PHONETIC RENDERINGS IN TURKISH ARABISMS AND FARSISMS

Kamil STACHOWSKI[*]

Abstract: *The paper is a summary of phonetic renderings found in borrowings from Arabic and Persian in Turkish. Based on 1748 loanwords, it gives an overview of which adaptations are typical in both groups, and which are unusual. For the latter, the specific cases are listed and briefly discussed. The focus is on renderings of individual phonemes rather than processes or the influence of the phonetic surrounding.*

Keywords: *Turkish, Arabism, Farsism, loanword adaptation, phonetics*

Arapça ve Farsça'dan Türkçe'ye Alınan Kelimelerdeki Fonetik Uyarlanma Değişmeleri

Öz: *Bu çalışmada, Arapça ve Farsça'dan Türkçe'ye alınan kelimelerde geçen fonetik uyarlanma değişmeleri toparlanıp tartışılır. 1748 alıntıya dayanarak, her iki grupta da hangi uyarlamaların tipik ve hangilerinin olağandışı olduğuna dair genel bir bakış sunar. İkincisi için, özel durumlar listelenir ve kısaca tartışılır. Odak noktası, süreçler veya fonetik çevrenin etkisinden ziyade bireysel ses birimlerinin tercümeleri üzerinedir.*

Anahtar Kelimeler: *Türkçe, Arabizm, Farsizm, Ödünç Kelime Uyarlaması, Fonetik*

1   Introduction

There exist a number of works devoted to the phonetic adaptation of Turkish Arabism and Farsisms, but they tend to be based on a limited number of examples, and do not necessarily provide a reliable picture of which renderings are usual, and which ones exceptional. General works in this field include dictionaries of borrowings (ALOT, PLOT, Pomorska 2013, Rocchi 2016–2017), as well as etymological dictionaries (GTS, KEWT, NS, TDES, TETTL). Approaches focusing on phonetics range from more comprehensive (İşler 2008, K. Stachowski 2015, M. Stachowski 2012a) to rather detailed (Al-Hashmi 2016, Baran 2019, Oytun Altun 2012, Öztekten 2001 and 2013, M. Stachowski 2012b), as well as those which adopt a different perspective (Erdem 2013, Korkmaz 2007, Sağ 2019) or include material beyond just Arabisms and Farsisms (Karaca 2012, Özkan 2011). Since the beginning of the 21st century, however, the interest appears to have somewhat shifted away from Turkish and towards other Turkic languages (Akın 2011, Cumakunova 2008, Karasoy 1998, Oytun Altun 2015, Pekaçar 2006 and 2007, Sarıkaya 2005, Yazıcı Ersoy 2006).

The presentation in this paper is organized by Arabic and Persian phonemes, and consists of two parts: a table detailing the number of examples which exhibit the different renderings of the given phoneme, and a list of the specific words in which unusual renderings can be observed, followed by a brief commentary. The lists are generally in the same order as the tables, which is alphabetical.

---

[*] PhD, Jagiellonian University, Chair of General and Indo-European Linguistics. Cracow / POLAND. E-mail: kamil.stachowski@gmail.com ORCID ID: 0000-0002-5909-035X

with ‚Arabic and Persian words are transcribed using the DIN/DMG transcription ة the exception for which is transcribed ‹ŧ›. Where the Arabic and Persian systems differ, the Arabic variant is given precedence, so e.g. ذ is spelt ‹ḏ› rather than ‹ẓ›. Turkish words are given in the official orthography, but with several non-standard additions. **Small caps** on voiced consonants denote 'positionally voiced' consonants, i.e. a consonant that is voiced in the presence of a vowel-initial suffix but devoiced when in auslaut, as e.g. in *cilD, -di*. **Stress mark** before a vowel denotes a vowel that is stressed despite not being in the final syllable, as in *bˈelki*. **Acute** marks 'functionally front' consonants, i.e. consonants adjacent to back vowels which nonetheless act as if they were front, such as the ones in *biĺĺahi* or *saaí, -ti*. **Grave** denotes vowels that are short in the nominative but lengthened when a vowel-initial suffix is added, as in *zamàn, -ānı*. **Subscribed zero** denotes those vowels which are present in the nominative but are deleted in the presence of a vowel-initial suffix, like e.g. *ili$_0$m, -lmi*. **Subscribed two**, analogously, denotes consonants that are single in the nominative and doubled when a vowel-initial suffix is added, as in *hak$_2$, -kkı*. Lastly, *ʔ* denotes a glottal stop, such as in *Kur'an*.

2 Renderings

Data have been collected from the newest etymological dictionary of Turkish which is KEWT. It contains close to 6000 entries in total; of this number, 1234 are clear Arabisms and 514 clear Farsisms. This is after discarding words with alternative phonetic variants (e.g. T *ateş* < P *ātaš ~ ātiš*), only attested in compounds (e.g. *ahdetmek*), or otherwise unclear or raising doubts about their phonetic adaptation. One or two misprints have been corrected here but are not marked. KEWT only gives Turkish words in the contemporary orthography, so phonetic details needed to be supplemented from GTS and TRT. The several words that were missing from both have been omitted. The extraction of individual adaptations, of words which exhibit them, preparation of the contingency tables, etc., have been performed in *R* using the *soundcorrs* package (K. Stachowski 2020).

The tables take a mechanistic approach, providing only phoneme-to-phoneme correspondences. The term *phoneme*, however, must be understood loosely in this case. Arabic and Turkish are essentially treated phonologically but the transcription of Persian words primarily reflects the orthography, distinguishing e.g. between *t* ت and *ţ* ط. There are two reasons for this. Firstly, at least in some cases the Turkish rendering, especially that of vowels, may have been influenced by Persian spelling. Secondly, and this is a general remark, I chose to distinguish more 'phonemes' rather than fewer. In case of doubts about this decision, the appropriate columns in the tables below can be simply summed up. The reverse operation would not have been possible. Exceptions have been made for geminates – which are treated as separate 'phonemes', and for *k*, *kk*, *g*, *l*, and *ll* where an alternative phonetic description has been also given beside a phonological one.

The phoneme-to-phoneme approach means the focus is on the effect of adaptation rather than its mechanism or motivation. It should be noted that focusing tables strictly on the effect does not render them useless. Indeed, a simple algorithm, based exclusively on the data from the tables, correctly classifies 88% of words included here, as either Arabisms or Farsisms (K. Stachowski [forthcoming]). Below the tables, the specific words that exhibit renderings which are found in up to seven examples, are singled out, and where this seemed reasonable, the phonetic surrounding is presented in

a broader context.

At the end of each subsection, if only the given 'phoneme' has more than one rendering in Turkish, I mention the entropy.[1] This is a measure notorious for its multiple interpretations; in this case, it is perhaps best read as an index of unpredictability of the outcome of adaptation. The range is from 0 to 1, where 0 means that the Turkish rendering can be predicted with absolute certainty, and 1 that such prediction is entirely impossible. Perhaps the results for vowels have proven the most interesting.

The tables below presents the number of times which all the individual renderings appear in our set of 1234+514=1748 words, i.e. they represent occurrences, not words. For example, A ʿumūm > T umum counts as two, not just one case of A m > T m.

2.1 Ø

	a	e	G	h	ι	ι₀	i	i₀	t	u	u₀	ü	ü₀	y
Arabisms	1	2	1	0	8	18	3	42	0	2	5	1	6	2
Farsisms	2	2	0	1	0	0	6	3	1	1	0	0	2	2

Insertions of vowels occur primarily in consonant clusters in auslaut. These are slightly more frequent in Persian than in Arabic (8 vs 6% of words), but Persian clusters tend to be preserved while Arabic ones not. The reasons for this different treatment are not clear to me. The specific unsual cases are: insertion[2] of *a*: A *qadr* > T *kadar*, and P *lāfzan* > *lafazan*, *mušt* > *muşta*; ¶ insertion of *e*: A *mustamlak* > T *müstemleke*, *balsam* > *pelesenk*, and P *rišta* > T *erişte*, *nārgīl* > *nargile*; ¶ insertion of *i*: A *laḥm* > T *lehim*, *rahn* > *rehin*, *saṭl* > *sitil*, and P *āškār* > T *aşikar*, *āšnā* > *aşina*, *baḫtjār* > *bahtiyar*, *ḫudāvandgār* > *hüdavendigar*, *jādgār* > *yadigar*, *siḥrbāz* > *sihirbaz*; ¶ insertion of *i₀*: P *pādzahr* > T *panzehir*, *zahr* > *zehir*, *šahr* > *şehir*; ¶ insertion of *u*: A *ḥawḍ* > T *havuz*, *šurb* > *şurub*; and P *toḥm* > T *tohum*; ¶ insertion of *u₀*: A *lutf* > T *lütuf*, *nuṭq* > *nutuk*, *quṭb* > *kutup*, *quṭr* > *kutur*, *ufq* > *ufuk*; ¶ insertion of *ü*: A *quds* > T *kudüs*; ¶ insertion of *ü₀*: A *ğurm* > T *cürüm*, *ḥukm* > *hüküm*, *ḥusn* > *hüsün*, *kufr* > *küfür*, *šukr* > *şükür*, *zulm* > *zulüm*, and P *muhr* > T *mühür*, *zulf* > *zulüf*.

