

KAMIL STACHOWSKI
Jagiellonian University in Kraków
kamil.stachowski@gmail.com
ORCID: 0000-0002-5909-035X

DETECTING PERSIAN MEDIATION IN ARABISMS IN TURKISH

Keywords: etymology, Turkish, Arabic, Persian, loanword adaptation

Abstract

There is a sizeable group of words in Turkish whose ultimate origin is known to be Arabic but whose direct donor language is unclear. The paper analyses 69 such words, and compares the phonetic adaptations present in them, to those attested in Arabisms as well as to those found in Farsisms, in order to determine the probability of them belonging to one group or the other. The results are compared to the opinions of the main etymological dictionaries of Turkish, splitting them into two camps.

1. Introduction

It is a well-known problem in Turkology, that Turkish has borrowings from Arabic and Persian, but also Persian contains a large number of loanwords from Arabic, and Arabic some from Persian. Establishing the direct donor language is not always possible using solely traditional philological methods, but a quantitative approach to phonetics can help shed light on some of the problematic cases.

Below is a concise description of the data on which the results presented here are based, and a short explanation of the method which produced them. The results themselves can be found, together with a brief discussion, in Section 2. Summary and conclusions are in Section 3. In the Appendix, a list of words is given whose phonetic adaptation suggests that their established etymologies might be incorrect.

The data consist of two datasets: one that serves as the foundation, and one that contains unclear words. The first set contains 1234 Arabisms and 514 Farsisms whose etymologies do not involve alternative phonetic shapes or incomplete attestations, and do not appear to be in any other way doubtful or contested.

A fuller presentation of this set can be found in K. Stachowski (2020a). The other set contains 69 words for which the direct donor, Arabic or Persian, has either not been conclusively established, or is at least theoretically uncertain. Both sets have been extracted from KEWT, but since KEWT cites words in the contemporary orthography, the phonetic details such as accentuation, non-harmonic palatalization, etc. needed to be supplemented from other sources; GTS and TRT have been used for this purpose. Beside the etymologies proposed in KEWT, the opinions of GTS, NS, TDES, and TETTL have been adduced. A comparison of their views can be found in Section 3.

The method used to obtain the results presented here is a relatively heavily modified variant of naïve Bayes classification. The classifier is trained on the “foundation” dataset by extracting from it all the phonetic adaptations attested in it, and calculating for each its relative frequency in Arabisms, and in Farsisms. Then, for a word whose etymology is to be established, the frequencies of the adaptations present in that word are combined into a single number which is effectively a measure of how typical such a set of adaptations is for Arabisms, versus for Farsisms. These values are reported here as $P(A)$ and $P(F)$, respectively. The final score is their ratio, BF'_{AF} , which is proportional to the probability that our word has been borrowed from Arabic or from Persian – at least so far as phonetic adaptation is concerned. Table 1 contains a guide to the interpretation of BF'_{AF} . (The above description omits several details which are necessary mathematically, but not important linguistically. A full explanation can be found in K. Stachowski [forthcoming].)

BF'_{AF}	Interpretation
> 10	Very strong evidence for Arabic origin
6 to 10	Strong evidence for Arabic origin
2 to 6	Positive evidence for Arabic origin
-2 to 2	Anecdotal evidence
-2 to -6	Positive evidence for Persian origin
-6 to -10	Strong evidence for Persian origin
< -10	Very strong evidence for Persian origin

Table 1. Interpretation of BF'_{AF} . Adapted from Kass, Raftery (1995: 777) and K. Stachowski [forthcoming].

2. Unclear words

Below is a list of 69 words where KEWT provided both an Arabic and a Persian etymon, and neither had alternative phonetic variants. In some cases one was simply given as the source for the other, in others one was suggested as a potential mediator between the other and Turkish. All situations have been treated equally here.

Entries are structured as follows:

entry | Arabic and Persian etyma | opinions of GTS, KEWT, NS, TDES, and TETTL
| quantitative indices | my opinion | comments.

Roman numerals in *entry* refer to entries in KEWT. Opinions of the five dictionaries are abbreviated to just a chain of languages, without marking the discrepancies between the phonetic shapes of etyma which sometimes arise between them. Quantitative indices include $P(A)$, $P(F)$, and BF'_{AF} (see Section 1). The purpose of the first two is to give a picture of the absolute probability of each etymology, and of the latter, a picture of their probability relative to each other. My own opinion (A or F) is decorated with one of three symbols: $-$, \pm , $+$, which indicate that I am hardly certain, fairly certain, or quite certain of it.

