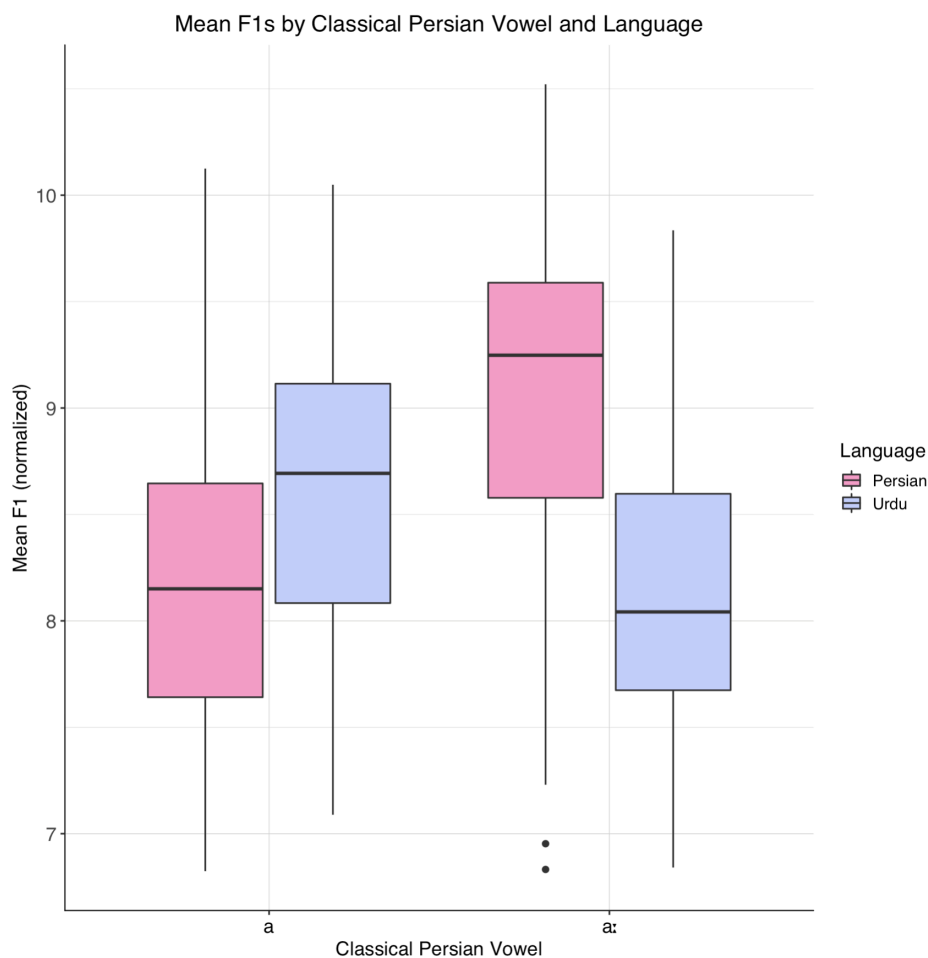


Figure 1. Mean Normalized F₁s by Classical Persian Vowel and Language

Figure 1 depicts the average normalized F₁s of each language's surface vowels for the Classical Persian /a/ and /a:/. The medians of the Persian /a/ and /a:/ F₁ values are 8.151 and 9.248 Barks respectively, while the Urdu /a/ and /a:/ F₁ values are 8.042 and 8.693 Barks. A regression model found a significant interaction effect of language and Classical Persian vowel ($p < .0001$), suggesting that the collected F₁s for each Classical Persian vowel are significantly different between the two languages. However, this model's Multiple R-squared value is only 21.56%, meaning that nearly 80% of the variation in the data cannot be attributed to Classical Persian vowel, language, or their interaction as factors. The standard errors across each vowel's F₁s are similarly too large to unambiguously conclude the accuracy of the attested difference.

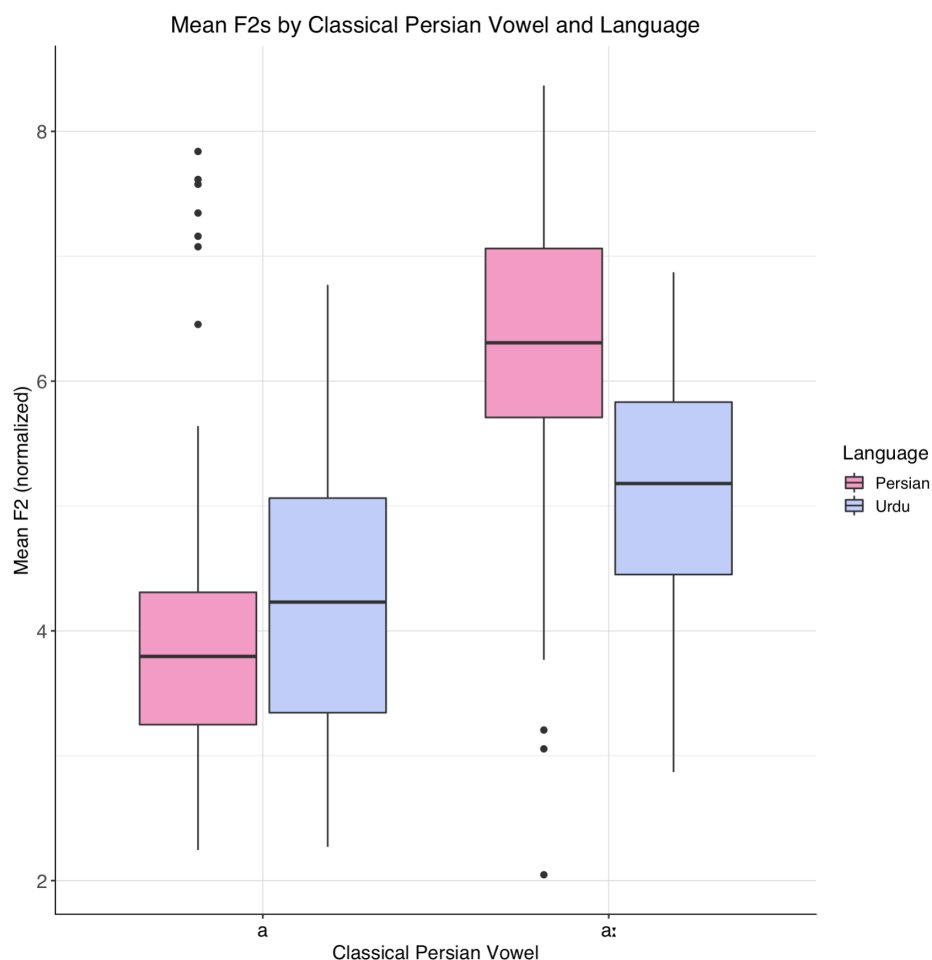


Figure 2. Mean Normalized F₂s by Classical Persian Vowel and Language

Figure 2 depicts the average normalized F₂s of each language's surface vowels for the Classical Persian /a/ and /a:/. The medians of the Persian /a/ and /a:/ F₂ values are 3.795 and 6.307 Barks and the Urdu /a/ and /a:/ F₂ values are 4.231 and 5.180 Barks. A regression model again found a significant interaction effect of language and Classical Persian vowel ($p < .0001$), suggesting that the collected F₂s for each Classical Persian vowel are also significantly different between the two languages. This model's R-squared value is still only 41.67%, so a little over half of the variation in the data cannot be attributed to Classical Persian vowel, language, or their interaction as factors. Once more, the standard errors across each vowel's F₂s are too large to unambiguously conclude the accuracy of the attested difference.