The one case of insertion of *G* is in A *balsam* > T *pelesenk*. GTS derives the word from A *belesān* without explanation; KEWT supposes the influence of such words as *tüfe(n)k*, *cenk*, and *renk*; NS proposes P *balasān* without explanation; TDES does not contain *pelesenk*. KEWT's proposition may be correct, though perhaps *pelesenk* 'an unnecessarily repeated word' (~ *persenk* < P *pārsang*) might have been an even greater influence that led to the addition of *-k*.

The single case of inserted *h* is in P *āvang* > T *hevenk*. The adaptation *ā > e* is also uncommon (cf. Subsection 2.4 below). A more understandable rendering would have been **āvenG*, as in P *āhang* > T *ahenk*.

---

[1] Specifically, normalized Shannon entropy. Given the counts of renderings $X=\{x_1, \dots x_n\}$,

$$H(X) = -\frac{1}{\log_e n} \cdot \sum p(x_i) \cdot \log_e p(x_i)$$

[2] *Insertion* is used here as an umbrella term. For example, in the case of *a*, two words exhibit an epenthesis, and one a paragoge; I see little benefit in the explicit naming of the process for every example, especially when it would have to be at the cost of legibility of the presentation. See also Footnote 3.

The insertion of **t** is in P *pāsbān* > T *pazvant*. See TDES for references, and note that this word is also the sole example for P *b* > T *v* (cf. Subsection 2.5 below).

The four cases of insertion of **y** are in A *anāī* > T *enayi*, *qīmaï* > *kıymet*, and P *panīr* > T *peynir*, *zīrak* > *zeyrek*. Diphthongisation is quite an unusual adaptation for a long vowel, and even more so for a short one (in *peynir*), see the appropriate subsections below. The reasons for it are not clear to me.

Overall, the insertions in Arabisms have the entropy of 0.7 while those in Farsisms 0.92, meaning the latter are a little more diversified and therefore less predictable.

2.2 ʾ ʿ

	∅	ʔ	v	y
Arabisms	69	2	1	5
Farsisms	0	0	0	0

The unusual cases here are: with *ʔ*: A *ḥamāʾil* > T *hamail*, *Qurʾān* > *Kurʾan*; ¶ with *v*: A *māʾī* > T *māvi*; ¶ with *y*: A *ʿaǧāʾib* > T *acayip*, *ḫalāʾiq* > *halayık*, *lāʾiq* > *layık*, *qadāʾif* > *kadayıf*, *sāʾis* > *seyis*. To an extent, renderings with *ʔ* and *y* are mixed in the actual usage, as in careful speech [ažajip] may be replaced by [ažaʔip], and in less careful pronunciation [χamaʔil] by [χamajil], etc. In this paper, the normative advice of GTS has been followed.

The entropy in Arabisms is 0.31, and in Farsims, naturally, 0.

2.3 *a*

	∅	a	ʿa	à	ā	e	ʿe	ē	ı	i	o	ö	u
Arabisms	7	434	7	3	34	676	3	4	4	3	1	1	2
Farsisms	1	89	1	1	4	309	10	0	4	4	0	0	0

The cases of **deletion**[3] are: A *ḍīq an-nafas* > T *tıknefes* (cf. Subsection 2.10), *ʿuqqijaï* > *okka*, *in šāʾa Allāh* > *inşallah*, *kajlaï* > *kile*, *māʿadā* > *mada*, *tabaʿaï* > *teba*, and P *girifta* > T *girift*. The latter is the only such rendering among more than a hundred Persian words in -*a*, including a dozen in -*ta*, and three in -*fta* (the other two being *hafta* > *hafta*, and *jāfta* > *yafta*).

The cases of uncommon stress are: with *ʿa*: A *ḍarṭaï* > T *zarta*, *falaqaï* > *falaka*, *ḫalīfaï* > *kalfa*, *masḫaraï* > *maskara*, *raġman* > *rağmen*, *šamʿaï* > *şama*, *ṭafraï* > *tafra*, and P *čagāla* > T *çağla*; with *ʿe*: A *albatta* > *elbette*, *ġabran* > *cebren*, *walḫāṣil* > *velhasıl*.

Lengthening occured in a relatively large number of words, uncommon among which are: in *à*: A *adabīǰaï* > T *edebiyat*, *saʿj* > *say*, *šaʾn* > *şan*, and P *feġan* > *figan*; ¶ in *ā*: P *labbalab* > T *lebalep*, *načar* > *naçar*, *taʿlīmḫāna* > *talimhane*; ¶ in *ē*: A *maʾḍūn* > T *mezun*, *maʾmur* > *memur*, *mabdaʾ* > *mebde*, *taʾsīs* > *tesis*. The specificity of the last group is not that the *aʾ* sequence yielded a long vowel, but that this vowel is *ē* rather than *ā*; long *ā* has been the outcome in eight times as many words.

---

[3] Like *insertion*, *deletion* is used here as an umbrella term for aphaeresis, syncope, and apocope. See Footnote 2.

Cases of raising resulting in *ı*: A *ḫaṣm* > T *hısım*, *manǧanīq* > *mancınık*, *manqabaẗ* > *menkıbe*, *minṭaqaẗ* > *mıntıka*, and P *čādar* > *çadır*, *čanār* > *çınar*, *ḫadiv* > *hıdiv*, *tarāš* > *tıraş*; ¶ and in *i*: A *fatīl* > T *fitil*, *raġā* > *rica*, *saṭl* > *sitil*, and P *anǧīr* > *incir*, *pahlavān* > *pehlivan*, *taftīk* > *tiftik*, *tarīt* > *tirit*.

Lastly, rounding yielded *o* in A *mawlā* > T *molla* (cf. Footnote 4), ¶ *ö* in A *tawbaẗ* > T *tövbe*; ¶ and *u* in A *baḫūr* > T *buhur*, and *maslūk* > *musluk*.

Fronting is overall considerably more common in Farsisms where it happened in 76% of words as opposed to 58% of Arabisms. It is also greatly more frequent than in the case of long *ā*. The diversification of the two groups is nevertheless nearly identical: entropy is 0.373 for Arabisms and 0.374 for Farsisms. These are low values for a vowel, and it may be surprising that they are lower than for long *ā*. (See Subsection 2.4, and also subsections for all the other vowels: *e* (2.13, 2.14), *i* (2.26, 2.27), *o* (2.38, 2.39), *u* (2.56, 2.57).)

2.4   *ā* ١

	∅	a	ʿa	à	ā	ʿā	e	ı	u
Arabisms	2	117	6	145	281	0	6	1	1
Farsisms	3	135	3	35	109	1	27	0	0

Cases of **deletion** include: A *ištihāʾ* > T *iştah*, *maʿārif* > *marif*, and P *čagāla* > T *çağla*, *rāzijāna* > *rezene*, *salāḫāna* > *salhane*. The Arabic sequence *Vʿā* can be found in nine words, and in all the remaining eight cases it yields T *aā*, *ua*, or *uā*.

Unusual place of stress can be seen in short *ʿa* in A *biṭānaẗ* > T *badana*, *ḫālaẗ* > *hala*, *maġāraẗ* > *mağara*, *qinnāraẗ* > *kanara*, *ṣandālijjaẗ* > *sandalye*, *ṭāʾbījaẗ* > *tabya*, and P *čārpāra* > T *çalpara*, *jāfta* > *yafta*, *nišāsta* > *nişasta*; ¶ and in long *ʿā* in P *šājed* > T *şayet*.

The remaining two unusual renderings are: six cases of *e*: A *ǧallābi* > T *çelebi*, *handāsaẗ* > *hendese*, *hiǧāʾ* > *hece*, *karbatān* > *kerpeten*, *kattān* > *keten*, *sāʾis* > *seyis*; ¶ one of *ı*: A *ġilāf* > T *kılıf*; ¶ and one of *u*: A *mihmāz* > T *mahmuz*.

Fronting is far less common than in the case of short *a*, as it only occurred in 9% of Farsisms and 1% of Arabisms. The diversification of both groups is fairly similar, the entropy of Arabisms being 0.56 and that of Farsisms 0.66. Strangely, this means that the renderings of long *ā* are less predictable than those of short *a* (see Subsection 2.3 above).

2.5   *b* ب

	b	ʙ	f	m	p	v
Arabisms	137	49	0	0	21	0
Farsisms	62	5	1	2	15	2

When *b* was in auslaut in the etymon, positional preservation of its voicedness (*ʙ*) has been considerably more common in Arabisms where this occurred in 49 out of 54 cases, as opposed to Farsisms where this only happened in five out of ten words: P *āb* > T *ab*, *dolāb* > *dolap*, *girdāb* > *girdap*, *māhtāb* > *maytap*, *sinǧāb* > *sincap*. In all five, the final *-b* was preceded by *ā*; the sequence *-āb* only occurs in two more words, P

čirkāb > T *çirkef*, *partāb* > T *pertav*, while in the remaining three cases where *-b* has been fully devoiced, it was preceded by a phoneme other than *ā* (P *čōb* > T *cop*, *labbalab* > *lebalep*, *turb* > *turp*).

All the other unusual renderings are also in Farsisms: *f* and *v* in *çirkef* and *pertav* mentioned above, as well as in *pāsbān* > *pazvant* (cf. Subsection 2.45 below); ¶ and *m* in *bandkaša* > *menteše*, *bīša* > *meše*.