As was mentioned above, the method employed here is based on sound correspondences. In this particular application, phoneme-to-phoneme correspondences have been used, whereby the term *phoneme* is understood loosely. In Turkish, “positionally voiced consonants” and “positionally long vowels” (as in *cevap*, *-ābi*) are distinguished as separate entities, different from p/b , a/\bar{a} , and so are various other quirks such as an unusual place of stress, a palatalized consonant adjacent to a back vowel, etc. In Persian, the transcription follows both pronunciation and orthography simultaneously because the use of e.g. emphatic or non-emphatic variant in the spelling would be likely to have influenced the Turkish rendering.

All the calculations, extraction of examples, etc. have been performed in *R* using the *soundcorrs* package (K. Stachowski 2020b).

abes ‘useless; absurd’ | A ‘*abaṭan* ‘in vain’ ~ P ‘*abaṭ* ‘useless; absurd’ | GTS: $\langle A$; KEWT: $\langle P \langle A$; NS: $\langle A$; TDES: missing; TETTL: $\langle P \langle A$ | $P(A)=0.06$, $P(F)=0.61$, $BF'_{AF}=-27.16$ | Here: $\langle P^+$ | It is primarily A $n > T \emptyset$ that determines the fate of the Arabic etymology, followed in distant second place by A $a > T \emptyset$. There are seven Arabic words in our dataset that end in *-an*, and nine Persian ones, and in all of them the *-an* sequence has been preserved, regardless of whether it is the adverb suffix like in *abaṭan* or just part of the word. Had it not been for the *-an*, BF'_{AF} would in fact slightly favour a direct borrowing from Arabic (1.01).

acaba ‘I wonder ...’ | A ‘*ağaban* ‘how strange ...’ ~ P ‘*ağabā* id. | GTS: $\langle A$; KEWT: $\langle P \langle A$; NS: $\langle A$; TDES: missing; TETTL: $\langle P \langle A$ | $P(A)=0.07$, $P(F)=0.42$, $BF'_{AF}=-22.79$ | Here: $\langle P^+$ | By far the largest contributor to the score here is $n > \emptyset$. Unlike in *abes* above, however, its removal would not change the verdict ($BF'_{AF}=-1.83$) as the final *-a* in *acaba* is in fact long, and lengthening of short *a* is a very rare adaptation.

acur ‘Armenian cucumber’ | A ‘*ağūr* ‘sponge gourd’ ~ P *angūr* ‘grape’ | GTS: no etymology; KEWT: $\langle A \langle P$; NS: $\langle A$; TDES: $\langle A$; TETTL: $\langle A$ | $P(A)=0.67$, $P(F)=0.14$, $BF'_{AF}=14.27$ | Here: $\langle A^+$ | It is the deletion of n that determines the score here. Even if the Persian etymon did not have it, however, BF'_{AF} of 4.27 would still favour a direct borrowing. One adaptation had to be omitted: P $g > T$ c , as it is not attested in the “foundation” dataset and its frequency is unknown. This only weakens the probability of Persian mediation even further.