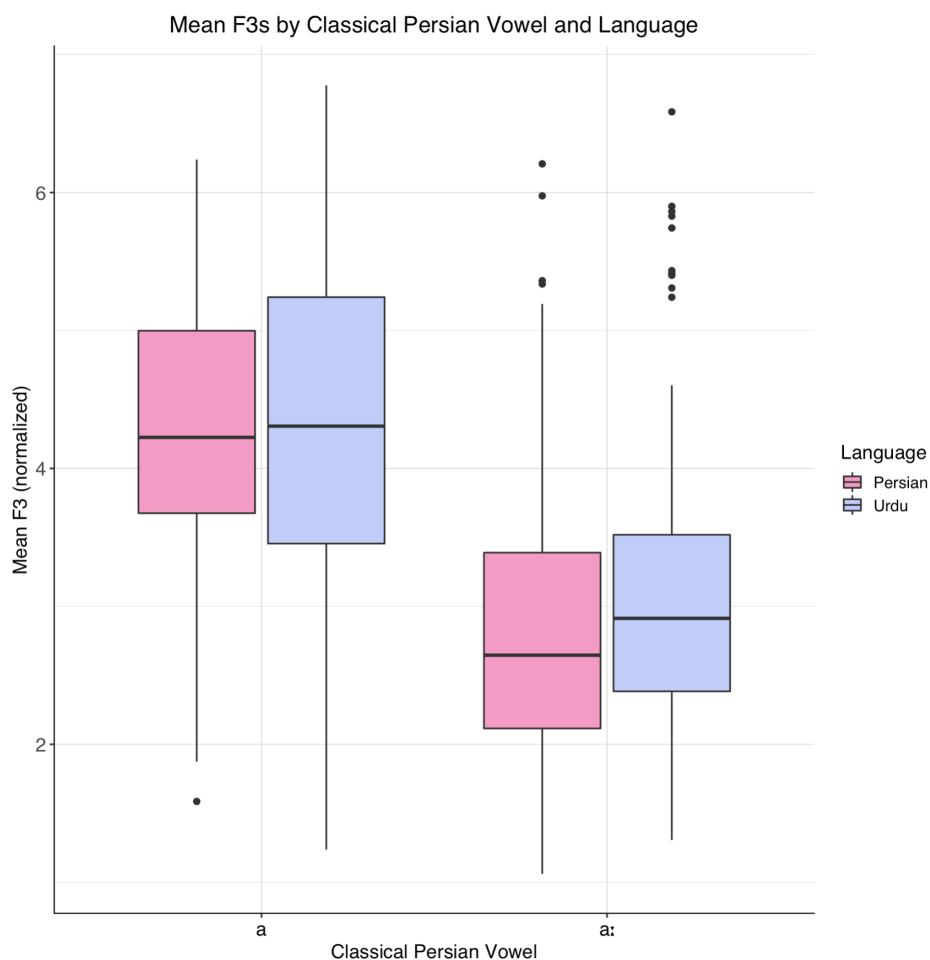


Figure 3. Mean Normalized F₃s by Classical Persian Vowel and Language

Finally, Figure 3 depicts the average normalized F₃s of each language's surface vowels for the Classical Persian /a/ and /a:/. While the normalized F₁ and F₂ were visibly different by language and vowel (see: fig. 1, fig. 2), the medians of the Persian /a/ and /a:/ F₃ values are quite alike to the Urdu counterparts. The Persian measurements are 4.225 and 2.646 Barks for /a/ and /a:/, respectively; the Urdu /a/ and /a:/ F₃ values are 4.307 and 2.913 Barks respectively. A regression model did not find a significant interaction effect of language and Classical Persian vowel ($p = .328$), providing evidence that the collected F₃s for each Classical Persian vowel are not significantly different between the two languages. This model's R-squared value is 29.11%, so two-thirds of the variation in the data cannot be attributed to Classical Persian vowel,

language, or their interaction as factors. Further summary of data measurements related to normalized formants can be found in the appendix (see appendix, 7.3).

