Loanword adaptation in Persian: A Core-Periphery model approach

Koorosh Ariyaee
University of Toronto

This study discusses the adaptation of loanwords in Persian from donor languages of Arabic, French and English. The main claim is that loanwords adapt to the host language phonology based on the duration of the time they have been used in that language (Kemmer 2017). Thus, I propose that Arabic loans which are borrowed earlier are more nativized than recent loans, like Russian, French and English words. Compared to the former studies (Shademan 2002; Perry 2005, 2011; among others), this is the first account that aims to show the hierarchical and gradual nativization process of loans in Persian by use of Core-Periphery model (Itô and Mester 1999). This model categorizes words from core stratum to periphery stratum based on the satisfaction of constraints in the host language as systematic comparison criteria of older loans with recent ones. In this study I show that the Persian lexicon is stratified into three strata: core, middle and periphery. This lexicon stratification reflects the gradual nativization of loanwords where the core stratum includes frequent native items, the middle stratum includes older loans (Arabic) and the periphery stratum includes recent loans (Russian, French and English). Furthermore, this paper shows that highly frequent Arabic loans satisfying all constraints have become completely nativized and thus are categorized as part of the core showing the great influence of the donor language.

1   Introduction

This study with the purpose of comparing the adaptation process of loanwords addresses how these lexical items from different source languages such as French, Russian, English and Arabic are adapted to the phonology of Persian (Farsi), a Southwestern Iranian language (Windfuhr 2009). The main postulation is that borrowings adapt to the host phonology based on the duration of the time that they have been borrowed and used in the host language. Hence, it is expected that Arabic loans show more adaptation to Persian phonology than Russian, French and English borrowings (hereafter European loans) since Arabic loanwords were borrowed into Persian after the Arab conquest of Persia (Iran) ca 651 AD (Morony 1986) whereas the European loanwords are recent borrowings which entered Persian in the 19th century (Kłagisz 2013: 39; Deyhime 2000). This assumption is in line with Kemmer's (2017) claim who posits the longer a loanword has been frequently used in the borrowing language, the more it resembles the native words of the language in terms of phonology. In addition, this gradual nativization process is an assertion of Kiparsky (1969) who posits that loans show different graded degrees of conventionalization. Thus, in order to capture this hierarchical nativization process of loanwords in Persian, I adopt the Core-Periphery model (Itô and Mester 1999) which is proposed to capture the stratified character of the lexicon by comparing the adaptation pattern of early-borrowed loans (Arabic items) with the recent borrowings (European words) in Persian.

   In order to present systematic comparison criteria of older loans with recent ones and stratify these items from core to periphery, in the Core-Periphery model, which is in the framework of Optimality Theory

(Prince and Smolensky 1993/2004), certain constraints that are specific to the phonology of the host language are introduced. Based on the satisfaction of these constraints, lexical items are stratified into different strata. The core stratum includes native words which satisfy constraints maximally. The other strata, as we move outwards the core, are assimilated loans and unassimilated foreign items forming the periphery. Moving toward the outer strata, we see more violation of constraints, until we see words at the periphery that satisfy a small subset of constraints. It will be shown that Persian includes three strata: core, middle and periphery. This lexicon stratification reflects the gradual nativization of loanwords where the middle stratum includes older loans (Arabic words) and the periphery stratum includes recent loans (European items). Furthermore, this paper shows that certain Arabic loans satisfying all constraints have become completely nativized and are categorized as part of the core showing the great influence of the donor language.

This stratification of the lexicon by comparing lexical items from different sources which has not been attested in the former accounts (Shademan 2002; Perry 2005, 2011; Paraskiewicz 2015; Deyhime 2012, among others) shows that loanword adaptation is not random; rather, this is a systematic process which is reflected in the hierarchical character of lexicon stratification.

1.1 Roadmap

In the following section the theoretical framework which is the Core-Periphery model will be introduced and discussed. Section 3 which addresses the theoretical analysis discusses the application of this model on Persian data. Finally, section 4 concludes the paper.

2 Theoretical framework: the Core-Periphery model

This section discusses how loanwords would be stratified in terms of their nativization to the phonology of the host language via the Core-Periphery model proposed by Itô and Mester (1999). In addition, the purpose of presenting this model will be addressed.

Itô and Mester (1999), following Kiparsky (1969), state that one of the reasons of presenting the Core-Periphery model is that a flat stratification of the lexicon of a language does not capture the gradual and hierarchical conventionalization of loanwords in a language. As displayed in (1), Itô and Mester (1999) refer to Kiparsky (1969) who argues that lexical items do not come neatly packaged into groups labeled either [+foreign] or [–foreign]. So, instead of partitioning foreign words into parallel and disjoint [+foreign] and [–foreign] sublexica, loans should be stratified based on a hierarchy of foreignness.