- afyon** ‘opium’ | A *afjūn* id. ~ P *afjūn* id. | GTS: <A; KEWT: <A ~ P; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.2, P(F)=0.4, BF'_{AF}=-7.55$ | Here: <P⁺ | The two main contributors here are A *i* > T \emptyset , and A *ū* > T *o*. The latter is also an atypical adaptation for a Farsism but to a lesser degree than it is for an Arabism.
- ait** ‘concerning; belonging to’ | A ~ P ‘ā’id id. | GTS: <A; KEWT: <P<A; NS: <A; TDES: missing; TETTL: <P<A | $P(A)=0.48, P(F)=0.36, BF'_{AF}=2.93$ | Here: <A[±] | The devoicing of *d* is overall a relatively rare adaptation, but in the auslaut it is only to be expected. The other renderings are generally a little more typical for Arabisms than for Farsisms, and this is reflected by the score which is not decisive but positive evidence in favour of a direct borrowing.
- arz** ‘presentation, show, offer’ | A ~ P ‘arḍ id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.74, P(F)=0.44, BF'_{AF}=4.14$ | Here: <A[±] | There is no single major contributor to the score here. Rather, the adaptations are overall more typical for a direct borrowing from Arabic than for one from Persian.
- avrat** ‘woman (*pej.*)’ | A ‘awrat ‘genitals, esp. feminine’ ~ P ‘avrat id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.1, P(F)=0.26, BF'_{AF}=-11.86$ | Here: <P⁻ | The final *-t* in T *avrat* is voiceless when in auslaut, but it becomes voiced in the presence of a vowel-initial suffix. There is only one other example of voiceless *t* being rendered as such “positionally voiced” *t*, and it is P *tarīt* > T *tirit*, *-di*. This adaptation is what determined the score in the case of *avrat*; had its final consonant been a simple voiceless *t*, BF'_{AF} would be 2.87. I am not certain to what degree this result is reliable.
- bayat** ‘stale’ | A *bā’it* id. ~ P *bajāt* id. | GTS: no etymology; KEWT: <P (?<A); NS: <A; TDES: <A ~ P; TETTL: <A | $P(A)=0.19, P(F)=0.57, BF'_{AF}=-10.8$ | Here: <P⁺ | The main contributor to the score here is the unlikely adaptation of A ‘i as T *ya*. This is in accordance with my human intuition.
- beden** ‘body’ | A ~ P *badan* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.68, P(F)=0.78, BF'_{AF}=-1.3$ | Here: <P⁻ | The renderings of *b* are more diversified than perhaps would be expected. See K. Stachowski (2020a), but in short, simple preservation is slightly more common in Farsisms. So is the fronting of *a*, and the score which only indecisively favours Persian mediation, reflects that.
- beytūmal** ‘treasury’ | A *bajtul-māl* id. ~ P *bajtulmāl* id. | GTS: <A; KEWT: <P<A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.34, P(F)=0.42, BF'_{AF}=-3.54$ | Here: <P⁺ | The largest contributor here is the adaptation of *a* as *e* that is stressed despite not being in the final syllable. It is an unusual adaptation for a Farsism, but even more unusual for an Arabism. Had the *e* not been stressed, BF'_{AF} would be 0.44, inconclusive. In opposition to *fakat* below, however, this unusual place of stress cannot be explained by semantics, which adds weight to this “argument from stress”.
- billur** ‘crystal’ | A ~ P *billūr* id. | GTS: <A; KEWT: <A ~ P; NS: <A ~ P; TDES: missing; TETTL: <P | $P(A)=0.5, P(F)=0.15, BF'_{AF}=12.13$ | Here: <A[±] | This result has been primarily determined by the adaptation of *ll* as *ll̥*, i.e. *l* that is palatal despite being followed by a back vowel. This is a rendering that does not occur among

Farsisms at all. It needs to be noted, however, that geminated *ll* is generally rare in Persian, and the dataset contains only two examples of it: P *çilla* > T *çille*, and *kalla* > *kelle*, so the strength of evidence in favour of a direct borrowing is in fact less than the score would suggest.

cenup ‘south’ | A ~ P *ġanūb* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.52$, $P(F)=0.24$, $BF'_{AF}=7.95$ | Here: <A[±] | The strongest contributors here are positionally long *u* and positionally voiced *p* (*cenup*, *-ūbu*). These are relatively frequent adaptations in Arabisms, and rare in Farsisms.

cephē ‘front’ | A ~ P *ġabha* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.49$, $P(F)=0.56$, $BF'_{AF}=-1.34$ | Here: ? | The adaptation that contributed most to the score here is the devoicing of *b*. It is generally not very common, and proportionally rarer in Arabisms. In the position directly before *h*, however, such an outcome is all but inevitable. With this one rendering removed, BF'_{AF} equals -0.26 . The result must be considered inconclusive.

cevap ‘answer’ | A *ġawāb* id. ~ P *ġavāb* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.5$, $P(F)=0.31$, $BF'_{AF}=4.57$ | Here: <A[±] | The biggest contributor here is the adaptation of final *-b* as positionally voiced (*cevap*, *-bi*). In Arabisms, such adaptation has been applied in 88% of cases of *-b* in the auslaut of the etymon, while in Farsisms only in 50%. Cf. *şarap* below. Positional preservation of length in *a* (*cevap*, *-ābi*) is also more typical for borrowings from Arabic.

ceza ‘punishment’ | A *ġazā* id. ~ P *ġazā* id. | GTS: <A; KEWT: <(P<) A; NS: <A ~ P; TDES: missing; TETTL: <A | $P(A)=0.74$, $P(F)=0.64$, $BF'_{AF}=1.37$ | Here: <A⁻ | No adaptation is here particularly more typical for Arabisms or Farsisms. They are generally slightly more common in the former group, but not sufficiently for the result to be considered significant.

cezire ‘island’ | A *ġazīra* id. ~ P *ġazīra* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.65$, $P(F)=0.65$, $BF'_{AF}=0.01$ | Here: ? | This is the second least decisive result in the entire dataset. The Turkish shape is precisely what would be expected from both A *ġazīra* and P *ġazīra*, and no conclusion can be made based on phonetics alone.