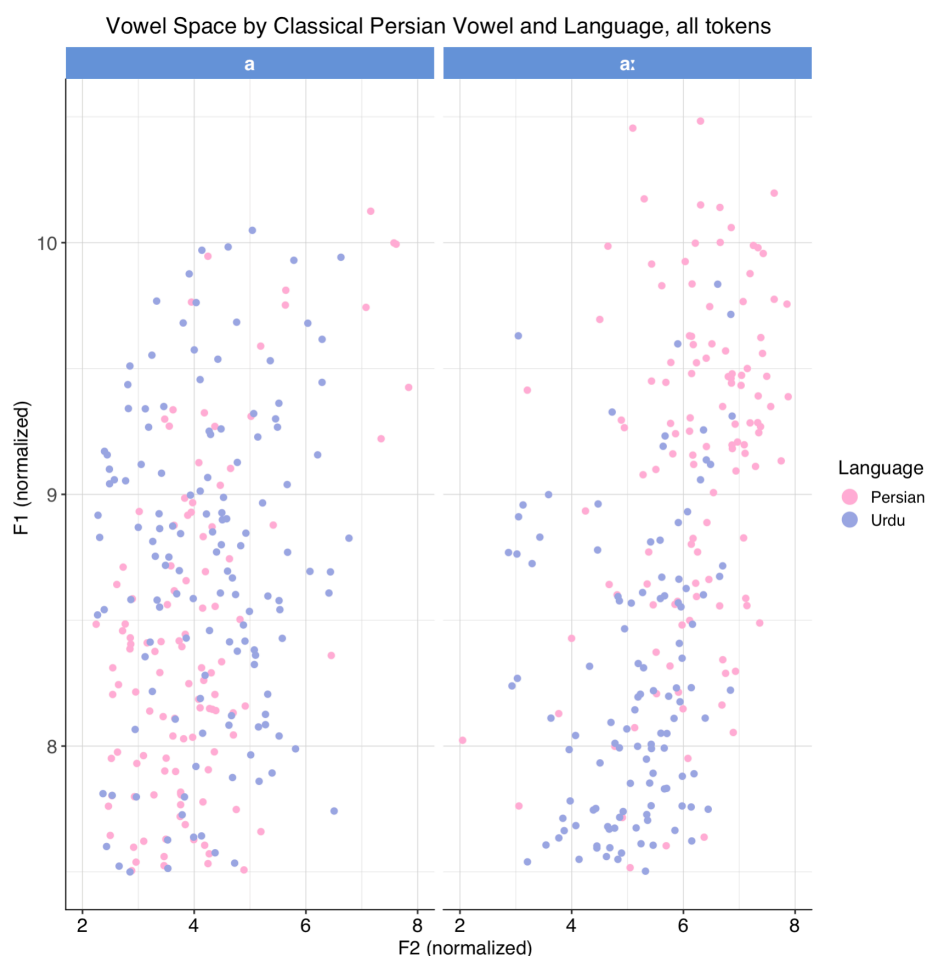


Figure 4. Vowel Space by Classical Persian Vowel and Language, all tokens

Figure 4 is a plot comprising of every collected non-outlying token from the project, separated by the Classical Persian Vowel and colored by speaker language. The tokens are organized among normalized F_1 and F_2 values. The spread of data is impressive. The /a:/ plot hints at the clustering of Persian tokens as having higher F_1 and F_2 Barks values (thus populating the upper-right of the vowel space) and the Urdu tokens having conversely lower values (populating the lower-middle). Even so, some Persian /a:/-derived tokens are as low as the Urdu counterparts. The /a/ plot has far less apparent patterning. The Urdu /a/-derived tokens range

fairly evenly between 7.5-10 Barks on the F_1 and 2-6 Barks on the F_2 , occupying a huge span of space. Any signal to be gleaned from the data is, unfortunately, quite faint.

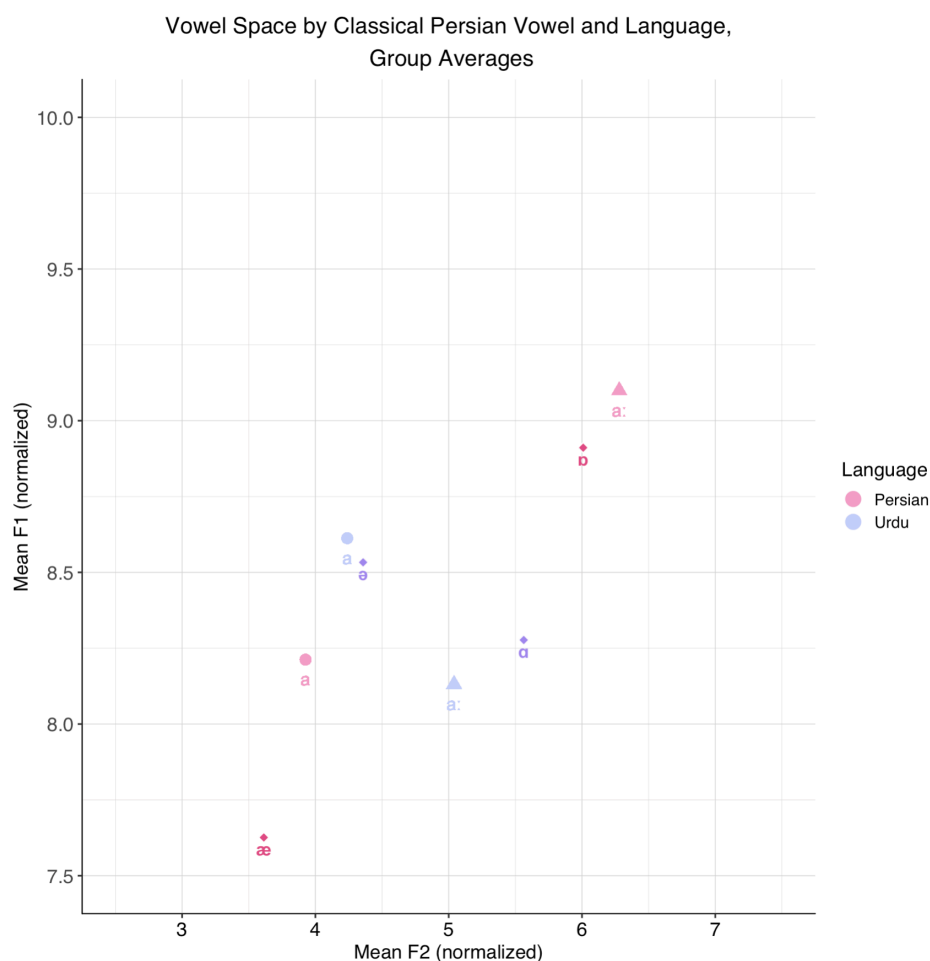


Figure 5. Vowel Space by Classical Persian Vowel and Language, Group Averages

Figure 5 is a plot of the *total* average normalized F1 and F2 value of each language's current derived vowel for /a/ and /a:/. As Figure 4 made it evident that the spread of data is quite wide, the averages in Figure 5 do not necessarily reflect an accurate vowel position for every speaker, but is valuable for two reasons. One, it loosely summarizes the position of each current vowel per language and thus enables some amount of comparison between them. Persian's current vowel for /a:/ is further apart from Urdu's current vowel for /a:/ than their respective

surface vowels for /a/, which are relatively close to one another. Furthermore, Persian's vowel space for the low vowels seems more spread out than Urdu's.

Figure 5 moreover includes the normalized F_1 and F_2 values of each language's purported current vowel, taken from existing phonetic papers on Persian (Ansarin 2; Mohamadi 83) and Urdu vowel space (Fatima 74). They are shown with diamond points, and in colors slightly darker than their associated language's key color. This was included to provide additional context and review whether the vowels are in the expected regions that previous vowel movement theories have suggested. Indeed, the collected current vowel averages largely match the relationship that the canon vowels hold: the current Persian /a:/ is higher and further right than all of the other vowels; the Urdu /a/ is the second-highest and second from the left; the Urdu /a:/ is second from the left, and the Persian /a/ is further left than all of the other vowels. However, the Persian /a/ should also be the lowest, a position that is instead occupied by the Urdu /a:/, which should be the second-lowest. This lack of correspondence in height suggests that while it is supported that the current form of the Persian /a:/ is generally in the region of /ɒ:/, the Urdu /a:/ generally in the region of /ɑ:/, and the Urdu /a/ fairly close to /ə/, the current form of the Persian /a/ may not be much like /æ/ but rather closer to a /ə/, /ɛ/, or /ɜ/. Concurrently, it must be reiterated that these group averages vastly oversimplify the collected data—some Persian speaker's current vowel for /a/ may be exactly where /æ/ is, whereas others may have something as high as /e/, creating an average resembling, say, /ɛ/.