Overall, Arabisms are less predictable than Farsisms, the entropies being 0.77 and 0.51, respectively.

2.6 *bb*

	b	*bb*	*p*
Arabisms	0	2	3
Farsisms	1	3	0

Here, all the examples fall under our 'definition' of unusual renderings, as none occurs more than seven times. Shortening to *b* happened in P *labbalab* > T *lebalep*; ¶ preservation of *bb* in A *muḥabbaï* > T *muhabbet*, *qubbaï* > *kubbe*, and P *ʿabbāsī* > T *abbasi*, *dabba* > *debbe*, *šabbūj* > *şebboy*; ¶ and rendering as *p* in A *ḥabb* > T *hap*, *šabb* > *şap* and, surprisingly, also in *qabbān* > *kapan*. Perhaps a contamination with *kapan* 'trap'?

Overall, the predictability of both groups is relatively similar and very poor, the entropy being 0.97 for Arabisms and 0.81 for Farsisms.

2.7 *č* چ

	c	*C*	*ç*
Arabisms	0	0	0
Farsisms	1	2	29

The only unusual cases here are: in *c*: P *čōb* > T *cop*, ¶ and in *C*: P *girač* > T *kireç*, *mārpič* > *marpuç*. Strangely, the latter group contains two out of three Persian words with final *-č* (the third one is P *čāč* > T *çeç*), making voicing appear to be the preferred adaptation. Based on just three words, of course, this would be a far-fetched conclusion.

The entropy for Arabisms is naturally 0, and for Farsisms 0.34.

2.8 *d* د

	Ø	*d*	*D*	*n*	*t*
Arabisms	0	122	39	0	14
Farsisms	1	97	11	1	16

Both special cases here are among Farsisms: **deletion** in *ğūdmard* > *cömert*; ¶ and rendering as *n* in *pādzahr* > *panzehir*. As for *cömert*, the sequence *-dm-* does not appear in any other Persian etymon, and in only one Arabic word in which it has been preserved: A *idmān* > T *idman*. See also Subsection 2.57 below as the behaviour of *ū* is

also unusual in this word. Regarding *panzehir*, GTS, KEWT, and NS agree on the etymon but do not provide any explanation for the unusual change; TDES does not discuss the word. NS adduces from Meniński's *Thesaurus* (1680): "*pādzehr* vulg. *panzeher*".

The entropy of Arabisms is overall quite higher than that of Farsisms: 0.72 vs 0.47, respectively.

2.9 *dd*

	D_2	*dd*
Arabisms	3	9
Farsisms	0	0

The three cases of D_2 are: A *ḍidd* > T *zıt*, *ğadd* > *cet*, *radd* > *ret*. These are the only three words where *dd* appears in auslaut.

The entropy for Arabisms is 0.81, and for Farsisms, of course, 0.

2.10 *ḍ* ض

	d	*s*	*t*	*z*
Arabisms	3	1	1	31
Farsisms	0	0	0	0

The three adaptation as *d* are: A *ʿaḍalaï* > T *adale*, *fuḍūl* > *fodul*, *qāḍī* > *kadı*; ¶ the one as *s* is *iḍṭirāb* > *ıstırap*; ¶ and the one as *t* is *ḍīq an-nafas* > *tıknefes*. The one case of *s* could effectively be counted as *z*, as it is the only etymon in which *ḍ* is adjacent to a voiceless consonant. This renders the *t* in *tıknefes* even more unusual, as indeed also are the other changes in its anlaut, and strengthens M. Stachowski's currently improvable but, it seems, very likely suspicion (p.c.) that the adaptation of this word was not phonetic but semantic in nature, and caused by the association of *dıkannefes 'shortness of breath' with *tıkamak* 'to choke, plug, clog'.

The entropy in Arabisms is 0.39, and in Farsisms 0.

2.11 *ḏ* ذ

	z
Arabisms	20
Farsism	1

The one time that *ḏ* appears in a Persian etymon is in *sarguḏašt* > *sergüzeşt*.

2.12 *ḏḏ*

	zz
Arabisms	1
Farsisms	0

The only word in this group is A *laḏḏaï* > T *lezzet*.

2.13 *e*

	e	*i*	*ī*	*ö*
Arabisms	0	0	0	0
Farsisms	15	3	1	1

The three cases of raising to *i* are P *feġan* > T *figan*, *ferdaws* > *firdevs*, *fešang* > *fişek*, the only three words with the sequence *fe* in any position.

The one case of *ī* is P *direġ* > T *diriğ*. It is one of two words with the sequence *eġ* (the other one being *figan* above), and one of a dozen Persian words with *Vġ* or *V̄ġ*. The only other case where this yielded a long vowel in Turkish is P *lāġar* > T *lagar*. The same sequences appear in thirteen Arabisms and is never rendered with a long vowel.

The rounding to *ö* is in P *šelān* > T *şölen*. The initial sequence *še-* does not appear in any other word; *ša-* is present in 32 Arabisms, but never rendered as *şö-*.

The entropy for Arabisms is, naturally, 0, and for Farsisms 0.58. This is the only vowel where the entropies for the short and long variant cannot be compared (see Subsection 2.14 below, and also Subsection 2.3 above).

2.14 *ē*

	e	*ı*
Arabisms	0	1
Farsisms	1	0

The one Farsism here is *tēz* > *tez*, and the one Arabism *zēh* > *zıh*. KEWT and TDES explain the presence of *ē* in an Arabic etymon by assigning it to the Syrian dialect. GTS, on the other hand, derives T *zıh* from P *zeh* 'catgut, cord, bowstring, …'. NS does not discuss the word. Data on the phonetic adaptation of borrowings from Syrian Arabic are not available to me; the adaptation of P *e* as T *ı* does not seem to be impossible, although the dataset used here does not contain any examples that could support it (see Subsection 2.13 above).

As only single adaptations are attested for both Arabisms and Farsisms, entropies cannot be calculated and compared to those of short *e*. This is the only such vowel (see Subsection 2.3 above).

2.15 *f* ف

	f	*f́*	*v*
Arabisms	165	1	0
Farsisms	27	0	1

The unusual adaptation as a functionally front *f́* is in A *ḥarf* > T *harf*, *-fî*. The other three words with the *-rf* sequence in auslaut are A *ʿurf* > T *örf*, *ṣarf* > *sarf*, *zarf* > *zarf*. ¶ The one rendering as *v* is in P *kafgīr* > T *kevgir*. The only other Farsism with *f* directly before a voiced consonant is *ṭarāfdār* > *taraftar*. Among Arabisms, *fl* and *fr* appear twice each, but never lead to voicing on Turkish ground.

The entropies are predictably low, 0.05 for Arabisms and 0.22 for Farsisms.

2.16 *ff*

	f	*ff*
Arabisms	1	3
Farsisms	0	0

The only case of shortening to *f* is in A *ṣaff* > T *saf*; ¶ the three preservations of *ff* are in A *ʿiffaẗ* > T *iffet, talaffuẓ > telaffuz, šaffāf > şeffaf*.

The entropy for Arabisms is 0.81, and for Farsisms 0.

2.17 *g* گ

	g	*G*	*ǵ*	*ğ*	*Ǧ*	*k*
Arabisms	0	0	0	0	0	0
Farsisms	27	5	3	5	2	12

The five words in *G* are: P *āhang* > T *ahenk, āvang > hevenk, ǧang > cenk, nārdang > nardenk, rang > renk*. ¶ There are only two more words with -*g* in auslaut, and they both yielded *Ǧ*: P *fešang* > T *fişek, čirāg > çırak*.

The cases of *ǵ* are P *āgāh* > T *agah, ḫudāvandgār > hüdavendigar, jādgār > yadigar*, against four words where *gā* did not result in *ǵ* (i.e. 'functionally front' g, see below; P *čagāla* > T *çağla, kārgāh > gergef, pargār > pergel, pājgāh > peyke*).

The five *ğ*'s are: P *agar* > T *eğer, čagāla > çağla, dīger > diğer, ǧigar > ciğer, magar > meğer*. There are six more words with the *ga ~ ge* sequence, and their Turkish renderings are always *ge* except in *gama > kama*.

Out of the dozen *k*'s, eight are in anlaut. This may seem meaningful at first, but there are in total 25 Farsisms with initial *g-* in the etymon, so this must be put down to chance. Similarly, no regularity could be found in the following vowel.

It must be emphasized that in this paper ‹*ǵ*› is only used for *g*'s that are palatal despite being adjacent to back vowels. Phonetically, in fact, P *g* never yields velar g in Turkish, only [k, ḱ, ǵ], and the two readings of ‹*ğ*› (i.e., all the 27 ‹*g*›'s in the table are in a front context). Overall, only three words have a velar rendering: P *čagāla* > T *çağla, čirāg > çırak, gama > kama*.

For Arabisms, the entropy is undecided. For Farsisms, it is quite high, 0.78 when counted phonologically, and a little lower, 0.67, when counted phonetically.

2.18 ʿع

	∅	*h*	*i*	*y*
Arabisms	179	3	2	1
Farsisms	3	1	0	0

The three cases of **deletion** from Persian words are: P *ʿabbāsī* > T *abbasi, šamʿdān > şamdan, taʿlīmḫāna > talimhane*. ¶ The four examples with *h* are: A *ʿajbaẗ* > T *heybe, ṭāliʿ > talih, saʿtar > zahter, ṭamaʿkār > tamahkar*. ¶ The one with *y* is A *ḍaʿīf* > T *zayıf*.