dahi I ‘genius’ | A ~ P *dāhī* ‘smart, cunning’ | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.59$, $P(F)=0.49$, $BF'_{AF}=1.51$ | Here: <A⁻ | This word lacks phonemes that would be adapted in a significantly different way in Arabisms and in Farsisms. While the BF'_{AF} does lean somewhat in the direction of a direct borrowing from Arabic, this result cannot be considered conclusive.

davul ‘drum’ | A ~ P *ṭabl* id. | GTS: <A; KEWT: <P<A; NS: <A ~ P; TDES: missing; TETTL: <A | $P(A)=0.02$, $P(F)=0.15$, $BF'_{AF}=-22.73$ | Here: <P⁺ | The rendering of *b* as *v* occurs in two borrowings from Persian (*pāsbān* > *pazvant*, and *partāb* > *pertav*), and none from Arabic, and it is it that contributes most to the score in the case of *davul*. However, the other adaptations also suggest Persian mediation and even if the Turkish shape were **dabul*, BF'_{AF} would be -5.91 , fairly firmly indicating indirect borrowing.

dehliz ‘corridor’ | A ~ P *dihlīz* id. | GTS: <P; KEWT: <A ~ P; NS: <A ~ P; TDES: missing; TETTL: <A | $P(A)=0.38$, $P(F)=0.5$, $BF'_{AF}=-3.3$ | Here: <P[±] | It is primarily

$i > e$ that tips the scales in favour of Persian mediation here. Without it, the score would be -0.41 , inconclusive. The BF'_{AF} of -3.3 is not conclusive as such, but it is an indication.

- dut** I ‘mulberry’ | A ~ P *tūt* ‘berry, esp. mulberry’ | GTS: <P; KEWT: <A ~ P; NS: <A; TDES: <P; TETTL: <P<A | $P(A)=0.01$, $P(F)=0.14$, $BF'_{AF}=-13.83$ | Here: <P⁻ | Voicing of initial *t*- is a very rare adaptation, there is in fact only one example of it in the “foundation” dataset: P *tağar* > T *dağar*. This is scant evidence, and the word is simply too short to provide much more, so despite the relatively high $|BF'_{AF}|$, I would consider this result less than conclusive, an indication.
- ehemmiyet** ‘importance’ | A *ahammījajī* id. ~ P *ahammījat* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.6$, $P(F)=0.69$, $BF'_{AF}=-2.2$ | Here: <P⁻ | There is no single major contributor here; instead, most adaptations are just a little atypical for an Arabism and fairly typical for a Farsism. The negative, but not very large BF'_{AF} , reflects this situation quite accurately.
- elmas** ‘diamond’ | A *’almās* id. ~ P *almās* id. | GTS: <A; KEWT: <A ~ P; NS: <A ~ P; TDES: missing; TETTL: <A | $P(A)=0.65$, $P(F)=0.76$, $BF'_{AF}=-1.81$ | Here: <P⁻ | The main hint here is the shortening of *ā*, an adaptation that is rarer among Arabisms than among Farsisms. The other renderings are fairly neutral, they are generally simply the most common ones in both groups. Had the *ā* been either fully preserved or at least positionally preserved (**elmas*, *-āsi*), BF'_{AF} would be positive but still below the threshold of significance (0.26, and 1.55, respectively).
- erbab** ‘master’ | A ~ P *arbāb* id. | GTS: <A; KEWT: <P<A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.47$, $P(F)=0.32$, $BF'_{AF}=3.9$ | Here: <A[±] | This result has been mostly informed by two adaptations: positional preservation of length, and positional preservation of voicedness (*erbab*, *-ābi*), both rarer occurrences in Farsisms than they are in Arabisms.
- etraf** ‘environs’ | A ~ P *aṭraf* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.67$, $P(F)=0.55$, $BF'_{AF}=1.98$ | Here: <A⁻ | Positional preservation of vowel length (*etraf*, *-āfi*) is more typical in borrowings from Arabic than in those from Persian. The other renderings present in this word are fairly typical for both groups, producing a positive but relatively low score.
- evvel** ‘first, initial’ | A *awwal* id. ~ P *avval* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.69$, $P(F)=0.77$, $BF'_{AF}=-0.75$ | Here: ? | The rendering P *vv* > T *vv* needed to be omitted here as geminated *vv* does not occur in Farsisms in the “foundation” dataset, and the frequencies of its Turkish reflexes are not known. This reduces the Persian etymology to just two adaptations: twice $a > e$, and once $l > l$, which is not sufficient base to draw any decisive conclusions.
- fakat** ‘but’ | A ~ P *faqaṭ* id. | GTS: <A; KEWT: <A>P; NS: <A; TDES: missing; TETTL: <P<A | $P(A)=0.65$, $P(F)=0.5$, $BF'_{AF}=3.9$ | Here: <A[±] | The largest contributor here is the rendering of the first *a* as an *a* that is stressed despite not being in the final syllable. This is a rare occurrence in Arabisms, and an even rarer one in Farsisms. The unusual place of stress may be emphatic here and connected to semantics (*fakat* ‘but’), but even without it BF'_{AF} would point in the direction of a direct borrowing from Arabic, only not as decisively (2.58).