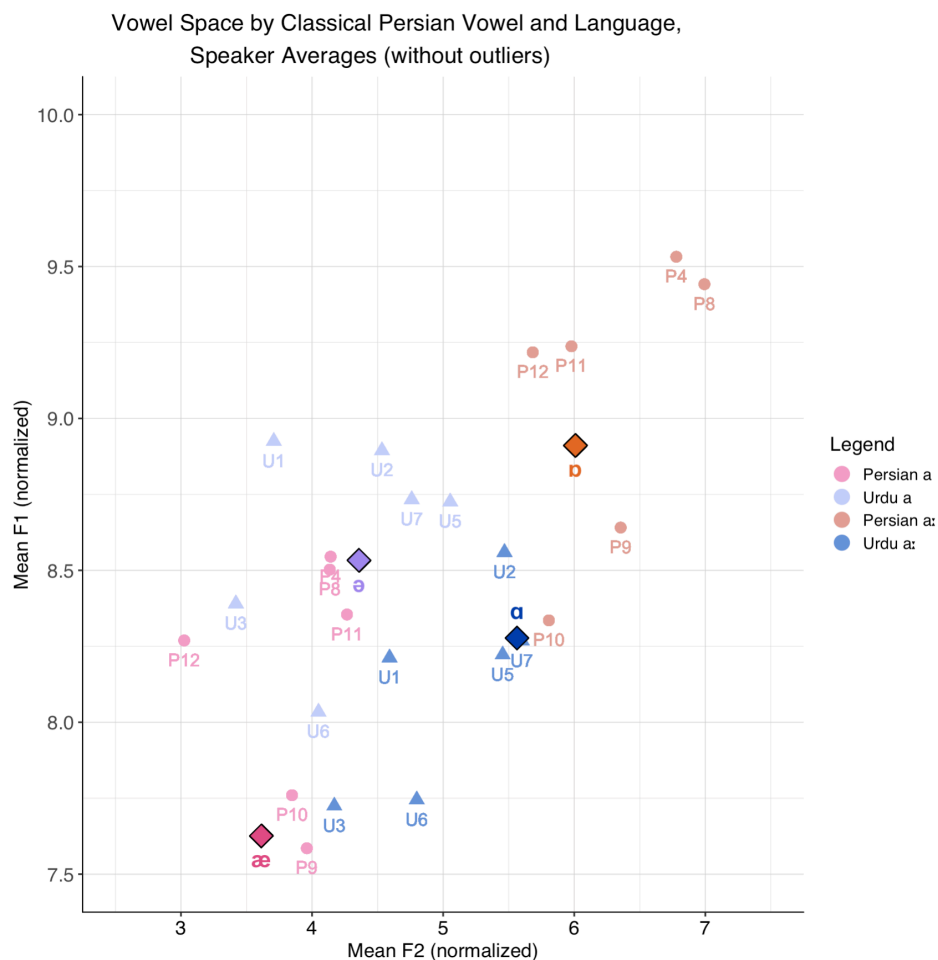


Figure 6. Vowel Space by Classical Persian Vowel & Language, Speaker Avgs (without outliers)

Figure 6 looks at the same vowel space as Figure 5 with attention to individual speakers' current derived vowels. Points are colored according to a combination of speaker language and Classical Persian vowel, and once again plotted alongside the canonical positions of the theorized current vowels. The theorized current vowels are colored in a darker hue than their collected data point counterparts, drawing attention to the location of the *actual data* versus where it should *ideally* be were it to affirm aforementioned vowel movement theories. Even with outliers eliminated, there is still a sizable span of data only vaguely in the vicinity of their respective theorized position.

Curiously, while Urdu /a/-derived speaker averages seem to circle the canonical /ə/ at nearly equidistant intervals, there are three Persian /a/ derived speaker averages (P4, P8, and P11) that are far closer to /ə/ than any Urdu /a/-derived average. Contrastively, two Persian speaker averages are directly adjacent to /æ/ (P10, P9). P12 is intermediately spaced on the normalized F₁ between these two clusters, but far lower on the normalized F₂ than either. Several Urdu /a:-derived speakers are incredibly proximate to /ɑ:/ (U5, U7). Two were far closer to /æ/ (U3, U6). Once again, the collected vowel spaces overtly overlap between the languages.

Of considerable interest is examining each speaker's derived vowel averages in tandem. Persian /a:-derived speaker averages for participants P8 and P4 are the highest and furthest-right on the plot, a fair deal away from the canonical /ɒ:/. Simultaneously, the same participants' Persian /a/-derived averages are the highest and furthest-right from the canonical /æ/. Keeping in mind that higher normalized F₁s and F₂s implicate *lower* actual formant values, this suggests that these participants—P4 and P8—may have a higher and backer vowel space overall. There are several more cases where both of a speaker's vowel averages have a similar spatial relationship to the theorized canonical position on at least one dimension, if not both (U1, U2, U3, U6, P11, P12). This illustrates the variety of different vowel spaces in the participants.

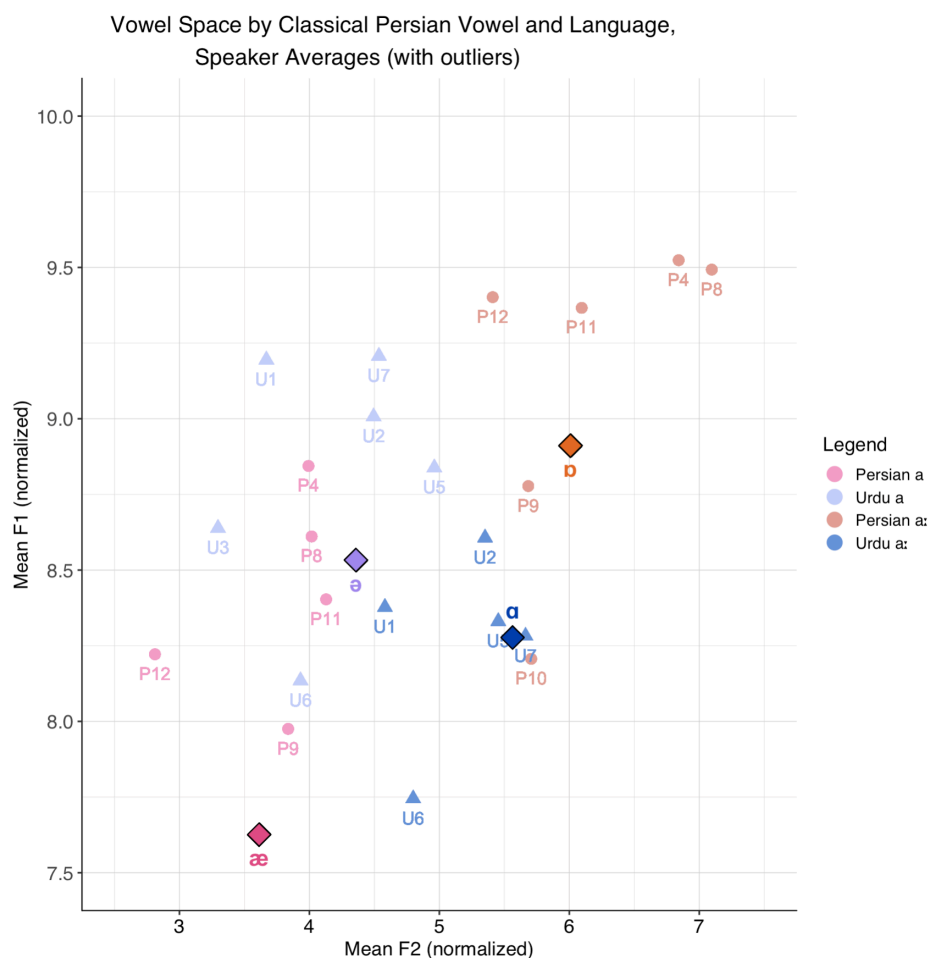


Figure 7. Vowel Space by Classical Persian Vowel and Language, Speaker Avgs (with outliers)