(1)
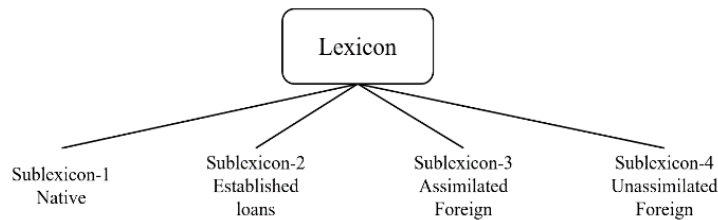
(Itô and Mester 1999: 64)

In order to capture this hierarchical structure of the lexicon, they propose their model of loanword adaptation in which based on host language constraints satisfaction, lexical items and subsequently donor languages are divided into different strata. In their model which investigates loanwords in Japanese, the core stratum of this model is native lexical items (Yamato). The other strata, as we move outwards the core, are Sino-Japanese, Assimilated foreign, and unassimilated foreign language lexical items, respectively. In this stratification, as displayed in (2), the core lexical items (Lex^0) fulfill markedness constraints maximally.

And the more we move outward toward the outer layer from the core, we see more violation of these markedness constraints.

(2)

Lex^{max}

Lex^2 "Unassimilated foreign"

Lex^1 "Assimilated foreign"

Lex^0 "Established loans"
(= "Native")
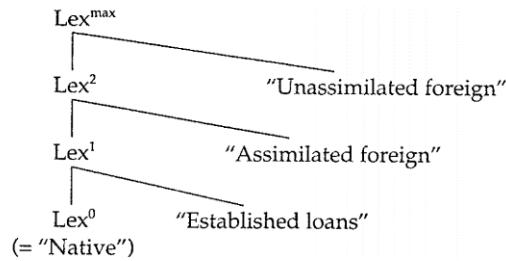
(Itô and Mester 1999: 65)

The structural organization of the lexicon is set inclusion, leading from innermost lexical core (Lex^0) to the most inclusive set (Lex^{max}) which comprises all lexical items. The above structures, displayed in (2), are built out of a network of implicational relations shown in (3): items that are subject to constraint A are always subject to constraint B, but not all items that are subject to B are subject to A. In this way, A would be a constraint with a more restricted domain than B (A's domain is properly included in B's domain).

(3)

B

z

y

A

x
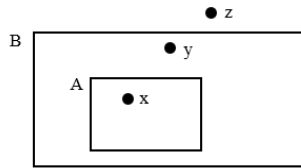
(Itô and Mester 1999: 65)

As shown above, *x* which is a native word and part of the core stratum is in the domain of A and of B; hence, it is impossible for an item to be in A's domain without being in the domain of B. *y* is in the domain of B but not of A. *z* is neither in the domain of A nor in the domain of B.

If we consider the same pattern for lexical items and constraints, then the relations of core-periphery will make sense. *x* which is in the core and a native word satisfies both A and B constraints. *y* which is an assimilated loanword only satisfies B and violates A indicating that it is not completely nativized. *z* which is an unassimilated foreign item by violating both A and B, shows that it is not nativized and, at the cost of violating the markedness constraints of the host language, is satisfying the faithfulness constraints of the source language.

Given the core-periphery relations above, Itô and Mester present their four markedness constraints for Japanese, shown in (4).

(4)
    a.   SYLLSTRUCTURE: Japanese syllable structure constraints (*COMPLEX, etc.)
    b.   NOVOICEDGEMM (NO-DD):   no voiced obstruent geminates (*dd, *gg, etc.)
    c.   NOVOICELESSLAB (NO-P):   no singleton-p: a constraint against nongeminate [p]
    d.   NONASAL⌢VOICELESS (NO-NT): postnasal obstruents must be voiced (*nt, *mp)

Only native items forming the core stratum satisfy all of these constraints. The more we move outwards from the core, the more violation of these constraints is observed until we reach the unassimilated foreign items at the periphery which satisfy one or none of these markedness constraints.

The abovementioned interaction of words from different source languages is shown in (5) exhibiting the argument ranking and the violation of these constraints by items from each stratum.
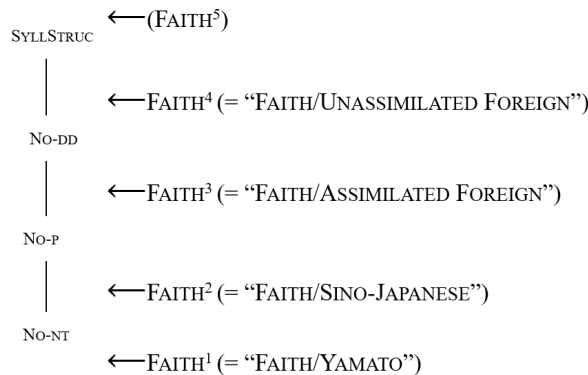
(5)

	SYLLSTRUC	NO-DD	NO-P	NO-NT
Yamato (Core)	✓	✓	✓	✓
Sino-Japanese	✓	✓	✓	violated
Assimilated foreign words	✓	✓	violated	violated
unassimilated foreign words (Periphery)	✓	violated	violated	violated

The tableau in (5) reveals that, as expected, native items satisfy all native Japanese markedness constraints maximally. The Sino-Japanese words form the closest stratum to the core since they satisfy all markedness constraints except the low-ranked constraint NO-NT. The assimilated foreign words form the penultimate stratum as they violate more constraints than Sino-Japanese words and less than the unassimilated foreign words. Finally, foreign words are stratified as items of the periphery stratum as they satisfy only the top-ranked markedness constraint and violate the rest of the constraints. In this way, the core-periphery relation in Japanese is justified.