fedā ‘sacrifice’ | A ~ P *fidā* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.33$, $P(F)=0.32$, $BF'_{AF}=0.32$ | Here: ? | Lowering of *i* to *e* is an atypical adaptation for both Arabisms and Farsisms, but more so for the former. The difference, however, is not sufficient to single-handedly produce a decisive score, all the other adaptations being neutral.

fil ‘elephant’ | A *fil* id. ~ P *pīl* id. | GTS: <A; KEWT: <A<P; NS: <A ~ P; TDES: missing; TETTL: <A<P | $P(A)=0.68$, $P(F)=0.63$, $BF'_{AF}=0.34$ | Here: <A⁺ | This word provides little evidence as it is simply just very short, and made even shorter by the necessity to omit one adaptation from the Persian etymology, as the rendering of *p* > *f* does not occur in any word in the “foundation” dataset, and its frequency is unknown. On the one hand, this results in a score that is far from conclusive, but on the other, it is an information that the Persian etymology involves a particularly unusual rendering.

gam ‘sorrow’ | A *ḡamm* id. ~ P *ḡam* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <A | $P(A)=P(F)=0.41$, $BF'_{AF}=-0.0003$ | Here: ? | This is the least conclusive result in the “unclear” dataset. The word is short, so there are few adaptations to analyze, and effectively all are among the most common ones.

gavur ‘giaour’ | A *kāfir* id. ~ P *gabr* id. | GTS: <P; KEWT: <P ~ A; NS: <P; TDES: missing; TETTL: <P | $P(A)=0.001$, $P(F)=0.11$, $BF'_{AF}=-39.63$ | Here: <P⁺ | The particularly low score of the Arabic etymology is due to two adaptations: *f* > *v*, and *i* > *u*. But direct borrowing is in fact even more unlikely as the adaptation *k* > *ḡ* which it entails (*g* that is palatal despite being followed by a back vowel) is entirely unattested in the “foundation” dataset (and as such did not count towards the score here). The adaptations assumed by the Persian etymology are also some of the rarer ones but they are all attested in at least one example. Incidentally, the relation between the Arabic and Persian words is in fact unclear; see KEWT s.v., and further Paraskiewicz (2017).

his ‘feeling’ | A ~ P *hiss* id. | GTS: <A; KEWT: <P<A; NS: <A; TDES: missing; TETTL: <A | $P(A)=0.87$, $P(F)=0.86$, $BF'_{AF}=0.04$ | Here: ? | The rendering of *ss* as positionally geminated *s* (*his*, *-ssi*) is not attested in the dataset and had to be omitted here, effectively reducing this word to just two adaptations. Both are simply the most common ones in borrowings from Arabic and Persian alike, so no conclusion can be made based on them.

hokka ‘small box, small pot’ | A ~ P *huqqa* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing; TETTL: <P<A | $P(A)=0.23$, $P(F)=0.18$, $BF'_{AF}=1.62$ | Here: <A⁻ | The adaptation *qq* > *kk* is missing from the “foundation” dataset and so had to be omitted. The score was primarily informed by *u* > *o* which is rare in both Arabisms and Farsisms, but rarer in the latter. It is, however, attested in two borrowings from Persian (*durdī* > *tortu*, and *silāḥşur* > *silahşor*), so this result can only be considered a mild indication of a direct borrowing.

kavga ‘quarrel’ | A *ḡawḡā* ‘turmoil’ ~ P *ḡavḡā* id. | GTS: <P; KEWT: <(P<) A; NS: <P; TDES: missing | $P(A)=0.38$, $P(F)=0.4$, $BF'_{AF}=-0.48$ | Here: ? | The rendering of *ḡ* as *k* is generally unusual, but more so in an Arabism. It is not, however, impossible, as it is attested in three examples, all borrowings from Arabic (*ḡilāf* >