In examining the same space as Figure 6 with the addition of outliers, patterns within the data fade out and appear more random (figure 7). All previous figures have represented the data without incorporating said outliers. It is of relevance to understand how ostensibly significant findings can wash away with consideration of these outliers; relatedly, all previous regressions referenced for Figure 1, 2 and 3 do not conclude significant language and Classical Persian vowel interaction effects when considering participants as a random intercept. Therefore, it cannot be unambiguously concluded that significant differences in the vowel space are attributable to language as opposed to simply individual speaker variation.

5.2 Consonants

Consonant data was not phonetically analyzed, but rather impressionistically reviewed to examine basic variation within speakers. In this section, I will be answering the following questions:

1. How are [q], [ɣ], [x], [k] and [g] used among Urdu and Persian?
2. Are [ɣ] and [q] merged in modern Persian?
3. Do Urdu speakers maintain distinction between Classical Persian's /ɣ/ and /q/?

List 2 included words that started with [q], [ɣ], [x], [k] and [g] in the Classical Persian form; the first three consonants [q], [ɣ] and [x] are variably rated as present in Urdu, while [q] and [ɣ] are considered a merged consonant in Persian that was apart from [x], which is definitively present. The second two consonants, [k] and [g], are considered shared between the two languages, and also a likely replacement for such sounds as [q], [ɣ] and [x] in casual Urdu speech. These velar stops were also included to determine if Urdu speakers overperformed speech sounds, such as replacing words that began with [k] and [g] natively with [q], [ɣ] and [x].

345 consonant tokens were collected total (172 for Urdu), comprising of 66 [g] tokens from 6 words, 82 [k] tokens from 7 words, 72 [x] tokens from 6 words, 67 [ɣ] tokens from 6 words, and 58 [q] tokens from 5 words.

5.2.1 [k] and [g]

Among all Urdu and Persian speakers, [k] and [g] were present and separate categories, and among Urdu speakers, these velar stops were never confused or replaced by a more “Persian” sound. One Persian speaker (P4) used /tʃ/ in place of /k/ and /dʒ/ in place of /g/ for a handful of words, solely preceding vowels that derived from the Classical Persian /a/, although

the dataset is too small to generalize this as any sort of rule. This speaker is reportedly from Tehran, but no other Tehrani participants had a similar alternation pattern.

5.2.2 [x]

All Persian speakers had a phoneme [x] that was used for the expected words, while only 5 of the 6 Urdu speakers produced [x] accurately (i.e. used for a word that started with [x] in Classical Persian) and regularly (i.e. used for all such words). The one Urdu speaker who did not use [x] instead produced [k^h] in its place for all instances.

5.2.3 [ɣ] and [q]

5 out of 6 Persian speakers had a merged category for [ɣ] and [q], surfacing mostly as /q/, although at times this was more alike to /g/, following Rees's account (128). As determining the voicing status of these consonants is outside the scope of this paper, it is sufficient to conclude that some uvular stop—either /q/ or /g/- is the merged category for [ɣ] and [q] within the majority of the Persian-speaking participants. One Persian speaker did not seem to have the category fully merged, as he used /ɣ/ accurately twice, while using /q/ in the other three instances of a Classical Persian /ɣ/. This speaker also happened to be the oldest of the entire participant group (77 years old) and was also the sole speaker from Isfahan, suggesting that the merger is not total for all generations and/or all regional dialects.

Contrariwise, 5 of the 6 Urdu speakers did produce the phoneme /ɣ/ accurately and regularly. This does not immediately mean that the Classical Persian distinction between /ɣ/ and /q/ is maintained in these speakers, as many also typically produced /k/ in the place of /q/. This observation is harder to quantify, as speakers used /q/ accurately for only some words; 3 speakers for *qanun* (law), all 5 speakers for *qalam* (quill pen), and 2 for *qatra* (drop of liquid), and 1 for *qainchi* (scissors) and *qismat* (fate). Moreover, some productions of what I perceived as /k/

seemed intermediate between /k/ and /q/ on closer inspection, implying that Urdu productions of /q/ are closer to /k/ than the equivalent Persian productions, which are further apart.

One speaker did not produce /q/ or /ɣ/ at all (the same speaker who lacked [x] in 5.2.2), regularly replacing them with [k] or [g].

5.2.4 *The Role of Prestige*

The status of a prestigious education was only relevant for Urdu speakers, where consonant quality could vary on prestige value. Based on the results from the demographic survey, education was deemed prestigious if their schooling was in English, was private, and had at least 20% of their graduating class go to college in the Western world. All Urdu speakers, save for one, had a prestigious high schooling experience. This is unsurprising, as international students attending American private liberal arts colleges are known to typically have formal education of a high standard in their home countries. Three speakers all graduated from the same Cambridge-style high school in Karachi, Pakistan that included Urdu as a formal subject in the curriculum. Such schooling likely emphasized the prescriptive standard for Urdu productions of Persian loanwords. 5 of the 6 speakers were generally accurate and regular with their velar and uvular consonants. However, one participant did not produce any of these phonemes save for [k] and [g], and had trouble with a good number of words (although he recognized them once I produced them) despite having grown up in Lahore. When I asked him why he could recognize but not produce the words he missed earlier, his reason for most words were that they were not used in daily speech, and would perhaps be used in poetic and academic settings.

6. Conclusion and Discussion

6.1 *Main Findings*

I set out to study two phenomena related to Persian loanwords in Urdu: One, whether the current low vowels in Urdu and Persian deriving from Classical Persian's /a/ and /a:/ are significantly phonetically different and matching existing theories of vowel change from Kuczkiewicz-Fraś; and two, how the dorsal consonants [q], [ɣ], [x], [k] and [g] are differently produced by Urdu and Persian speakers.

The collected data suggests that there are significant differences between the F₁ and F₂ values of each language's /a/ and /a:/-derived vowels, but the spread of data and the inability of regression models to explain the majority of the variation insinuate that the purported difference does not solely represent the interaction effect of an “underlying” Classical Persian vowel and language. The language group F₁ and F₂ averages for each vowel have a similar spatial relationship to the canonical position of their theorized current vowel—that is, /æ/ for the Persian /a/-derived vowel, /ə/ for the Urdu /a/-derived vowel, /ɒ/ for the Persian /a:/-derived vowel, and /ɑ:/ for the Urdu /a:/-derived vowel—this relationship is not wholly accurate, as the Persian /a:/-derived vowel has far a higher average normalized F₁ value than /æ/. Further inspection of speaker averages by language and Classical Persian vowel reveal once more that the data spans a large vowel space, where some speaker averages are quite close to canonical positions while others are comprehensively shifted and still others dissimilar in seemingly random ways. Seeing as significant differences wash away with the inclusion of a random intercept variable, this study did not find unambiguous evidence that the Classical Persian vowel /a/ went to /æ/ in Persian and /ə/ in Urdu, not that the Classical Persian /a:/ went to /ɒ/ in Persian and /ɑ:/ in Urdu as suggested by previous non-phonetic research.