The hierarchy of the above constraints clarifies the point that this lexical stratification comes about through different faithfulness constraints rankings within this fixed hierarchy of markedness constraints. As shown in (6), there are five positions where faithfulness constraints marked as FAITH$_1$ through FAITH$_5$ are located.

(6)

$$\text{SYLLSTRUC} \longleftarrow (\text{FAITH}^5)$$

$$\longleftarrow \text{FAITH}^4 (= \text{"FAITH/UNASSIMILATED FOREIGN"})$$

$$\text{NO-DD}$$

$$\longleftarrow \text{FAITH}^3 (= \text{"FAITH/ASSIMILATED FOREIGN"})$$

$$\text{NO-P}$$

$$\longleftarrow \text{FAITH}^2 (= \text{"FAITH/SINO-JAPANESE"})$$

$$\text{NO-NT}$$

$$\longleftarrow \text{FAITH}^1 (= \text{"FAITH/YAMATO"})$$

(Itô and Mester 1999: 72)

The faithfulness constraint of the core stratum ranks below all the other strata's markedness constraints indicating that it cannot interfere with the demands of these markedness constraints. In this hierarchy, the closer to the periphery, the more satisfaction of the source language constraints is observed at the cost of violation of the markedness constraints of the host language. That is why words at the periphery satisfy the source language faithfulness constraints (almost) maximally and satisfy the host language markedness constraints at minimum. In fact, the faithfulness constraint of the unassimilated foreign word stratum (FAITH/UNASSIMILATED FOREIGN) is only subordinate to the top-ranked markedness constraint of the host language (SYLLSTRUC). Respectively, the faithfulness constraint of the Assimilated foreign word stratum (FAITH/ASSIMILATED FOREIGN) is subordinate to NO-DD constraint. Likewise, the FAITH/SINO-JAPANESE constraint subordinates NO-P constraint. Finally, the FAITH/YAMATO constraint subordinates the low-ranking NO-NT constraint. From this argument ranking it can be deduced that the highest-ranked faithfulness constraints represent the most peripheral lexical items which have not been assimilated. And the lowest-ranked faithfulness constraints represent the most core vocabulary, the native

Japanese. Thus, when a host language markedness violation is preferred over a source language faithfulness violation, we have "periphery behavior", and when a source language faithfulness violation is preferred over a host language markedness violation, we have the "core behavior". Given this, the overall argument ranking of these constraints would be as follows in (7).

(7)

SYLLSTRUC >> FAITH/UNASSIMILATED FOREIGN >> NO-DD >> FAITH/ASSIMILATED FOREIGN >> NO-P >> FAITH/SINO-JAPANESE >> NO-NT >> FAITH/YAMATO

This argument ranking reduces what seems to be a random nativization process to a model with a systematic set of markedness constraints and indexed faithfulness constraints stratified and interleaved at different points in different layers (Itô and Mester 1999: 76).

In the following section the application of this model on recent and older loanwords will be addressed. By applying the model and introducing native Persian markedness constraints in this model, it will be shown that older loans (Arabic) satisfy more constraints than recent ones indicating Arabic items are more nativized than non-Arabic loans in Persian. Thus, Arabic items are expected to form the stratum that is closer to the core than non-Arabic ones. It will be shown that the majority of Arabic loans pattern as expected; however, there are certain Arabic loans that show complete nativization since they satisfy all native Persian markedness constraints at the cost of violating source language faithfulness constraints. This core behavior suggests that these Arabic loans are so nativized that they are now part of the core stratum of the Persian lexicon. On the other hand, I will show that there are certain native Persian items that are not part of the core stratum as they do not satisfy all Persian markedness constraints. This different patterning of native words and Arabic loans will be addressed in more detail in the following section.

3 Persian and the Core-Periphery model

In this section, in order to show the gradual nativization process of loans in Persian, the interaction of recent and older loans with native Persian markedness constraints within the framework of the Core-Periphery model is addressed. But before that, some brief information about Persian vowel and consonant inventory will be displayed in the following subsection. After this, the markedness constraints will be introduced individually. This section concludes with the full argument ranking and the stratification of items from core to the periphery strata based on these constraints.

3.1 Persian vowel and consonant inventory

According to Majidi and Ternes (1999: 124) in the Handbook of IPA, the vowel and consonant chart of Persian is as shown in (8) and (9), respectively.