The entropy for Arabisms is 0.13, and for Farsisms 0.81.

2.19 *ġ* غ

	∅	g	ǧ	Ǧ	h	k
Arabisms	0	26	11	1	1	3
Farsisms	1	6	8	0	1	2

The only case of **deletion** is in P *čaġbut* > T *çaput* which is one of three Farsisms with the *ġC* sequence (the other two are *bāġča* > *bahçe*, and *jaġma* > *yağma*). This sequence is also present in six Arabisms, and always yields *ǧ* in Turkish.

The six examples with **g**: P *feġan* > T *figan*, *lāġar* > *lagar*, *pejġāmbar* > *peygamber*, *rovġan* > *rugan*, *sīmurġ* > *samurg*, *šalġam* > *şalgam*. ¶ The one word with *Ǧ* is A *mablaġ* > T *mablak*, one of five Arabisms, and five Farsisms, where -*ġ* is in auslaut (including eight where it is preceded by a vowel). ¶ The two cases of **h** are A *ġirār* > T *harar*, and P *bāġča* > T *bahçe*. ¶ Finally, the **k**'s are in A *ġilāf* > T *kılıf*, *ġurnūq* > *kırnak*, *ṣamġ* > *zamk*, and P *ġoza* > *koza*, *ġulāmbāra* > *kulampara*.

The entropy is 0.63 in Arabisms, and 0.8 in Farsisms.

2.20 *ǧ* ج

	c	C	ç	ş
Arabisms	103	11	3	1
Farsisms	18	4	5	0

Original voicing in auslaut is almost always preserved positionally (*c*): in eleven out of twelve cases among Arabisms (¶ the one exception with **ş** being *kalaǧ* > *keleş*), and in four out of five cases among Farsisms: A *bāǧ* > T *baç*, *biriṅ* > *pirinc*, *tāǧ* > *taç*, *turunǧ* > *turunç* versus only *ḫāǧ* > *haç*. ¶ The remaining examples where ǧ yielded *ç* are: A *ǧallābi* > T *çelebi*, *ḫanǧar* > *hançer*, *iǧtimāʿ* > *içtima*, and P *ǧāmašūr* > *çamaşır*, *ǧuvāl* > *çuval*, *ǧuvāldūz* > *çuvaldız*, *panǧa* > *pençe*.

Renderings in Arabisms are considerably more predictable, with entropy of 0.34 versus 0.79 in Farsisms.

2.21 *ǧǧ*

	c	cc
Arabisms	1	2
Farsisms	0	0

According to our 'definition' of up to seven cases, all three examples here count as unusual: in *c*: A *ḫaǧǧī* > T *hacı*; ¶ and in *cc*: A *saǧǧādaï* > *seccade*, *tuǧǧār* > *tüccar*.

The entropy is understandably high for Arabisms (0.92) and low (0) for Farsisms.

2.22 *h* ه

	∅	**f**	**h**	**y**
Arabisms	0	0	85	0
Farsisms	2	1	45	1

The two **deletions** are in P *pīhsuz* > T *pesüs*, *pājgāh* > *peyke*; ¶ the one case of *f* is in P *kārgāh* > T *gergef*; ¶ and the one of *y* in P *māhtāb* > T *maytap*.

The entropy of Farsisms is 0.27.

2.23 *ḥ* ح

	∅	**h**
Arabisms	1	154
Farsisms	0	5

The only case of **deletion** is A *ṣaḥiḥ* > T *sahi*, versus sixteen examples with final -*ḥ* > *h*; ¶ the five examples in borrowings from Persian are: *darḥāl* > T *derhal*, *ḥīlakār* > *hilekar*, *nāḥaq* > *nāhak*, *siḥrbāz* > *sihirbaz*, *silāḥšur* > *silahşor*.

The entropy of Arabisms is minimal, 0.06.

2.24 *ḥḥ*

	hh
Arabisms	1
Farsisms	0

The only word with a geminated *ḥḥ* in the entire dataset is A *ṣiḥḥat* > T *sıhhat*, adapted exactly as it could be expected based on the renderings of *ḥ* (see Subsection 2.23 above).

2.25 *ḫ* خ

	h	**k**
Arabisms	63	2
Farsisms	41	1

The three unusual cases in *k* are: A *ḫalīfaẗ* > T *kalfa*, *masḫaraẗ* > *maskara*, and P *baḫir* > *bakır*. As might be expected, the sequence *ḫa* appears in many more Arabisms, exactly 25 in total. Besides *kalfa* and *maskara*, the only cases where it is not rendered as *ha* in Turkish, are A *ḫalā* > T *hela*, *ḫandaq* > *hendek*, *ḫaṣm* > *hısım*. As for *bakır*, *ḫi* is not attested in any other borrowing from Persian, though I do not suspect that it could be the influence of *i* that caused this atypical rendering of *ḫ*. Rather, it is something to do with the history of both words as those are not in fact entirely clear.

In *maskara*, a contamination might be suspected. GTS and KEWT only cite the Arabic etymon, but NS divides our word into two: *maskara* 'clown' < A *masḫaraẗ*, and

maskara 'mascara' either < Engl./Fr./It. *mascara ~ maschera* – or also < A *masḫaraï*. TDES does not discuss the word. The variant with *k* likely only arose after 1680 (cf. K. Stachowski 2015: 301), so it seems probable that the word had been first borrowed from Arabic, with *h* and in the meaning 'clown', only later joined by a borrowing from one of the European languages, with *k* and in the meaning 'mascara', and finally, since the two words sound exceptionally similar and have in fact compatible semantics, they merged into *maskara* with two meanings.

Bakır is also not entirely clear; see KEWT and NS for remarks and references.

The adaptations in both Arabisms and Farsisms are very predictable, yielding entropy of 0.2 and 0.16, respectively.

2.26 *i*

	∅	a	e	ʾe	ɩ	ʾɩ	i	ʾi	i₀	ī	u	ü	y
Arabisms	4	13	6	0	65	1	337	1	1	13	0	2	2
Farsisms	1	0	4	1	6	0	53	2	0	0	2	3	0

The **deletions** are in A *ğuġrāfijaï* > T *coğrafya*, *kīmijā* > *kimya*, *ṣandālijjaï* > *sandalye*, *sīmijā* > *simya*, and P *zirih* > T *zırh*. As for the Arabic examples, there are twenty more with the *ij* sequence, and in all it has been rendered as *Vy* in Turkish.

The other atypical examples are: with *e*: A *ijālaï* > T *eyalet*, *hiğāʾ* > *hece*, *ḥisāb* > *hesap*, *maʿdin* > *maden*, *sijāḥaï* > *seyahat*, *taḏkiraï* > *tezkere*, and P *bikār* > T *bekar*, *bijābān* > *beyaban*, *girdak* > *gerdek*, *sihpāj* > *sehpa*; ¶ with *ʾe*: P *rāzijāna* > T *rezene*; ¶ with *ɩ*: P *baḫir* > T *bakır*, *čirāg* > *çırak*, *piḫtī* > *pıhtı*, *rīḫtim* > *rıhtım*, *tiğ* > *tığ*, *zirih* > *zırh*; ¶ with *ʾɩ*: A *ṭibq* > T *tıpkı*; ¶ with *ʾi*: A *in šāʾa Allāh* > T *inşallah*, and P *rišta* > T *erişte*, *zinhār* > *zinhar*; ¶ with *i₀*: *raḥim* > *rahim*; ¶ with *u*: P *mārpič* > T *marpuç*, *mihtar* > *muhtar*; ¶ with *ü*: A *mumkin* > *mümkün*, *muškil* > *müşkül*, and P *ğumbiš* > T *cümbüş*, *gūgird* > *kükürt*, *čūnki* > *çünkü*; ¶ and with *y*: A *fāʾidaï* > T *fayda*, *ṭāʾifaï* > *tayfa*. The only obvious regularity in the phonetic surrounding is in the last two examples; in total there are, however, sixteen words with the *Vʾi* sequence, and its renderings are: *āi* (eight cases), *ayı* and *ay* (two cases each), *āʔi*, *āyı*, *āyi*, *eyi* (one case each; cf. Subsection 2.2 above).

Backing occurs in 18% of Arabisms and 11% of Farsisms. Despite this, the renderings in Arabisms are generally a little more predictable than those in Farsisms, with entropies of 0.37 and 0.5, respectively. These are fairly low values for a vowel, and surpringly, lower than for long *ī* (see Subsection 2.27 below, and also Subsection 2.3 above).