The comparison of modern vowels between the two languages provides a remarkable way to track theorized vowel change. While the data collected did not definitively affirm and unite the theories of Classical Persian to Urdu/Persian vowel shifts, future research that further controls for such variation as stress, age, gender, and participant level of fluency may yield better results. Even so, there is the possibility that vowel change is taking place at this very moment, and that this project's inability to unearth a one-to-one correspondence of a Classical Persian vowel to modern vowel is reflective of the ongoing change.

The nativization process is similarly fascinating to examine in the purview of diachronic methods. Researchers do not have unambiguous evidence of the average formant values of the Classical Persian vowels, but solely conjecture based on reconstruction, which is in turn based on comparative analysis. Rees tracked the development of older Persians into the current range of Persians, including Iranian, Dari, and Tajik by comparing which mergers and changes did and did not take place between the three, and subsequently theorizing which sounds were underlying or preceded. Kuczkiewicz-Fraś's work in combination with this project still suggests that Urdu loanwords are another possible site of analysis for CP reconstruction. Furthermore, it shows that the nativization process is only one event in a long history of sound change. The purportedly connected sounds of [ɑ:] and [ɒ:] or [ə] and [æ] may become more and more distal with time as sound change naturally occurs.

With respect to dorsal consonants, all Persian speakers produced /k/, /g/ and /x/ accurately and regularly. 5 out of 6 Persian speakers never produced /ɣ/, instead possessing the merger of /ɣ/ and /q/ that surfaces as /q/--although at times this surface segment sounded closer to /g/, as has been suggested by certain studies (Rees 128). One speaker had half of his /ɣ/ productions matching that of the /ɣ/ and /q/ merger, i.e., surfacing as /q/, but also used /ɣ/ for the

other half of such words. As such, I conclude that [ɣ] and [q] are largely merged in modern Persian to /q/ but not completely for all speakers.

Likewise, all Urdu speakers produced /k/ and /g/ accurately and regularly, while 5 of these 6 speakers could also produce /x/ and /ɣ/ accurately and regularly. These same 5 speakers had variable production of /q/, using it accurately but only semi-regularly, and otherwise replacing it with [k]; additionally, some productions of /q/ seemed to exist between /k/ and /q/. This suggests that Classical Persian's /ɣ/ is retained, but the explicit Classical Persian distinction between /ɣ/ and /q/ has been lost as the latter segment is rare or moved to the velar position. One speaker regularly substituted [k^h] for /x/, [g] for /ɣ/ and [k] for /q/. These replacements were all in accordance with the Rule 3, 6, and 7 of Kuczkiewicz-Fraś's observations (see Summary 2. Consonants in Section 2.2.2). This does not indicate this speaker never produces /x/, /ɣ/, or /q/; when I asked about these segments during the interview he simply indicated that was not the manner in which he would normally speak as it sounded "formal" to him. These findings support existing research positing that the dorsal consonants /x/, /ɣ/, or /q/ are rarer and have prestige value associated with their use in Urdu. This portion of the project was meant to survey how borrowed "Persian" consonants are used in Urdu such that future research with explicitly sociolinguistic design can further elaborate on these patterns.

6.2 *Limitations and Future Research*

The background research and collection of data were not completely separate phases during the creation of this project. This led to certain design errors that could have contributed to variation in the data. Pakistani Urdu speakers were sought after due to the status of Urdu as the official language of Pakistan; it was later that I realized many of the Urdu speakers in Pakistan are L2 speakers due to the fact that languages indigenous to Pakistan are Punjabi, Sindhi, Pashto,

and Balochi (Khan 277)—Urdu was considered a foreign import from Northern Indian Muslim immigrants to Pakistan. Furthermore, it was only through conversation that I discovered two of my Urdu speaking participants were actually Sindhi L1 speakers, which could easily affect vowel quality. In the same vein, Persian speakers spanned a huge variety of age and origin; again, while I sought out Tehrani Persian speakers, in examining the survey results after interviews I determined that while the majority participants grew up mostly in Tehran, others moved around within Iran during their youth. Finally, despite the fact that nativization occurred some time ago, the age range difference between the groups may have played a key part in the heterogeneity of vowel spaces even within languages. Although this study was intending to study a diachronic incident, the scope of the project prevented me from individually examining very recent and/or current sound changes in Persian that could account for the lack of uniformity in the data.

Kuczkiewicz-Fraś provided an incredible resource by creating a dictionary of most every loanword of Persian origin in Hindustani. It is about 945 pages long, with about 7 entries per page, suggesting a total of around 6600+ loanwords that are up for investigation. As my project morphed with time, I was unable to most accurately select words from the dictionary that contained /a/ and /a:/ in Classical Persian and had similar statuses of stress and local environment, again contributing to a mass of data that was manifold. *Saadaa* (meaning plain or simple), which canonically is pronounced as /'sa:.da:/ in Urdu, was at times pronounced as /'sa:.də/ due to the status of the second syllable as unstressed. I did not account for stress in my models. Future work would do well to better select for controlled target words.

Moreover, the quality of my phonetic data was somewhat degraded and faint due to a lack of experience conducting phonetic research. The Barks Difference Method was therefore

risky to employ as it relies so heavily on already faint F_3 values to normalize data, and the particular haziness of my data only exacerbated the problem.

Other limitations of this project were of course precipitated by the goals of the project being focused on linguistic phenomena from languages that are most commonly spoken in Southwest and South Asia. The ideal version of this project would take place in Tehran and Delhi, where concerns about bilingual interference and lack of control for age or gender could be washed out with a large enough population or simply more prerequisites for participation in the project. Relatedly, much of the background literature I reviewed over the course of this project conflicted each other, leading to difficult decisions as a novice researcher regarding which theories, models and inventories to take more seriously than others. At times, incredibly pertinent research was written in other languages and I was unable to translate from, or only available at institutions in South and Southwest Asia itself.