(8)
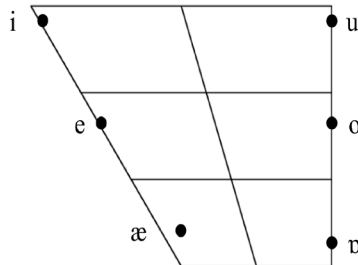
(Majidi and Ternes 1999: 124)

Note that the [+low, +back] vowel they introduce in Persian inventory is [+round] whereas others, including Hodge (1957), Modaressi (1979), Jahangiri (1980), Sadeghi (2001), Rohany (2012), among others, do not consider this vowel as [+round] and use [–round, +back] vowel [ɑ] in their studies on Persian; hence, in order to be consistent with the major body of literature in this regard, [ɑ] is used in this paper.[1]

(9)

	bilabial	labiodental	dental	alveolar	post alveolar	palatal	velar	glottal
plosive	p b			t d			k g	ʔ
nasal	m			n				
fricative		f v		s z	ʃ ʒ		x ɣ	h
affricate					t͡ʃ d͡ʒ			
trill				r[2]				
approximant						j		
lateral approximant				l				

(Majidi and Ternes 1999: 124)

In the following subsection each markedness constraint and its interaction with words from different source languages, including native Persian words, is addressed. In the end, the final argument ranking within the framework of the Core-Periphery model will be discussed.

3.2   Persian lexicon stratification

The lexicon of each language displays a certain degree of internal stratification of lexical items. These strata reveal a synchronic impact in that they reflect a general partitioning of the whole lexicon into different strata (Itô and Mester 1999). The lexical items of each stratum behave similar to each other based on certain criteria within the grammar of the language. In other words, this stratification helps us understand how the classified vocabularies interact with each constraint of a language, and it is based on this interaction that the lexical items can be categorized in a structured and organized system. Hence, in order to show this structured classification of the Persian lexicon, I introduce certain Persian markedness constraints, then show what words from what sources would either satisfy or violate these constraints. This interaction between words and constraints results in a systematic classification of items revealing their hierarchical degrees of nativization.

Given this explanation, in this study, I propose that the Persian lexicon is divided into three different strata. The core stratum consists of words which are native Persian lexical items. This stratum contains words that are frequently used in colloquial speech of Persian speakers. As mentioned earlier, there are certain loanwords in Persian that are borrowed from Arabic which have been nativized and conventionalized. These loans are also frequently used in Persian daily, colloquial speech. These nativized Arabic loans which have become part of the Persian lexicon are part of this core stratum. This stratum can be entitled the "Spoken Persian-Arabic". The closest stratum to the core, as we move outwards toward the periphery, are elevated Persian words. This stratum consists of native words that are not used in colloquial speech. These words are Persian words that are used in a highly formal register. In addition, this stratum consists of Arabic borrowings which are not (or rarely) used in daily spoken Persian but are used in formal or literary contexts. This stratum is entitled the "Written Persian-Arabic" stratum. The words of this stratum are less nativized in comparison to core stratum items. So, "Written Persian-Arabic" words (the middle stratum) violate a markedness constraint that is satisfied by lexical items of the core stratum which is expected since as we move outwards toward the periphery, more markedness constraints will be violated; as a result, "Written Persian-Arabic" stratum satisfies more markedness constraints than the most peripheral stratum which will be introduced below.

---

[1] It is beyond the scope of this study to investigate the true nature of this vowel. For further information in this regard, see Jones (2019).

[2] In Persian /r/ varies with [ɾ] and [ɹ] (Majidi and Ternes 1999: 125).

The third stratum contains the unassimilated foreign words from different languages, except Arabic. For the purposes of this study, loanwords from English and French and Russian have been chosen. Following Itô and Mester (1999), these words will be entitled the unassimilated foreign words. These borrowings are the least nativized words which exist in Persian lexicon; in other words, the loanwords of this stratum violate Persian markedness constraints more than the words of the two aforementioned strata. In the following subsection this stratification based on the markedness constraints is discussed.

Prior to the lexicon stratification, the constraints and their interaction with the native words and loans are addressed in the following subsection.

3.3   Persian markedness constraints

The abovementioned core-periphery stratification is formed based on the violation of three markedness constraints in the phonology of Persian. These constraints are displayed in (10) based on their ranking.

(10)

a.	PERSIANPHONOLOGY	cover constraint for a set of loanword adaptation constraints (see below)
b.	*CCC#	no triple consonant clusters at the ends of the words
c.	*[ɑN]	no [ɑ] preceding any nasal consonant

PERSIANPHONOLOGY is a cover term for a set of constraints such as ONSET (10b and 10d), *COMPLEXONSET (13), among others. These constraints are satisfied by all loans regardless of their source or duration of time being used in Persian justifying why it is the top-ranked constraint. In addition, based on the PERSIANPHONOLOGY constraint, each loanword that is borrowed into Persian must adapt to the phonological inventory of this language. If a segment is shared both in the source and the host languages, then that segment surfaces faithfully, like /m/ or /n/ whose inputs and outputs are the same. But if a segment of the source language does not exist in the host language, then that segment undergoes some repairs so that it could map to the phonology of the borrowing language; for instance, there is no pharyngeal segment in the inventory of Persian (see 9); hence, /ħ/ loses its pharyngeal feature and surfaces as [h] when borrowed into Persian, as in (11a).

Note that regardless of the source of a loanword, similar segments of different words from different source languages adapt to Persian phonology the same as each other, and thus in the output they surface as the same segments; for instance, a segment like the velarized /l/, which does not exist in the phonology of Persian, surfaces as non-velarized /l/, no matter whether the word with the velarized /l/ is borrowed from Arabic or English. Examples in (11) will illustrate more.