2.27 *ī* ي

	∅	e	ʾe	ɩ	ʾɩ	i	ʾi	ì	ī	u	ü
Arabisms	2	0	1	10	1	90	0	25	68	1	1
Farsisms	1	5	1	9	0	39	2	3	19	3	1

The **deletions**: A *ḫalīfaï* > T *kalfa*, *ṭāʾbījaï* > *tabya*, and P *kulīča* > *külçe*. ¶ The renderings as *e*: P *bīša* > T *meşe*, *pīhsuz* > *pesüs*, *pīšīn* > *peşin*, *rīčāl* > *reçel*, *zīrak* > *zeyrek*; ¶ as unusually stressed *ʾe*: A *īj wa-llāh* > T *eyvallah*, and P *pīštaḫta* > T

peştahta; ¶ as unusually stressed ʾ*ı*: A *ḍīq an-nafas* > T *tıknefes* (cf. Subsection 2.10); ¶ as unusually stressed ʾ*i*: P *ankīšta* > T *enişte*, *mīnā* > *mine*; ¶ as positionally long *i*: P *temhīr* > *temhir*, *zamīn* > *zemin*, *zarrīn* > *zerrin*; ¶ rounding to *u*: A *ṣūfī* > T *sofu*, and P *dīvār* > T *duvar*, *durdī* > *tortu*, *turšī* > *turşu*; ¶ rounding to *ü*: A *kursī* > T *kürsü*, and P *dūrbīn* > T *dürbün*. I was not able to find any phonetic regularities in these examples.

Backing occurred in 6% of Arabisms and 14% of Farsisms, which is the opposite proportion to the adaptations of short *i*. The entropies are 0.59 for Arabisms and 0.69 for Farsisms, meaning the renderings of long *ī* are in fact less predictable those of short *i* (see Subsection 2.26 above).

2.28 *j* ي

	Ø	*i*	*y*
Arabisms	2	1	78
Farsisms	3	0	43

The cases of **deletion** are: A *ʿuqqijaẗ* > T *okka*, *ğajb* > *cep*, and P *darjā* > T *dere*, *rāzijāna* > *rezene*, *sihpāj* > *sehpa*. ¶ The one rendering as *i* is in A *kajlaẗ* > T *kile*. These five words, *cep*, *kile*, *okka*, *rezene*, and *sehpa*, are all the examples that there are for monophthongization of a falling diphthong wiht *j*.[4] The case of *okka* is not in fact entirely clear. It can be seen from the table in Subsection 2.26 above, that rendering A *i* as T *a* is not an especially rare phenomenon, and the position directly after *q* may perhaps be contributive to such an adaptation: not counting *okka*, it occurs in four out of twelve cases of *qi*, but cf. also A *aqriba* > T *akraba*, *ğirār* > *harar*. However, the deletion of -*ja*- (or -*ij*-?) is difficult to explain without appealing to irregular development due to frequency of use.

The entropy is 0.17 for Arabisms and 0.35 for Farsisms.

2.29 *jj*

	y	*yy*
Arabisms	15	4
Farsisms	0	0

The four cases of preservation of geminated *jj* are A *ajjām* > T *eyyam*, *muqajjad* > *mukayyet*, *sajjāḥ* > *seyyah*, *sajjāraẗ* > *seyyare*. In all four, *jj* is preceded by *a* or *ā* whereas in the fifteen cases where *jj* has been shortened, it was always preceded by *i* or *ī*. The following vowel is in all nineteen cases *a* or *ā*.

The entropy for Arabisms is 0.74.

2.30 *k* ك

	Ø	*g*	*ǧ*	*k*	*ḱ*
Arabisms	0	0	10	85	16

---

[4] There are two more where such a diphthong has been otherwise changed: A *sīmijā* > T *simya*, and *ṭāʾbījaẗ* > *tabya*. Falling diphthongs with *w* are only monophthongized in the sense that *w* is adapted as *v*, except in A *mawlā* > T *molla*.

Farsisms	3	1	13	49	12

Only four cases to be mentioned here: three **deletions** in P *ankīšta* > T *enişte*, *kašk* > *keş*, *bandkaša* > *menteşe*; ¶ and one voicing to **g** in P *kārgāh* > T *gergef*. In the last example, one could suspect the influence of *g*. There are three more etyma with *k* and *g* simultaneously; P *girdak* > T *gerdek* does not shed any light on the issue, but P *kangar* > T *kenger*, and *kafgīr* > *kevgir* suggest that there must have been another reason for the voicing in *gergef*.

It needs to be stressed that in this paper the opposition ‹k› : ‹ḱ› is phonological rather than phonetic, meaning that ‹ḱ› is only used for *k*'s which are palatal despite being adjacent to a back vowel. Phonetically speaking, renderings as velar *k* are rare, with just five cases among Arabisms (*dallāk* > *tellak*, *ḫirkaï* > *hırka*, *nafakaï* > *nafaka*, *maslak* > *maslak*, *maslūk* > *musluk*), and three among Farsisms (*čābuk* > *çabuk*, *čākū* > *çakı*, *zambak* > *zambak*).

Phonologically, the entropies are quite high for a consonant: 0.56 for Arabisms, and 0.62 for Farsisms. The phonetic approach considerably increases their predictability: 0.24 for Arabisms, 0.46 for Farsisms.

2.31 *kk*

	kk	*ḱḱ*
Arabisms	3	1
Farsisms	0	0

The four examples here are: in *kk*: A *murakkab* > T *mürekkep*, *sikkaï* > *sikke*, *takkaï* > *tekke*; ¶ and in *ḱḱ*: A *dukkān* > T *dükkan*. Note that similarly to single *k* (Subsection 2.30 above), ‹ḱ› is used here phonologically. Phonetically all the renderings are [ḱḱ].

Phonologically, the entropy for Arabisms is 0.81. Phonetically it is undecided as there is only one category.

2.32 *l* ل

	l	*ĺ*
Arabisms	210	95
Farsisms	56	10

Here, no group of renderings counts as unusual under the 'definition' given at the beginning of this section. It needs to be noted, however, that despite the first impression given by the table, both Arabic and Persian *l* is much more commonly rendered as palatal *l* in Turkish. As was explained in Section 1, the symbol ‹ĺ› is only used here for *l* that is palatal despite being adjacent to a back vowel, so ‹l› covers both velar *l*'s, as well as palatal *l*'s adjacent to front vowels. If the focus were on phonetics instead of phonology, renderings as velar [ł] would only be found in 17% of Arabisms and 24% of Farsisms. No phonetic pattern is evident in their distribution, however.

When counted phonologically, the entropies are 0.89 for Arabisms and 0.61 for Farsisms. Phonetic counting inverses the proportion: 0.66 for Arabisms and 0.80 for Farsisms.

2.33 *ll*

	l	*ll*	*íí*
Arabisms	6	11	3
Farsisms	1	1	0

Shortening to single *l* occurred in A *balluṭ* > T *pelit*, *ǧallab* > *celep*, *ǧallābi* > *çelebi*, *qullaï* > *kule*, *sallaï* > *sele*, *siǧill* > *sicil*, and in P *čilla* > T *çile*. ¶ The one case of preservation of *ll* among Farsisms is in *kalla* > *kelle*. ¶ Lastly, the three palatalizations to *íí* are in A *billāhi* > T *billahi*, *bismillāh* > *bismillah*, *ǧallād* > *cellat*.

Similarly to *l* (Subsection 2.32 above), the proportions change when the distinction between *l* and *í* is counted phonetically rather than phonologically. In that case, rendering as velar [ł] occurs only in three examples in total, which are in fact just variations of one word: A *in šāʾa Allāh* > T *inşallah*, *īj wa-llāh* > *eyvallah*, *wa-llāhi* > *vallahi*.

In the case of *ll*, entropies do not change depending on whether counting is performed phonologically or phonetically. For Arabisms, it is always 0.89, and for Farsisms, naturally, 1.

2.34 *m* م

	m	*n*
Arabisms	387	1
Farsisms	99	0

The sole unusual rendering here is in A *balsam* > T *pelesenk*. See Subsection 2.1 above.

2.35 *mm*

	m	*m₂*	*mm*
Arabisms	4	1	8
Farsisms	0	0	0

Shortening to single *m* occurred in A *ḥammāl* > T *hamal*, *ḥammām* > *hamam*, *muhimm* > *mühim*, *summāq* > *sumak*. ¶ Positionally geminated *m₂* can be found in A *ḍamm* > T *zem*; this is one of just two words with -*mm* in the etymon, the other being *mühim* above.

The result is a high degree of unpredictability in Arabisms, with entropy of 0.78.

2.36 *n* ن

	∅	*m*	*n*	*nn*	*ń*
Arabisms	0	3	199	1	0
Farsisms	1	2	168	0	1

The one case of **deletion** is P *fešang* > T *fişek*. ¶ The renderings as *m* are in A *ʿanbar* > T *amber*, *inbīq* > *imbik*, *tanbīh* > *tembìh*, and P *ǧānbāz* > T *cambaz*,

zanpāra > zampara. It is only in A *anbijā > T *enbiya* that the *nb ~ np* sequence did not yield *m*. ¶ Gemination to **nn** occurred in A *ḫunāq > T *hunnak*, one of 44 words with the sequence *nā*, ten with *un*, and three with *unā* (the other two are A *munāsabaẗ > T *münasebet*, and P *gunāh > T *günah*). ¶ The one rare case of palatalized *ń* is P *tuvānā > tüvana* (see GTS).

Overall, the entropies are nearly identical, and very low: 0.098 for Arabisms, and 0.097 for Farsisms.