The project's requirement of having shared Classical Persian words between Urdu and modern Persian lead to the most interesting observations. I was only able to design the target word lists with my own proficiency in Urdu in mind, and could therefore not systematically confirm beforehand whether said words were still used in Tehrani Persian in similar contexts (if at all!). While Urdu speakers were able to guess nearly all the words by the second inducing, Persian speakers struggled far more and gave variable judgements on the meaning of the word. The vast majority of the recognized words for Persian speakers were produced through the literary form rather than through my own verbal cues. One speaker also knew Dari, and he noted that many of the words I was asking for were not often used in Persian and had antiquated meanings, but were used in the fashion I was requesting by Dari speakers. That is, Iranian Persian speakers no longer used *teiyaaar* to mean ready, although Afghan Dari speakers would in

the same way Pakistani and North Indian Urdu speakers would. Semantic shift was clearly playing a part in the disconnect between Iranian Persian and Urdu definitions of the shared words, but it was unclear which language had retained a definition that was more alike to the Classical Persian meaning. For example, *takleef* (“trouble, difficulty, or inconvenience” in Urdu) more so meant “assignment” in Persian, and *andar* (a generic “inside” in Urdu) was an old preposition mostly used in poetry in Persian. To my knowledge, there has been no research examining in detail these diverging meanings and if and how they came from a shared root. It is not ridiculous to surmise that Persianate elites solely using high register Persian resulted in Urdu’s Persian loanwords having nearly exclusively prestige associations in Iran. The import of the Classical Persian *yaar* (meaning “companion”) is used nearly sarcastically in everyday Hindustani in Mumbai, akin to “Really, dude?” (Dasgupta, “Arre Yaar, Churidaar Are Now in Oxford Lexicon”), whereas in Iran *yaar* is a relic of older poetry. Such departures in denotation and connotation are rife with history and abundant sources of future research.

Kuczkiewicz-Fraś’s opening statement for a chapter in *Perso-Arabic Loanwords in Hindustani. Part II. Linguistic Study* was the following: “For a linguist, South Asia remains both a paradise and a hell.” This sentence precisely reflects my sentiment, too. There is a profusion of linguistic phenomena to study in the region, and much of it is fraught with ideology and inextricable from incredibly specific history, contributing to messy data. This is, of course, why it’s all the more precious.

7. Appendix

7.1 Stimuli

List 1. Types for vowel analysis.

<i>bagh</i> (garden)	<i>kam</i> (less)
<i>jaan</i> (dear)	<i>asman</i> (sky)
<i>zindabad</i> (long live...!)	<i>aram</i> (calm)
<i>shahzadi</i> (princess)	<i>bas</i> (enough)
<i>vatn</i> (country)	<i>yaar</i> (friend)
<i>khat</i> (letter in the mail; calligraphy)	<i>rang</i> (color)
<i>zabaan</i> (tongue)	<i>jawab</i> (answer)
<i>bazaar</i> (open-air market)	<i>panka</i> (fan)
<i>zyaada</i> (more)	<i>namaaz</i> (Muslim prayer)
<i>haal</i> (condition)	<i>hamesha</i> (always)
<i>ya</i> (or)	<i>pyaaz</i> (onion)
<i>nazar</i> (evil eye, gaze)	<i>pardah</i> (curtains)
<i>shakl</i> (shape)	<i>sheher</i> (city)
<i>kagad</i> (paper)	<i>awaz</i> (voice)
<i>namak</i> (salt)	<i>nashta</i> (breakfast; time before first meal of the day)
<i>saadaa</i> (plain)	<i>khetem</i> (end)
<i>tabiyaat</i> (health condition; nature)	<i>anar</i> (pomegranate)
<i>saal</i> (year)	<i>agar</i> (if)
<i>taraf</i> (side)	<i>baraf</i> (ice/snow)

Formatted as such:

(*CP word* (Urdu or shared English translation; Persian translation, if different))

List 2. Types for consonant analysis.

<i>g garmi</i> (heat)	<i>g ganda</i> (dirty; foul)
<i>k kursi</i> (chair; heated ottoman on the ground)	<i>g gunah</i> (sin/fault)
<i>k kamar</i> (waist/back)	<i>k kinara</i> (banks of a river)
<i>y gazal</i> (type of poem)	<i>g gula</i> (throat)
<i>k kishmish</i> (raisin)	<i>q kalam</i> (pen)
<i>x kharch</i> (expense)	<i>y gam</i> (sorrow)
<i>g guftgu</i> (chit-chat)	<i>x khala</i> (mother's sister)
<i>y galti</i> (mistake)	<i>g gosht</i> (meat)
<i>x khud</i> (reflexive self)	<i>k kofta</i> (spiced meatballs)
<i>k kitab</i> (book)	<i>x khandan</i> (family line)
<i>k kami</i> (deficiency)	<i>q katra</i> (drop of liquid)
<i>q qanun</i> (law)	<i>y gayab</i> (vanished; absent)
<i>x khali</i> (empty)	<i>y ghareeb</i> (poor)
<i>x khargosh</i> (rabbit)	<i>q kainchi</i> (scissors)
<i>y ghusl</i> (ablutions)	<i>q kismat</i> (fate)

Formatted as such:

CP segment *CP word* (Urdu or shared English translation; Persian translation, if different)

7.2 Eliminated Stimuli

List 1 Removals: *Types that were removed from the initial body*

fikr (thought)	kameez
sabzi (green, vegetable)	gulaab
khoshki (dryness)	zakat (charity)
zameen (floor)	kamar (waist)
ki (that, relativizer)	diwaar (wall)
sabzii (green, vegetable)	shiad (maybe)
karna (to do)	saanf (clean)
kharab (spoiled)	yasmeen (jasmine)
baarish? (rain)	qasai (butcher)
badtameez (bad manners)	zeera
qanun (law)	shakar
verna (otherwise)	qainchi (scissors)
darzi (tailor)	raushanii (light)
kharch (to spend)	shorba
dukan (shop)	qabūl kardān (to accept)
bistr (bed)	hum/ham (we)
biryani (meat-rice dish)	asman (heaven)
diwaan (type of seating)	siyahi (ink)
jaldi (quickly)	meherbani (kindness, favor)
lekin (but)	yaani (meaning, i.e.)
angur (grape)	ghoom (lost)
sirf (only)	Khush-buu (scent)
kofta (meatballs)	kaafi (enough)
dana (a crumb, speck, fleck)	khareed (to buy)
istimaal (to use)	chatri (umbrella)
insaan (person)	gosht
khuda hafiz	pajama
zindagii (life)	badshah (king)
falooda	admi (a man)
saaf (clean)	istimaal (use)
nazdeeq (close)	kitab (book)
divai (medicine)	akl (mind)
hona (to be, to happen)	zukaam
tareeq (date)	safed (white)
khirki (window)	tasfeer/tasviir (photo)
saboon (soap)	khud (reflexive)
garam (hot)	umr (age)
intezaar (to anticipate)	takut (strength/power)
khush (happy)	jadoogar (magician)

zaroor (truly)	khoobsurat (beautiful)
nakhoon (nail)	qurbani (sacrifice)
shokria (thank you)	shaam (evening)
chadar (sheet)	andar (inside)
ready (taiyar)	pir (sufi saint)
kursi (chair)	achaar (pickle)
qalam (pen)	chehera
kishmish (raisin)	topi (hat)
magar (but, however)	nishaan (a marking)
dhuniya (universe)	muhabbat (adoration)
bimaar (sick)	takleef (difficulty/inconvenience)
dervaaaza (door)	ishq (love)