- kılıf, ġurnūq > kurnak, ŧamġ > zamk*). Had the Arabic etymon had an initial **k-*, BF'_{AF} would be 3.87, not only in its favour, but also much more decisively so.
- kös** I ‘kettledrum’ | A *kūs* ~ P *kōs* id. | GTS: <P; KEWT: <P>A; NS: <P; TDES: missing | $P(A)=0.02, P(F)=0.62, BF'_{AF}=-20.72$ | Here: <P⁺ | The decisive contributor here is *ū > ö* which is an adaptation entirely unattested in Arabisms. Its counterpart in the Persian etymology, *ō > ö* is also atypical, but it is supported by two examples, *ġōsāla > kösele*, and *kōsa > köse*.
- kullap** ‘hook’ | A *kullāb* id. ~ P *qulāb* id. | GTS: <A; KEWT: <A>P; NS: missing; TDES: <P<A | $P(A)=0.4, P(F)=0.27, BF'_{AF}=3.59$ | Here: <A⁺ | In terms of the score, the largest contributor is the adaptation of *-b* as positionally voiced (*kullap, -bi*), which is an indication of a direct borrowing from Arabic. Simultaneously working against the possibility of Persian mediation, is the fact that the adaptation *l > ll*, which only the Persian etymology entails, had to be omitted because it is not attested in any example in the “foundation” dataset.
- küme** ‘heap’ | A *kūma* id. ~ P *kuma* id. | GTS: no etymology; KEWT: <A ~ P; NS: <A; TDES: <A ~ P | $P(A)=0.31, P(F)=0.72, BF'_{AF}=-6.86$ | Here: <P[±] | Fronting of long *ū*, as assumed by the Arabic etymology, is considerably less likely than the fronting of a short *u*, as the Persian etymology requires. The other adaptations are all quite neutral.
- küp** ‘jug’ | A ~ P *kūb* id. | GTS: no etymology; KEWT: <P<A; NS: missing; TDES: missing | $P(A)=0.12, P(F)=0.24, BF'_{AF}=-4.43$ | Here: <P[±] | Fronting of long *ū* is overall a rare adaptation, but it is especially rare in Arabisms. Also the complete devoicing of final *-b* is more typical for a Farsism; in Arabisms, voicing tends to be preserved positionally in the presence of a vowel-initial suffix (**küp, -bü*).
- laden** ‘cistus; ladanum’ | A ~ P *lādan* id. | GTS: missing; KEWT: <A ~ P; NS: <A; TDES: missing | $P(A)=0.48, P(F)=0.51, BF'_{AF}=-0.68$ | Here: ? | No adaptation is here particularly unusual or characteristic for either borrowings from Arabic or those from Persian, so the result must be considered inconclusive.
- lal** ‘ruby; garnet’ | A *la^l* id. ~ P *lāl* id. | GTS: <A; KEWT: <(P<) A; NS: <A ~ P; TDES: missing | $P(A)=0.23, P(F)=0.2, BF'_{AF}=0.92$ | Here: <P⁻ | The rendering of *l* as palatal despite being adjacent to a back vowel, is a relatively more frequent adaptation in Arabisms than it is in Farsisms, and since it occurs twice in *lal*, it overpowers the lengthening of *a* which would be considerably more typical in a Farsism. Were one of the *l*’s adapted as dark *l*, BF'_{AF} would be -1.71 , and were they both, it would be -4.34 . Perhaps, Persian mediation is slightly more likely in this case despite the score.
- leġen** ‘washtub, washbowl’ | A *lakan* id. ~ P *laġan* id. | GTS: <P; KEWT: <P<A; NS: <A ~ P; TDES: <P | $P(A)=0.69, P(F)=0.53, BF'_{AF}=2.36$ | Here: <P⁻ | All adaptations are quite neutral here, except for the middle consonant. Strictly speaking, P *g* > T *ġ* is a relatively rare rendering, and it is it that tips the balance in favour of direct borrowing from Arabic, but two points must be taken into consideration. Firstly, this rendering is effectively just preservation of P *g* (the most common outcome), which only then turned into *ġ* simply because it happens to be positioned between vowels. Secondly, the counterpart in the Arabic etymology, *k > ġ*,

was entirely omitted in the calculation of BF'_{AF} because such an adaptation is not attested in any pair in the “foundation” dataset. In fact, there are no examples at all for the voicing of A *k*, intervocalically or otherwise. Considering all this, Persian mediation appears to be more likely than not, despite the score.

mabet ‘temple’ | A ~ P *maʿbad* | GTS: <A; KEWT: <(P < A) A; NS: <A; TDES: missing | $P(A)=0.36$, $P(F)=0.26$, $BF'_{AF}=3.99$ | Here: <A[±] | The major contributor here is the lengthening of *a*. This is a rare adaptation in Arabisms, and rarer still in Farsisms. Nearly all of the occurrences are where the *a* is adjacent to ʿ, which also explains why it is so unusual in Farsisms where ʿ is only pronounced as a glottal stop rather than a pharyngeal fricative. Also, positional preservation of voicedness (*mabet*, *-di*) is more common in Arabisms.