2.37   *nn*

	n	*n₂*	*nn*
Arabisms	3	2	5
Farsisms	0	0	0

Shortenings to **n**: A *ḍīq an-nafas > T *tıknefes* (cf. Subsection 2.10), *ğinn > cin*, *qinnāraẗ > kanara*. ¶ Positional preservation of gemination (**n₂**): A *fann > T *fen*, *ẓann > zan* (two of three cases in auslaut, the remaining one being *ğinn* above). ¶ Preservations of geminated **nn**: A *ğahannam > T *cehennem*, *ğannaẗ > cennet*, *ğinnaẗ > cinnet*, *minnaẗ > minnet*, *sunnaẗ > sünnet*.

Overall, the adaptations are exceptionally diversified, with the entropy for Arabisms being 0.94.

2.38   *o*

	o	*u*	*ù*	*ū*
Arabisms	0	0	0	0
Farsisms	9	1	1	1

The rounding to **u** occurred in P *rovġan > T *rugan*; ¶ rounding and positional lengthening (**ù**) in P *guroh > T *güruh*; ¶ rounding with lengthening (**ū**) in P *bosa > T *buse*.

The above amounts to the entropy of 0.6 for Farsisms, lower than for long *ō* (see Subsection 2.39 below, and also Subsection 2.3 above).

2.39   *ō*

	o	*ö*
Arabisms	0	0
Farsisms	3	2

The three cases of shortening to **o** are: P *čōb > T *cop*, *gōz > koz*, *pōst > post*; ¶ the two of shortening and fronting to **ö** are P *kōsa > T *köse*, *gōsāla > kösele*. This is too small a base for any phonetic pattern to emerge.

As with short *o*, the resulting entropy for Farsisms is very high: 0.97, higher than in the case of short *o* (see Subsection 2.38 above).

2.40 *p* پ

	b	*p*
Arabisms	0	0
Farsisms	2	55

The two unusual cases with *b* are P *pāpūš* > T *pabuç*, *taḫtapūš* > *tahtaboş*. The sequence *ap* ~ *āp* appears in two more words (P *čaprāst* > T *çapraz*, *sarāpā* > *serapa*), and so does *pu* ~ *pū* (P *pūl* > T *pul*, *pušt* > *puşt*).

The entropy for Farsisms is 0.22.

2.41 *q* ق

	g	*ǧ*	*k*
Arabisms	1	30	159
Farsisms	0	2	6

The one word with *g* is A *manqal* > T *mangal*, one of eight in which *q* is directly after a sonant, and one of two where it is directly after *n* (the other one is A *manqabaẗ* > T *menkıbe*). ¶ The two cases of *ǧ* are P *čārṭāq* > T *çardak*, *nāḥaq* > *nahak*; ¶ and the six with *k* are P *ḫāqān* > T *hakan*, *qahramān* > *kahraman*, *qalandar* > *kalender*, *qaltabān* > *kaltaban*, *qāqum* > *kakım*.

The entropy for Arabisms is 0.43, and for Farsisms 0.81.

2.42 *qq*

	k	*k₂*	*kk*
Arabisms	1	1	7
Farsisms	0	0	0

Here, we have one shortening to *k* in A *saqqāʾ* > T *saka*; ¶ one positional preservation of gemination (*k₂*) in A *ḥaqq* > T *hak* (the only word with -*qq* in the final position); ¶ and seven preservations as *kk* in A *baqqāl* > T *bakkal*, *diqqaẗ* > *dikkat*, *ʿuqqijaẗ* > *okka*, *naqqāš* > *nakkaş*, *raqqāṣ* > *rakkas*, *taraqqī* > *terakki*, *zaqqūm* > *zakkum*.

The entropy for Arabisms is 0.62.

2.43 *r* ر

	l	*r*	*ŕ*
Arabisms	0	362	0
Farsisms	3	240	1

Dissimilation to *l* can be observed in P *čārpāra* > T *çalpara*, *durgar* > *dülger*, *pargār* > *pergel*. It should be noted, however, that there are seven more Farsisms, and fourteen Arabisms, which contain two *r*'s, and no dissimilations occurred in them. ¶ The single case of functionally front *ŕ* is in P *jār* > T *yar*, -*ri*. Overall, P *r* does not appear to exert a strong fronting influence. Long vowels preceded by *r* preserve their

harmony in 92% of cases, though short vowels in the same position are fronted in 64% of examples. In Arabisms, the situation is similar, only more extreme in both cases.

Overall, the entropy is very low, only 0.08 for Farsisms.

2.44   *rr*

	r	*r₂*	*rl*	*rr*
Arabisms	4	1	0	6
Farsisms	0	0	1	1

The words with single ***r*** are: A *darrābaẗ* > T *taraba*, *karraẗ* > *kere*, *qārraẗ* > *kara*, *sarrāğ* > *saraç*; ¶ the one case of ***r₂*** is A *sirr* > T *sır* (the only occurrence of *rr* in auslaut); ¶ the one dissimilation to ***rl*** is P *šīrrauġan* > T *şırlağan*; ¶ and the preservations as ***rr*** are A *ḏarraẗ* > T *zerre*, *ğarrāḥ* > *cerrah*, *ğarrār* > *cerrar*, *mudarris* > *müderris*, *ṣarrāf* > *sarraf*, *tabarruʿ* > *teberru*, and P *zarrīn* > T *zerrìn*.

The entropy for Arabisms is 0.83, and for Farsisms, naturally, 1.

2.45   *s* س

	ç	*s*	*z*
Arabisms	0	162	1
Farsisms	1	89	2

The only example with ***ç*** is P *sirīš* > T *çiriş*. There are seven more Farsisms which contain both *s* and *š* – the combination does not occur in Arabisms – and in all of them *s* has been preserved. ¶ Renderings as ***z*** can be found in A *saʿtar* > T *zahter*, and P *čaprāst* > T *čapraz* (cf. Subsection 2.50 below), *pāsbān* > *pazvant* (cf. Subsection 2.5 above).

The entropies are low, 0.05 for Arabisms, and 0.15 for Farsisms.

2.46   *ss*

	s	*ss*
Arabisms	1	1
Farsisms	0	0

The two words here are: for ***s***: A *tamāss* > T *temas*; ¶ and for ***ss***: A *tabassum* > T *tebessüm*.

The entropy for Arabisms is of course 1.

2.47   *ş* ص

	s	*z*
Arabisms	84	1
Farsisms	0	0

There is only one exception here, A *ṣamġ* > T *zamk*. There are thirteen more words whose etyma begin with *ṣa-*, including nine where this is followed by a voiced

consonant, e.g. A *ṣabr* > T *sabır*, *ṣarf* > *sarf*, *ṣandūq* > *sandık*.

Naturally, the entropy for Arabisms is minimal, 0.09.

2.48   *ṣṣ*

	s	*ss*
Arabisms	2	2
Farsisms	0	0

The four words here are: with *s*: A *miqaṣṣ* > T *makas*, *qaṣṣāb* > *kasap*; ¶ and with *ss*: A *ḥiṣṣaẗ* > T *hisse*, *qiṣṣaẗ* > *kıssa*. No conclusions can be drawn from this small set.

The entropy for Arabisms is of course 1.

2.49   ش *š*

	C	*ç*	*s*	*ş*
Arabisms	0	0	1	82
Farsisms	1	1	0	70

The two renderings as *ç* are: positionally voiced *C* in P *pāpūš* > T *pabuç*; ¶ and entirely voiceless *ç* in P *šaltūk* > T *çeltik*. A similar change appears to have occurred in at least two words, *çakal* and *çorba*, which are however not included here because the entries in KEWT proposed alternative phonetic variants of their etyma (see the beginning of this section). NS suggests P *šaltūk ~ čaltūk* which, as a matter of fact, potentially renders the adaptations in *çakal*, *çorba*, and *pabuç*, even more unusual. ¶ The one remaining atypical adaptation as *s* is in A *furšat* > T *fırsat*.

For Arabisms, the entropy is 0.09, and for Farsisms 0.13.

2.50   *t* ت

	Ø	*d*	*D*	*t*
Arabisms	0	0	0	145
Farsisms	1	1	1	92

One **deletion** in P *čaprāst* > T *çapraz* (cf. Subsection 2.45 above); ¶ one voicing (*d*) in P *taġar* > *dağar*; ¶ and one positional voicing (*D*) in P *tarīt* > T *tirit*.

The entropy for Farsisms is 0.13.

2.51   *tt*

	t	*tt*
Arabisms	1	3
Farsisms	0	0

The shortening to *t* is in A *kattān* > T *keten*; ¶ the three *tt*'s in A *albatta* > T *elbette*, *ittifāq* > *ittifak*, *ittiḥād* > *ittihat*.

The resulting entropy for Arabisms is 0.81.

2.52 *ṭ* ط

	d	*D*	*t*
Arabisms	2	2	63
Farsisms	1	0	2

Three cases of *d*: A *biṭānaï* > T *badana*, *qaṭīfaï* > *kadife*, and P *čārṭāq* > T *çardak*; ¶ two of *D*: A *balluṭ* > T *pelit*, *šarīṭ* > *şerit*. ¶ The remaining two cases of P *ṭ* > T *t* are not so much unusual as just few because *ṭ* is simply rare in Persian: P *ṭamaʿkār* > T *tamahkar*, *ṭarāfdār* > *taraftar*.

The entropy for Arabisms is 0.24, and for Farsisms 0.92.