(11)

Arabic	→	**Persian**	
a. /ħalːaːd͡ʒ/	→	[hælɑd͡ʒ]	'male name'
b. /alːaːh/	→	[ʔælɑh]	'Allah'
c. /maðˤlum/	→	[mæzlum]	'oppressed'

English	→	**Persian**	
d. /iːmeɪɫ/	→	[ʔimejl]	'email'
e. /fʊɫ/	→	[ful]	'full'

Regarding the above examples, I should highlight a few points. First, gemination is hardly tolerated in Persian; that is why /lː/ in (11b) surfaces as [l] in Persian. Second, as Persian does not have any dental fricative segment, the segment /ðˤ/ changes into [z] when it is borrowed into Persian. This indicates that Persian does not keep the pharyngealized feature of the segment either, as mentioned before. This language preserves the voicing feature of the segment, as it is contrastive, but sacrifices the [+distributed] feature and changes it into [–distributed]. The reason could be that as Persian phonology does not have any dental

segment and considering that all dental consonants are [+distributed], this feature sounds redundant and thus disappears. Furthermore, in Persian each syllable must have an onset, and if a loanword which is borrowed into Persian is onsetless, this language repairs this onsetless syllable by inserting the [ʔ] segment at the beginning of the syllable to satisfy the ONSET constraint (Prince and Smolensky 1993/2004), as in (11b) and (11d).

One other characteristic of Persian phonology is that this language does not tolerate complex onsets. So, loanwords that have complex onsets undergo certain repairs in order to fit into Persian phonology and thus satisfy the *COMPLEXONSET (Golston 1996).

There are different strategies such as deletion and epenthesis to satisfy *COMPLEXONSET. Persian prefers the latter as the repair over the former. This type of repair is in line with what Kang (2011), quoting from Paradis and LaCharité (1997), refers to as the Preservation Principle, which dictates that the input material be preserved as much as possible, unless the cost of preservation is too extreme and thus epenthesis should generally be the preferred type of repair. Cross-linguistically, as for word-initial onset clusters, a survey of available cases shows that epenthesis is the predominant choice of repair in languages like Burmese, Egyptian Arabic, Persian, Fijian, Fula, Hindi (Kang 2011).

Persian uses three different epenthetic vowels of [i], [e] and [o] in different phonological environments. The following examples in (12) will illustrate more.

(12)

    Epenthetic [e]
a. /pɹaɪslɪst/ → [peɾɑjslist] 'pricelist'
b. /plæn/ → [pelæn] 'plan'
c. /plʌs/ → [pelɑs] 'plus'

    Epenthetic [i]
d. /fri:z/ → [fiɾiz] ~ [feɾiz] 'freeze'
e. /tɹɪp/ → [tiɾip] 'trip'

    Epethetic [o]
f. /bʁɔʃyʁ/ → [boɾoʃuɾ] 'brochure'
g. /pʁɔʒɛ/ → [poɾoʒe] 'project, plan'
h. /blɔ̃d/ → [bolond] ~ [belond] 'blond'

Based on the above data, it seems that the elsewhere condition epenthetic vowel is [e]. But the epenthetic vowel assimilates with its following vowel in certain environments. For instance, if that vowel is either /ɪ/ or /i/, then the epenthetic vowel assimilates with it and thus it is [i]. Or the epenthetic vowel is [o] if the first vowel of the loanword is /ɔ/ (and probably /o/); in other words, regarding the epenthetic vowel we can propose the following generalization shown in (13).

(13)

    a. Ø → i / ___ Ci
    b. Ø → o / ___ Co/ɔ
    c. Ø → e / ___ Ce/ɑ/æ

Further evidence to suggest that the elsewhere epenthetic vowel could be [e] is that in loanwords shown in (14) the epenthetic vowel is [e], no matter what the first vowel of the loanword is. The only difference between loanwords in (14) and ones in (13) is that the epenthetic vowel in (14) appears before the onset consonant cluster (at the left edge of the word), and then the default epenthetic segment [ʔ] inserts word-initially to satisfy the ONSET constraint.[3]

---

[3] Note that Arabic does not have complex onsets, so this constraint is already satisfied in the donor language.

(14)

    a.   /spɔːrt/   →   [ʔespoɾt]   'sport'
    b.   /spaɪdərmæn/ →   [ʔespɑjdeɾmæn]'Spiderman'

Kambuziya et al. (2010) argue that based on the sonority hierarchy principle, the insertion position of the epenthetic vowel will change. If the sonority of the complex onset is falling, the vowel will be inserted at the left edge of the syllable, before the cluster, as in (14). If the sonority of the onset cluster is rising, the vowel will be inserted between the two consonants, as in (12). Conversely, Gouskova (2001: 177) argues that epenthesis is not related to the sonority hierarchy principle within the syllable; rather, it is a matter related to the sonority hierarchy principle at the syllable boundary. She posits that sonority must not rise across syllable boundary. This topic of debate is beyond the scope of this study.