matem ‘mourning’ | A *maʿtam* id. ~ P *mātam* id. | GTS: <A; KEWT: <P<A; NS: <A; TDES: missing | $P(A)=0.5$, $P(F)=0.76$, $BF'_{AF}=-4.65$ | Here: <P⁻ | The Arabic etymology entails lengthening of *a* which is in general a rare rendering. In this case, the neighbourhood of ʾ could be suspected, and it would not count towards the score because here only phoneme-to-phoneme correspondences have been taken into account. However, ʾ does not exert as strong an influence as one might expect: out of eleven cases among Arabisms, lengthening occurred in seven. If the *aʾ* sequence were treated together, $P(A)$ would be 0.73, and BF'_{AF} would be -0.31, still in favour of Persian mediation, if only just.

maymun ‘monkey; ape’ | A ~ P *majmūn* id. | GTS: <A; KEWT: <A ~ P; NS: <A; TDES: missing | $P(A)=0.73$, $P(F)=0.62$, $BF'_{AF}=1.94$ | Here: <A⁻ | There is no single large contributor here. Preservation of back *a* is a little more likely in an Arabism, and shortening of long *ū* is just slightly more likely. All the other renderings are virtually neutral. The score reflects this situation, it suggests a direct borrowing, but somewhat tentatively.

mir ‘commander’ | A *amīr* ‘lord, emir’ ~ P *mīr* id. | GTS: <P; KEWT: <P<A; NS: <A ~ P; TDES: missing | $P(A)=0.23$, $P(F)=0.77$, $BF'_{AF}=-8.56$ | Here: <P⁺ | Deletion of *a* is in general a rare adaptation, and specifically apheresis is not attested in any example in the “foundation” dataset. The other renderings are essentially neutral.

mumya ‘mummy’ | A ~ P *mūmijā* id. | GTS: <P; KEWT: <A ~ P; NS: <A ~ P; TDES: missing | $P(A)=0.3$, $P(F)=0.34$, $BF'_{AF}=-1.5$ | Here: <P⁻ | Deletion of *i* is rarer in Arabisms than it is in Farsisms. To a much smaller degree, so is shortening of *ā*. Only the shortening of *ū* is marginally more typical for an Arabism than for a Farsism.

saniye ‘second (unit of time)’ | A *tānijāʾ* ‘second (numeral and unit of time)’ ~ P *tā-nije* ‘second (unit of time)’ | GTS: <A; KEWT: <(P < A) A; NS: <A; TDES: missing | $P(A)=0.73$, $P(F)=0.75$, $BF'_{AF}=-0.37$ | Here: ? | Apart from the preservation of long *ā* which is a little more common in Arabisms (*P ā* is shortened more often), practically all the adaptations in *saniye* are equally standard for both borrowings from Arabic and Persian.

santur ‘dulcimer’ | A *sanṭīr* id. ~ P *santūr* id. | GTS: <P; KEWT: <P<A; NS: <A ~ P; TDES: missing | $P(A)=0.35$, $P(F)=0.58$, $BF'_{AF}=-6.9$ | Here: <P[±] | Most adaptations

are relatively neutral here, but $A \bar{i} > T u$ is considerably less likely than $P \bar{u} > T u$. If the Arabic etymon had $-\bar{u}-$, BF'_{AF} would be in fact in its favour (2.67), though not as decisively.