Formatted as such: *CP word* (Urdu or shared translation; Persian if different)

The reason for elimination is color-coded according to the following:

- Already used in L2
- used differently by speakers within and/or between languages
- included sounds outside the current scope of the paper
- difficult for speakers to produce
- duplicate of a word earlier in the list

List 2 Removals: *Types that were removed from the initial body*

y gussa (angry)
x khair (wellbeing, well then, in that case)
q kassab (butcher)
q kulfi (ice cream)
y ghairat (shame)
k kurti
g gurdah (kidney)
k karam (generosity, kindness, grace, favor)
q kameez (shirt)
g gada (beggar)
y gaur (deep thought, deliberation)
g gulab (rose)
q kahva (coffee)
x khalas (done, finished)
x kharab (spoiled)

CP segment *CP word* (Urdu or shared English translation; Persian if different)

The reason for elimination is color-coded according to the following:

- used differently by speakers within and/or between languages
- difficult for speakers to produce
- accidentally removed

7.3 Tables

Table 1. Persian /a/ Normalized Formant Values Summary (Barks)

	F1	F2	F3
Minimum	6.824	2.245	1.586
1 st Quartile	7.641	3.249	3.675
Median	8.150	3.796	4.225
Mean	8.212	3.926	4.286
3 rd Quartile	8.646	4.309	4.998
Maximum	10.125	7.839	6.239

Table 2. Persian /a:/ Normalized Formant Values Summary (Barks)

	F1	F2	F3
Minimum	6.832	2.047	1.061
1 st Quartile	8.578	5.709	2.115
Median	9.248	6.307	2.646
Mean	9.099	6.280	2.819
3 rd Quartile	9.589	7.062	3.389
Maximum	10.521	8.367	6.208

Table 3. Urdu /a/ Normalized Formant Values Summary (Barks)

	F1	F2	F3
Minimum	7.089	2.270	1.237
1 st Quartile	8.084	3.345	3.455
Median	8.693	4.231	4.306
Mean	8.613	4.239	4.374
3 rd Quartile	9.114	5.063	5.242
Maximum	10.049	6.771	6.777

Table 4. Urdu /a:/ Normalized Formant Values Summary (Barks)

	F1	F2	F3
Minimum	6.841	2.869	1.306
1 st Quartile	7.674	4.451	2.384
Median	8.042	5.180	2.913
Mean	8.131	5.040	3.091
3 rd Quartile	8.597	5.832	3.519
Maximum	9.835	6.872	6.585

Works Cited

- Abbas, Mirza Mohd Ejaz. "The Role of Persian Language before Independence in India." *International Journal of Pure and Applied Researches*, vol. 1, no. 2, 15 May 2016.
- Alam, Muzaffar. "The Pursuit of Persian: Language in Mughal Politics." *Modern Asian Studies*, vol. 32, no. 2, 1998, pp. 317–349. *JSTOR*, www.jstor.org/stable/313001.
- Ansarin, Ali Akbar. "An acoustic analysis of modern Persian vowels." *9th Conference Speech and Computer*. 2004.
- Bashir, Elena. "Urdu and Linguistics: A Fraught but evolving relationship." *Annual of Urdu Studies*; 2011, Vol. 26, p97.
- Billah, Abu Musa Mohammad Arif. "Linguistic influence of Persian on South Asian languages: Special reference to Urdu." (2018). *South Asian Cultural Studies*.
- Brown, E. K., and Anne Anderson. *Encyclopedia of Language & Linguistics*. 2nd ed., Elsevier, 2006.
- Dasgupta, Priyanka. "Arre Yaar, Churidaar Are Now in Oxford Lexicon." *The Times of India*, India News, 25 June 2015, timesofindia.indiatimes.com/india/Arre-yaar-churidaar-are-now-in-Oxford-lexicon/articleshow/47821906.cms.
- Everaert, Christine. *Tracing the boundaries between Hindi and Urdu: Lost and added in translation between 20th century short stories*. Vol. 32. Brill, 2010.
- Fatima, Nakhat, and Rabea Aden. "Vowel Structure of Urdu." *Akhbar-e-Urdu*, Apr. 2003.
- Hakala, Walter N. *Negotiating Languages: Urdu, Hindi, and the Definition of Modern South Asia*. Columbia University Press, 2016. *JSTOR*, www.jstor.org/stable/10.7312/haka17830.
- Khan, Masud Husain. "Urdu." *Linguistics in South Asia*, edited by Murray B. Emeneau and Charles A. Ferguson. De Gruyter, 2016, pp 277-284.
- Khansir, Ali Akbar, and Nasrin Mozafari. "The Impact of Persian Language on Indian Languages." *Theory & Practice in Language Studies* 4.11 (2014).
- Kuczkiewicz-Fraś, Agnieszka. *Perso-Arabic Loanwords in Hindustani. Part I. Dictionary*. Księgarnia Akademicka, 2008.
- Kuczkiewicz-Fraś, Agnieszka. *Perso-Arabic Loanwords in Hindustani. Part II. Linguistic Study*. Księgarnia Akademicka, 2012.

- Majidi, Mohammad Reza, and Elmar Ternes. "Persian (Farsi)." *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*, 2nd ed., vol. 17, International Phonetic Association, 1999.
- Mir, Raza. "Voh Yar Hai jo Khusbu ki Tarah/Jis ki Zuban Urdu ki Tarah": The Friendly Association Between Urdu Poetry and Hindi Film Music." (2000).
- Mohammadi, Hiwa, et al. "Formant structure and vowel space in Persian vowels." *Bimonthly Audiology-Tehran University of Medical Sciences* 20.2 (2011): 79-85.
- Ohala, Manjari. "Hindi." *Journal of the International Phonetic Association* 24.1 (1994): 35-38.
- Prakāśam, Vennelakaṇṭi. *Encyclopaedia of the Linguistic Sciences: Issues and Theories*. 1st ed., Allied, 2008.
- Rees, Daniel A. *Towards Proto-Persian: An optimality theoretic historical reconstruction*. Diss. Georgetown University, 2008.
- Rosen, Friedrich, and Nāṣir al-Dīn Shāh . *Modern Persian Colloquial Grammar: Containing a Short Grammar, Dialogues and Extracts from Nasir-Eddin Shah's Diaries, Tales, Etc., and a Vocabulary*. Luzac, 1898. *WorldCat* <https://catalog.hathitrust.org/api/volumes/oclc/2657427.html>. Accessed 2 Nov. 2019.
- Schmidt, Ruth Laila. "Chapter Eight: Urdu." *The Indo-Aryan Languages*, edited by Dhanesh Jain and George Cardona. Routledge, 2007, pp 288-344. *WorldCat*, <https://www-taylorfrancis-com.ccl.idm.oclc.org/books/9780203945315>. Accessed 29 Oct. 2019.