The above-mentioned phonological requirements for the mapping of segments from various source languages to the Persian inventory and the syllable-related constraints such as ONSET and *COMPLEXONSET are top-ranked and inviolable; that is why the cover constraint for this set of constraints is top-ranked indicating all words of the three Persian lexicon strata satisfy this constraint. Given that, the OT tableau in (15) shows that all Persian lexicon strata satisfy this constraint.

(15)

	PERSIANPHONOLOGY
Spoken Persian-Arabic (core)	✓
Written Persian-Arabic (middle)	✓
unassimilated foreign words (periphery)	✓

From (15) it can be inferred that based on the PERSIANPHONOLOGY constraint there is no stratification of lexical items in terms of being assimilated to Persian phonology, as all words satisfy all the Persian phonology constraints. However, the markedness constraint *CCC#, which bans clusters with triple consonants word-finally, separates the periphery stratum from the middle and core strata.

The reason that the *CCC# constraint separates European items from native Persian and Arabic loans is that unlike Persian and Arabic which ban triple consonants at the ends of words in their phonological systems, European items allow such clusters with three consonants. The following data in (16) are loans from English and French which violate the *CCC# constraint.

(16)

    a.   /wɑːlts/   →   [vɑlts]   'waltz'
    b.   /hɜːrts/   →   [heɾts]   'hertz'
    c.   /ɔʀkɛstʀ/   →   [ʔoɾkestɾ]   'orchestra'
    d.   /desãbʀ/   →   [desɑmbɾ]   'December'

When a word with a three-consonant word-final cluster is borrowed from either English or French into Persian, that word violates the host language markedness constraint at the cost of satisfying the source language faithfulness constraint which is the periphery behavior, as in the periphery stratum it is the source language faithfulness that dominates the host language markedness constraints.

Given this explanation, the *CCC# constraint is dominated by the top-ranked constraint PERSIANPHONOLOGY which is satisfied by all the strata. This argument ranking is shown in (17).

(17)

	PERSIANPHONOLOGY	*CCC#
Spoken Persian-Arabic (core)	✓	✓
Written Persian-Arabic (middle)	✓	✓
unassimilated foreign words (periphery)	✓	violated

The above argument ranking only classifies the lexicon into two strata of the unassimilated foreign words and the rest of the items. In order to separate the core from the middle stratum the low-ranked markedness *[ɑN] constraint is introduced.

This constraint bans the sequence of [ɑN][4] and triggers the raising of the pre-nasal /ɑ/ to [u] in this environment. This pre-nasal vowel raising operation is only common among the core Persian words. Hence, only native Persian words and Arabic loans that are used frequently in spoken Persian undergo this alternation. The following examples in (18) illustrate more.[5]

(18)

UNDERLYING FORM			SURFACE FORM	SOURCE LANGUAGE	GLOSS
ɑrɑm			ʔɑrum	Persian	peaceful
xɑne			xune	Persian	home
bɑdɑm			bɑdum	Persian	almond
penhɑn			penhun	Persian	hidden
nɑn			nun	Persian	bread
musalma:n	→	mosælman	mosælmun	Arabic	Moslem
tama:m	→	tæmam	tæmum	Arabic	end
ama:nat	→	æmanæt	ʔæmunæt	Arabic	deposit
ħajwa:n	→	hejvan	hejvun	Arabic	animal
ħara:m	→	hæram	hærum	Arabic	haram

Unlike lexical items in the core stratum (spoken Persian-Arabic), words at the middle stratum (written Persian-Arabic) do not undergo the pre-nasal vowel raising alternation since the *a ~ u* raising is an operation that occurs only in words with high frequency in spoken Persian and even if a word used in written Persian appears in spoken Persian (in a formal register), the alternation does not apply on that lexical item. The following data in (19) show words that are used in written Persian and hence resist raising.

(19)

UNDERLYING FORM			SURFACE FORM		SOURCE LANGUAGE	GLOSS
rɑjgɑn			rɑjgɑn,	*rɑjgun	Persian	free
legɑm			legɑm,	*legum	Persian	harness
bɑstɑn			bɑstɑn,	*bɑstun	Persian	ancient ages
hengɑm			hengam,	*hengum	Persian	time
burha:n	→	borhan	borhan,	*borhun	Arabic	reason
dˤima:m	→	zemam	zemam,	*zemum	Arabic	harness
tˤaʕa:m	→	tæʔam	tæʔam,	*tæʔum	Arabic	food
istiħma:m	→	estehmam	ʔestehmam,	*ʔestehmum	Arabic	bathing
d͡ʒa:nib	→	d͡ʒaneb	d͡ʒaneb,	*d͡ʒuneb	Arabic	direction

---

[4] *N = m* and *n*

[5] It is beyond the scope of this study to address the the phonological, morphological and lexical factors which block the occurrence of pre-nasal vowel raising in spoken Persian (see Ariyaee 2018).