2.53 *ṭṭ*

	kt	*t₂*	*tt*
Arabisms	1	1	1
Farsisms	0	0	0

The three examples here are as follows: with **kt**: A *ʿaṭṭār* > T *aktar*; ¶ with **t₂**: A *ḫaṭṭ* > T *hat*; ¶ and with **tt**: A *baṭṭānījaï* > T *battaniye*.

The resulting entropy for Arabisms is naturally 1.

2.54 *ṯ* ث

	s
Arabisms	23
Farsisms	0

In the case of *ṯ*, all examples are perfectly regular. There are five cases of *ṯ*- in anlaut, six of -*ṯ* in auslaut, words with *ṯ* preceded by *a*, *ā*, *i*, *ī*, or a consonant, and there are examples of *ṯ* followed by *a*, *ā*, *ī*, or a consonant.

2.55 *ẗ* ة

	Ø	*t*	*t́*
Arabisms	161	132	15
Farsisms	0	0	0

The rendering of tāʾ marbūṭaï is not clear. My understanding of it has sadly not improved since K. Stachowski (2015: 315), and I can only confirm the observation made there, that the rendering as Ø or as *t* correlates with vowel harmony: words with rendering as Ø are back 36% of times, while words with *t* are back 18% of times if *aẗ* (as in *dikkat*, *-ti*) is counted as back-harmonic, and only 8% of times if it is counted as front-harmonic.[5]

The entropy is accordingly high, 0.77.

---

[5] In the first case, i.e. counting *aẗ* as back-harmonic, χ^2 (1, N=308) = 11.1, p = 8.5×10⁻⁴. Counting *aẗ* as front-harmonic, χ^2 (1, N=308) = 32.4, p = 1.3×10⁻⁸.

2.56 *u*

	∅	*a*	*ı*	*i*	*o*	*ö*	*u*	*ʾu*	*ù*	*ū*	*ü*	*ū̄*
Arabisms	0	1	2	3	6	2	82	2	1	4	68	1
Farsisms	1	0	3	2	2	0	20	2	0	0	44	0

The sole **deletion** is in P *šïrrauġan* > T *şırlağan*. ¶ The one rendering as *a* in A *fulān* > T *falan*. ¶ The five renderings as *ı* in A *furšat* > T *fırsat*, *ġurnūq* > *kırnak*, and P *bahādur* > T *bahadır*, *qāqum* > *kakım*, *partavsuz* > *pertavsız*. ¶ The five cases of *i* are in A *fulān* > T *fîlan*, *balluṭ* > *pelit*, *tadāruk* > *tedarik*, and P *nāzuk* > T *nazil*, *zardālu* > *zerdali*. ¶ The eight examples with *o* are A *ʿuqqijaï* > T *okka*, *ġuġrāfijaï* > *coğrafya*, *fuḍūl* > *fodul*, *luqmaï* > *lokma*, *nuqsān* > *noksan*, *nuqṭaï* > *nokta*, and P *silāḫšur* > T *silahşor*, *durdī* > *tortu*. ¶ The two words with *ö* are A *ʿurf* > T *örf*, *šuhraï* > *şöhret*. ¶ The four atypically stressed *ʾu*'s: A *Qurʾān* > T *Kurʾan*, *muṭlaqā* > *mutlakā*, and P *mušt* > T *muşta*, *ustura* > *ustura*. ¶ The one positionally long *ù* is in A *zujuf* > T *züyuf*. ¶ The four cases of long *ū*: A *muʿǧizaï* > T *mucize*, *muʿtabar* > *muteber*, *tabarruʿ* > *teberru*, *šuʿbaï* > *şube*. ¶ Lastly, long *ū̄* is in A *muʾmin* > T *mumin*. All these groups are unfortunately too small for any clear phonetic patterns to emerge.

Fronting occurred in 43% of Arabisms and 62% of Farsisms. In general, the number of different renderings is high but they are all rare, so the overall unpredictability is lower than it might at first seem, with entropy being 0.52 for Arabisms and 0.59 for Farsisms. It is, in fact, also lower than for long *ū*, see Subsection 2.57 below, and also Subsection 2.3 above.

2.57 *ū* و

	a	*e*	*ı*	*i*	*o*	*ö*	*u*	*ù*	*ū*	*ü*	*ʾü*
Arabisms	2	0	3	0	2	0	42	33	15	2	0
Farsisms	0	1	6	3	2	3	11	1	5	5	1

Delabialization can be observed in fifteen examples: two with *a*: A *ġurnūq* > T *kırnak*, *maṣrūf* > *masraf*; ¶ with *e*: P *zambūrak* > T *zemberek*; ¶ nine with *ı*: A *ḥarūn* > T *harın*, *ṣandūq* > *sandık*, *sāṭūr* > *satır*, and P *dādū* > T *dadı*, *māzū* > *mazı*, *bāzū* > *pazı*, *čākū* > *çakı*, *ǧāmašūr* > *çamaşır*, *ǧuvāldūz* > *çuvaldız*; ¶ and three with *i*: P *lablabū* > T *leblebi*, *tarāzū* > *terazi*, *šaltūk* > *çeltik*. It might be noticed that the adaptations to *ı* and *ı* all occur in the final syllable. However, this is unlikely to be the cause for these unusual renderings as these twelve cases constitute only 4% of Arabisms with *ū* in the final syllable, and 33% of analogously built Farsisms.

The renderings as *o* are in A *ṣūf* > T *sof*, *ṣūfī* > *sofu*, and P *taḫtapūš* > T *tahtaboş*, *šabbūǧ* > *şebboy*; ¶ as *ö* in P *ǧūdmard* > T *cömert* (cf. Subsection 2.8 above), *kūr* > *kör*, *nānkūr* > *nankör*; ¶ as *ù* in P *nawrūz* > T *nevruz*; ¶ as *ū* in P *āsūda* > T *asude*, *farsūda* > *fersude*, *fīrūza* > *firuze*, *rūznāma* > *ruzname*, *zabūn* > *zebun*; ¶ as *ü* in A *ajlūl* > T *eylül*, *rūʾjā* > *rüya*, and P *dūrbīn* > T *dürbün*, *kākūl* > *kakül*, *gūgird* > *kükürt*, *kūra* > *küre*, *lūla* > *lüle*; ¶ and lastly as *ʾü* in P *čūnki* > T *çünkü*.

As is usually the case with long vowels, fronting of long *ū* is considerably rarer than that of short *u*, as it only happened in 20% of Arabisms and 34% of Farsisms. Overall, however, these are more diversified adaptations than those of short *u* (see Subsection 2.56 above), and the entropies are 0.7 for Arabisms, and 0.88 for Farsisms.

2.58 *v* و

	∅	*v*
Arabisms	0	0
Farsisms	1	34

The only unusual case here is the deletion in A *rovġan* > T *rugan*. It is the only etymon with either the *ov* or the *vġ* sequence. The only two examples with any consonant at all directly following *v* are P *čāvdār* > T *çavdar*, *partavsuz* > *pertavsız*.

The entropy for Farsisms is 0.19.

2.59 *w* و

	l	*p*	*v*
Arabisms	1	1	104
Farsisms	0	0	2

The two highly atypical examples here are: in *l*: A *mawlā* > T *molla* (cf. Footnote 4); ¶ and in *p*: A *kiswaẗ* > T *kispet*. The *wl* sequence appears in six more words; *sw* is not attested in any other, and neither is *ṯw*, but *ṣw* is in A *taṣwīr* > T *tasvir*. ¶ The two cases of *w* in Persian are P *ferdaws* > T *firdevs*, *nawrūz* > *nevruz*.

The entropy for Arabisms is 0.1.

2.60 *ww*

	vv
Arabisms	2
Farsisms	0

The only two cases here are A *qūwwaẗ* > T *kuvvet*, *munawwar* > *münevver*.

2.61 *z* ز

	∅	*s*	*z*
Arabisms	0	0	68
Farsisms	1	1	76

The one case of **deletion** is in P *zorbāz* > T *zorba*; ¶ and the one case of devoicing to *s* is in P *pīhsuz* > T *pesüs*. There are in total 24 Farsisms and eighteen Arabisms with final -*z*.

The entropy for Farsisms is 0.12.

2.62 *zz*

	z	*zz*
Arabisms	1	1
Farsisms	0	0

Here, there is only one example with *z*: A *razzaï* > T *reze*; ¶ and one with *zz*: A *bazzāz* > T *bezzaz*.

The entropy, of course, is 1 for Arabisms.

2.63 *z̧* ظ

	z
Arabisms	23
Farsisms	0

The 23 examples include words with *z̧* in anlaut, in auslaut, and preceeded and followed by different vowels. There is only one example with *z̧* adjacent to a consonant: A *ḥifz̧* > T *hıfız*.

2.64 *z̧z̧*

	z_2
Arabisms	1
Farsisms	0

The one example here is A *ḥaz̧z̧* > T *haz*, which is a perfectly regular adaptation (cf. Subsection 2.63 above).

2.65 *ž* ژ

	j
Arabisms	0
Farsisms	4

The four examples here are P *aždar* > T *ejder*, *aždarhā* > *ejderhā*, *žāla* > *jale*, *mužda* > *müjde*; the one significant position that is unattested is -*ž* in auslaut.