The data in (19) prove that this operation separates core stratum (native Persian and Arabic loans in spoken Persian) from the middle stratum (native Persian and Arabic loans in written Persian). By the same token, European loans with the /ɑN/ sequence do not allow the raising of /ɑ/ to [u], regardless of whether they are highly frequent words in spoken Persian or not. The following data in (20) show some loans from French, English and Russian that are frequently used in spoken Persian without undergoing the alternation.

(20)

UNDERLYING FORM			SURFACE FORM		SOURCE LANGUAGE	GLOSS
ʀɛstɔʀɑ̃	→	restorɑn	restorɑn,	*restorun	French	restaurant
nɔvã̃bʀ	→	novɑmbr	novɑmbr,	*novumbr	French	November
ʃã̃s	→	ʃɑns	ʃɑns,	*ʃuns	French	chance
kamjɔ̃	→	kamjon	kamjon,	*kumjon	French	truck
ɑːnɬaɪn	→	ɑnlɑjn	ʔanlɑjn,	*ʔunlɑjn	English	online
mɑːnɪtər	→	manitor	manitor,	*munitor	English	monitor
ɑːnkaɬ	→	ɑnkal	ʔankal,	*ʔunkal	English	on-call
ædəmz	→	adams	ʔadams,	*ʔadums	English	gum
stækan	→	estekan	ʔestekan,	*ʔestekun	Russian	teacup
livan	→	livɑn	livɑn,	*livun	Russian	glass
van	→	vɑn	vɑn,	*vun	Russian	bathtub

The blockage of raising in frequent European loans and the occurrence of the raising in frequent Arabic loans can justify the gradual loanword nativization process. In other words, older loans (Arabic words) by satisfying the native Persian markedness constraint *[ɑN] show that they are more nativized than recent ones (French, Russian and English items). This comparison shows that the long duration of the time loans have been used in the host language and the high frequency of use are both required for a loan to undergo the pre-nasal vowel raising operation. The following figure in (21) shows that from among the majority of the frequent words used in spoken Persian 97% of native Persian items undergo the alternation, and 76% of frequent Arabic items used in spoken Persian show the alternation whereas only 4% of frequent European loans used in spoken Persian undergo the process of pre-nasal vowel raising.[6]

(21)
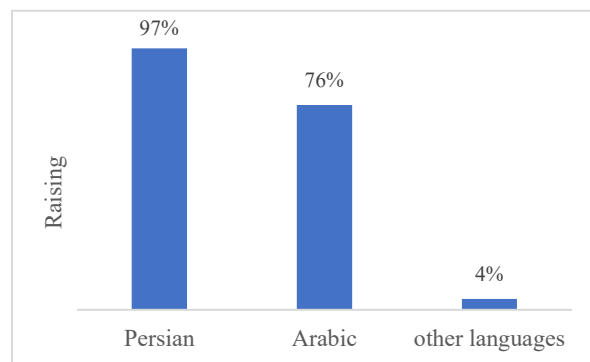
Pre-nasal raising in frequent items in Persian based on source language

---

[6] For the discussion of exceptional pre-nasal vowel raising and blocking, see Ariyaee (to appear).

The above figure, again, proves that Arabic loans are more nativized than non-Arabic borrowings which correctly reflects the gradual and hierarchical character of loanword nativization in host languages; older loans are more conventionalized than recent ones.

Given that the markedness constraint *[ɑN] is satisfied by the items in the core stratum only and is violated by the items of the periphery and middle strata, it is revealed that this is the low-ranked constraint. As a result, the finalized OT tableau of the Persian lexicon will be as follows in (22).

(22)

	PERSIANPHONOLOGY	*CCC#	*[ɑN]
Spoken Persian-Arabic (core)	✓	✓	✓
Written Persian-Arabic (middle)	✓	✓	violated
unassimilated foreign words (periphery)	✓	violated	violated

The above argument ranking is restated in (23) which is based on the set inclusion relationship proposed by Itô and Mester (1999), shown earlier in (3). Items in core stratum that are subject to constraint A, which is the *[ɑN] constraint in this study, are always subject to constraint B, which is the *CCC# constraint. But middle stratum items that are subject to B are not subject to A. In this way, A would be a constraint with a more restricted domain than B (A's domain is properly included in B's domain). Hence, it is impossible for Spoken Persian-Arabic words (Core) to be in A's domain without being in the domain of B. Finally, items of the periphery stratum are neither subject to constraint A nor to B. That said, the core-periphery relations can be justified in Persian.

(23)
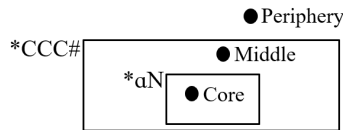
The argument ranking in (22) and the set inclusion relationship between the constraints and the stratified items in (23) reflect the gradual and systematic stratification of loanwords based on their degrees of nativization in Persian. This hierarchical system indicates that conventionalization of loans is structured, not random or haphazard.

4 Conclusion

This study, as the first account that compares older loans (Arabic items) with recent ones (Russian, French and English words) through the application of the Core-Periphery model (Itô and Mester 1999) showed that older loans are more nativized than recent ones highlighting the hierarchical character of Persian loanword nativization suggesting that there are different hierarchical degrees of nativization among loans in the lexicon of the host language. In order to capture this systematic hierarchy, this nativization stratification is accounted for by certain markedness constraints of the host language. The loans that are more nativized satisfy more host language markedness constraints as they are more assimilated with host language phonology. This outcome is in line with Kiparsky (1969) who argues that a flat stratification of loans does not show the gradual and hierarchical character of the lexicon; rather, there are different degrees of nativization and conventionalization among foreign words.

Besides showing the systematic and layered strata of the Persian lexicon which indicates that words do not nativize in a random way, another outcome of this study was that there are Arabic loans that have become completely nativized since they satisfy all markedness constraints of native Persian and thus are part of the core stratum. These Arabic loans are ones that are frequently used in spoken Persian. On the other hand, there are certain native Persian items that are part of the middle stratum, and not the core, since these items do not satisfy all native Persian markedness constraints (they violate one constraint). These

native items are used in written Persian and are infrequent in spoken Persian, unless used in a very formal register. This shows that Arabic is so nativized that it has permeated the core stratum of Persian, and native speakers prefer to use these Arabic loans in their daily speech, instead of formal native Persian items of the middle stratum. This outcome demonstrates the extensive influence of the donor language on the host language lexicon which has not been investigated systematically in Persian so far.

Acknowledgement

References

Algeo, John. 2009. *The origins and development of the English language*. Boston: Wadsworth publications.

Anttila, Arto, and Young-mee Yu Cho. 1998. Variation and change in Optimality Theory. *Lingua 104*(1–2): 31–56.

Ariyaee, Koorosh. 2018. Pre-nasal vowel raising in Tehrani Persian. Master's thesis, University of Toronto.

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

Boersma, Paul, and Clara Levelt. 2000. Gradual constraint-ranking learning algorithm predicts acquisition order. In *Proceedings of Child Language Research Forum* 30, 229–237. Stanford, CA: CSLI Publications.

Collins French Dictionary. Retrieved at https://www.collinsdictionary.com/dictionary/english-french.

Deyhime, Guity. 2000. France xvi. loanwords in Persian. In *Encyclopaedia Iranica* X/2, 181–184.

Farazandeh-pour, Farzane and Aliye Kord Zafaranlu Kambuziya. 2014. German loanwords adaptation in Persian: Optimality approach. In *International Journal of Humanities 20*(4): 23–40.

Golston, Chris. 1996. Direct optimality theory: Representation as pure markedness. *Language*, *72*(4): 713–748.

Gouskova, Maria. 2001. Falling sonority onsets, loanwords, and syllable contact. In *Proceedings of the Annual Meeting of the Chicago Linguistic Society (CLS) 37*(1): 175–185.

Hyman, Larry. 1970. The role of borrowing in the justification of phonological grammars. *Studies in African Linguistics 1*(1): 1–47.

Majidi, Mohammad Reza and Elmar Ternes. 1999. Persian (Farsi). In *Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet*, ed. Amalia Arvaniti, 124–125. Cambridge University Press.

Itô, Junko, and Armin Mester. 1999. The phonological lexicon. In *The handbook of Japanese linguistics*, ed. Natsuko Tsujimura, 62–100. Blackwell Publishers, Malden, Massachusetts, and Oxford, U.K.

Jones, Taylor. 2019. A corpus phonetic study of contemporary Persian vowels in casual speech. *University of Pennsylvania Working Papers in Linguistics 25*(1): 127–137.

Kang, Yoonjung. 2011. Loanword phonology. In *The Blackwell companion to phonology: general issues and segmental phonology volume 1*, ed. Marc van Oostendorp, 1–25. Malden, MA: Wiley-Blackwell.

Kemmer, Suzanne. 2017. Retrieved at http://www.ruf.rice.edu/~kemmer/Words/loanwords.html.

Kiparsky, Paul. 1969. Linguistic universals and linguistic change. In *Universals in linguistic theory*, ed. E. Bach and R. Harms, 170–210. New York, Holt.

Kord Zafaranlu Kambuziya, Aliye and Eftekhar Sadat Hashemi. 2010. Russian loanword adaptation in Persian; Optimal approach. In *Iranian Journal of Applied Language Studies 3*(1): 77–96.

Kłagisz, Mateusz. 2013. Hints on French loanwords in modern new Persian. In *Romanica Cracoviensia 13*(1): 38–51.

Morony, Michael G. 1986. Arab ii. Arab conquest of Iran. In *Encyclopaedia Iranica*, II/2, 203–210.

Paraskiewicz, Kinga. 2015. English loanwords in new Persian. In *Folia Orientalia*, *52*(2): 282–292.

Paradis, Carole, and Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. In *Journal of Linguistics 33*(2): 379–430.

Perry, John. R. 2005. Lexical areas and semantic fields of Arabic loanwords in Persian and beyond. In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic, and Turkic*, ed. Éva Ágnes Csató, Bo Isaksson, and Carina Jahani, 97–110. London; New York: Routledge.

Prince, Allen, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004.

Shademan, Shabnam. 2002. Epenthetic vowel harmony in Farsi. Master's thesis. University of California at Los Angeles.