3 Summary

Based on a dataset of 1748 words, the paper provides a quantitative summary of the commonness of various phonetic adaptations that borrowings from Arabic and Persian underwent in Turkish, as well as lists of cases with unusual renderings, accompanied by a brief discussion of the phonetic surrounding. Etymological remarks are given only occasionally. In a sense, however, all the words mentioned in the paper combine into a list of cases that do perhaps require a revision in this aspect, precisely because they are the ones that exhibit atypical adaptations.

Quantitative data can also find various applications. One example has been suggested in K. Stachowski 2020 in the form of an algorithm for detecting Persian mediation in Arabisms. They also pose at least one problem, however, as the results appear to indicate that the renderings of long vowels are less predictable than those of short ones (cf. Subsection 2.3).

Abbreviations and references

A = Arabic; **Engl**. = English; **Fr.** = French; **It.** = Italian; **P** = Persian; **T** = Turkish

AKIN, Cüneyt, (2011). "Kırgız Türkçesinde Alıntı Kelimelerde Ses Değişmeleri", **Turkish Studies**, 6(1): 621-635.

AL-HASHMI, Shadiya, (2016). ***The phonetics and phonology of Arabic loanwords in Turkish: residual effects of gutturals***, Unpublished doctoral dissertation, University of York, York, United Kingdom.

ALOT = S. Stachowski 1975–1986.

BARAN, Burhan, (2019). "Farsça Birleşik Sözcüklerin Türkiye Türkçesindeki Kullanilişlari Üzerine", **Uluslararası Sosyal Araştırmalar Dergisi**, 62: 60–65.

CUMAKUNOVA, Gülzura, (2008). "Kırgız Türkçesinde Arapçadan Alıntı Kelimeler", **Türkiyat Araştırmaları**, 9: 101–117.

ERDEM, İlhan, (2013). "Konuşma Dili Açısından Arapça ve Farsça Alıntı Kelimelerin Bugünkü Durumu", **Ana Dili Eğitimi Dergisi**, 1(1): 92–100.

EREN, Hasan, (²1999). ***Türk Dilinin Etimolojik Sözlüğü***. Ankara: Bizim Büro Basım Evi.

GTS = "Türk Dil Kurumu: Güncel Türkçe Sözlük", https://sozluk.gov.tr.

İŞLER, Emrullah, (2008). "Türkçedeki Arapça Alıntı Kelimelerde Ses Değişiklikleri", **Ekev Akademi Dergisi**, 35: 187–196.

KARACA, Veysel İbrahim, (2012). "Türkiye Türkçesindeki Alıntı Sözcüklerde Görülen Ses Olayları Üzerine Bir İnceleme", **Turkish Studies**, 7/4: 2059–2090.

KARASOY, Yakup, (1998). "Türkmen Türkçesinde Arapçadan Geçen Ayın'lı Kelimelerdeki Ses Değişmeleri", **Türk Dünyası Dil ve Edebiyat Dergisi**, 5: 397–432.

KEWT = M. Stachowski 2019.

KORKMAZ, Feryal, (2007). "Alıntı Kelimelerin Türkçeleşme Surecinde Bazı Arapça Kelimelerin Görev Değişikliğine Uğraması Üzerine", **İlmî Araştırmalar**, 23: 103–112.

NİŞANYAN, Sevan, (on-line), "Nişanyan Sözlük", https://www.nisanyansozluk.com.

NS = Nişanyan (on-line).

OYTUN ALTUN, Hilal, (2012). "Türkiye Türkçesindeki Arapça Alıntı Kelimelerde Ünlü Uyumsuzlukları", **Türkiyat Araştırmaları Dergisi**, 32: 55–71.

OYTUN ALTUN, Hilal, (2015). "Arapça Alıntı Kelimelerin Kazak Türkçesine Uyarlanması", **Türkiyat Araştırmaları**, 22, 167–201.

ÖZKAN, Nevzat, (2011). "Türkçedeki Yabancı Unsurları Tasnif Denemesi", **38. ICANAS**, (Vol.3; Ed. Z. Dilek et al.), Ankara: Atatürk Kültür, Dil ve Tarih Yüksek Kurumu, 1343–1359.

ÖZTEKTEN, Özkan, (2001). "Türkçedeki Arapça Alıntıların Fonolojisi Üzerine", **Ege Üniversitesi Türk Dili ve Edebiyatı Araştırmaları Dergisi**, 10: 315–320.

ÖZTEKTEN, Özkan, (2013). "Osmanlı Türkçesindeki Arapça Alıntılarda Yapımlık Ses Değişiklikleri. İlletli Kökler", ***Yalım Kaya Bitigi. Osman Fikri Sertkaya Armağanı***, (Ed. H. Şirin User & B. Gül), Ankara: Türk Kültürünü Araştırma Enstitüsü, 1–12.

PEKAÇAR, Çetin, (2006). "Kumuk Türkçesine Arapça ve Farsçadan Geçen Kelimelerdeki Ses Olayları", **Selçuk Üniversitesi Türkiyat Araştırmaları Dergisi**, 19: 53–71.

PEKAÇAR, Çetin, (2007). "Kumuk Türkçesindeki Arapça Alıntı Kelimelerde Son Ses /U/ Meselesi ve Türkçe-Arapça İlişkileri*", IV. Uluslararası Türk Dili Kurultayı Bildirileri II 24–29 Eylül 2000*, Ankara: Türk Dil Kurumu, 1465–1473.

PLOT = S. Stachowski 1998.

POMORSKA, Marzanna, (2013). ***Materials for a historical dictionary of New Persian loanwords in Old Anatolian and Ottoman Turkish from the 13th to the 16th century***, (= *Studia Turcologica Cracoviensia* 13), Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

R = R Core Team, (2020). ***R: A language and environment for statistical computing***. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

ROCCHI, Luciano, (2016–2017). "Addenda from Pre-Meninski transcription texts to Stanisław Stachowski's "Osmanlı Türkçesinde Yeni Farsça Alıntılar Sözüğü"", (Parts 1–11), **Studia Linguistica Universitatis Iagellonicae Cracoviensis**, 133–134.

SAĞ, Zehra, (2019). ***Farsça Üzerinden Türkçeye Alıntılanmış Arapça Menşeli Kelimeler***, Unpublished MA thesis, İstanbul Medeniyet Üniversitesi, Istanbul, Turkey.

SARIKAYA, Yakup, (2005). "Türkmen Türkçesindeki Arapça Alıntı Kelimelerde Görülen Ünlü İncelme Ve Kalınlaşmaları", **Türk Dünyası Dil ve Edebiyat Dergisi**, 19: 63–80.

STACHOWSKI, Kamil, (2015). "Phonetic adaptation of Arabic borrowings in Argenti's Ottoman

Turkish (1533). Part 1. Consonants and semivowels", ***Words and Dictionaries. A Festschrift for Professor Stanisław Stachowski on the Occasion of His 85th Birthday***, (Ed. In E. Mańczak-Wohlfeld & B. Podolak), Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 297–317.

STACHOWSKI, Kamil, (2020). "*soundcorrs*: Tools for Semi-automatic Analysis of Sound Correspondences", **Glottometrics**, 49: 66–86.

STACHOWSKI, Kamil, [forthcoming], "Detecting Persian mediation in Arabisms in Turkish".

STACHOWSKI, Marek, (2012a). "On the consonant adaptation of Arabic (and some Persian) loan-words in an Ottoman Turkish dictionary by Arcangelo Carradori (1650)", **Studia Linguistica Universitatis Iagellonicae Cracoviensis**, 129 supplementum: 373–381.

STACHOWSKI, Marek, (2012b). "Some features of the vowel adaptation of Arabic loan-words (along with a few remarks on their consonants) in an Ottoman Turkish dictionary by Arcangelo Carradori (1650)", **Folia Orientalia**, 49: 491–502.

STACHOWSKI, Marek, (2019). ***Kurzgefaßtes etymologisches Wörterbuch der türkischen Sprache***. Kraków: Księgarnia Akademicka.

STACHOWSKI, Stanisław, (1975–1986). ***Studien über die arabischen Lehnwörter im Osmanisch-Türkischen***, (= *Prace Komisji Orientalistycznej PAN* 13, 15, 17, 19), Wrocław, Warszawa, Kraków, Gdańsk, Łódź: Ossolineum.

STACHOWSKI, Stanisław, (1998). ***Osmanlı Türkçesinde Yeni Farsça Alıntılar Sözlüğü. Wörterbuch der neupersischen Lehnwörter im Osmanisch-Türkischen***, (Ed. M. Ölmez), İstanbul: Simurg.

TDES = Eren 1999.

TETTL = Tietze 2002–2009.

TIETZE, Andreas, (2002–2009). ***Tarihi ve Etimolojik Türkiye Türkçesi Lugatı. Sprachgeschichtliches und etymologisches Wörterbuch des Türkei-Türkischen***, (Vols. 1–2), İstanbul, Wien: Simurg, Österreichische Akademie der Wissenschaften.

TRT = "TRT Tellafuz Sözlüğü: Genel Sözlük", http://trttelaffuz.com.

YAZICI ERSOY, Habibe, (2006). "Başkurt Türkçesinde Arapça ve Farsçadan Alıntı Kelimelerdeki Ünlü ve Ünsüz Değişmeleri", **Türkbilig Türkoloji Araştırmaları Dergisi**, 12: 143–162.