- seme** 'stupid' | $A \sim P sama$ id. | GTS: <P; KEWT: <(P <) A; NS: missing; TDES: <A | $P(A)=0.76, P(F)=0.82, BF'_{AF}=-0.92$ | Here: ? | Effectively, only three adaptations are available here. The one that occurs twice, $a > e$, is slightly more typical for a Farsism, but the other two are neutral, and ultimately the result is inconclusive.
- sene** 'year' | $A \sim P sana$ id. | GTS: <A; KEWT: <(P <) A; NS: <A; TDES: missing | $P(A)=0.75, P(F)=0.84, BF'_{AF}=-0.91$ | Here: ? | The score has been mainly driven by the adaptation of a as e which occurs twice, but the difference between Arabisms and Farsisms is too small in this case to draw any conclusions.
- sihir** 'magic' | $A \sim P sihr$ id. | GTS: <A; KEWT: <(P <) A; NS: <A; TDES: missing | $P(A)=0.81, P(F)=0.64, BF'_{AF}=2.38$ | Here: <A⁻ | The main contributor here is the insertion of "positional" i (*sihir*, *-hri*). Consonant clusters are in fact more frequent in Persian etyma than in Arabic ones, both in auslaut and in other positions (43% of words vs 26%, and 8% vs 6%, respectively). Notwithstanding, "positional" epenthetic vowels such as the one here, are added considerably more often to loanwords from Arabic (22% of cases vs only 2%), which suggest a direct borrowing. Cf. also *yeşim* below.
- silah** 'weapon' | $A \sim P silāh$ id. | GTS: <A; KEWT: <(P <) A; NS: <A; TDES: missing | $P(A)=0.55, P(F)=0.54, BF'_{AF}=0.09$ | Here: ? | Shortening of \bar{a} is a little more typical in Farsisms, while the rendering of l as palatal \acute{l} despite back vocalism, a little more typical for Arabisms. No decisive conclusion can be made here.
- sinī** 'kind of tray' | $A \text{ } \acute{s}inī$ 'Chinese; id.' $\sim P sinī$ 'kind of tray' | GTS: <P; KEWT: <P <A; NS: <P; TDES: <P <A | $P(A)=0.67, P(F)=0.68, BF'_{AF}=-0.1$ | Here: ? | All the three renderings available here are virtually as common in Farsisms as they are in Arabisms. The rather low $|BF'_{AF}|$ reflects this uncertainty.
- şarap** 'wine' | $A \sim P şarāb$ id. | GTS: <A; KEWT: <(P <) A; NS: <A; TDES: missing | $P(A)=0.45, P(F)=0.35, BF'_{AF}=2.55$ | Here: <A⁻ | The main contributor here is the positional preservation of voicedness (*şarap*, *-bi*), an adaptation found more frequently in Arabisms than in Farsisms. Cf. *cevap* above.
- şey** 'thing' | $A \text{ } \acute{s}aj$ id. $\sim P \acute{s}ej$ id. | GTS: <A; KEWT: <(P <) A; NS: <A; TDES: missing | $P(A)=0.84, P(F)=0.88, BF'_{AF}=-0.36$ | Here: ? | All the renderings here are the most common ones in their respective groups, so no conclusions can be drawn from them.
- tabii** 'natural' | $A \sim P \acute{t}abī\acute{r}ī$ id. | GTS: <A; KEWT: <(P <) A; NS: <A; TDES: missing | $P(A)=0.57, P(F)=0.45, BF'_{AF}=2.9$ | Here: <A⁻ | No single adaptation is clearly the major contributor here. Overall, most renderings are a little more typical for borrowings from Arabic than for those from Persian.
- tacir** 'merchant' | $A \sim P tāğīr$ id. | GTS: <A; KEWT: <(P <) A; NS: <A; TDES: missing | $P(A)=0.8, P(F)=0.7, BF'_{AF}=1.43$ | Here: ? | All the renderings are very usual here. Combined, they may be a little more indicative of a direct borrowing than of Persian mediation, but the difference is not sufficient to consider the result conclusive.

- tarhun** ‘tarragon’ | A *tarḥūn* id. ~ P *tarḥūn* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: <A | $P(A)=0.72$, $P(F)=0.62$, $BF'_{AF}=1.85$ | Here: <A⁺ | In terms of frequencies of phoneme-to-phoneme correspondences, a more expected Turkish form would begin with **te-*. This is true for both Arabisms and Farsisms, but a little more so for the latter. However, it is only the Arabic etymology that provides a clear reason for the rendering with *a*, in the form of the preceding emphatic consonant. Also the shortening of *ū* is slightly more characteristic for borrowings from Arabic.
- tas** ‘bowl’ | A *tās* id. ~ P *tās* id. | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing | $P(A)=0.58$, $P(F)=0.74$, $BF'_{AF}=-1.45$ | Here: ? | This result hinges almost exclusively on the shortening of *ā*, and such an adaptation is a little more common in Farsisms. I do not think, however, that the difference is sufficient to view this result as decisive.
- temaşa** ‘observation; spectacle’ | A ~ P *tamāšā* id. | GTS: <P; KEWT: <A ~ P; NS: <A; TDES: missing | $P(A)=0.72$, $P(F)=0.66$, $BF'_{AF}=1.08$ | Here: ? | The largest contributor here is the preservation of length in both *ā*’s. This adaptation is a little more common in Arabisms, and since it appears twice in *temaşa*, BF'_{AF} slightly favours a direct borrowing. This is, however, scant evidence.
- temiz** I ‘clean’ | A *tamjiz* ‘distinction’ ~ P *tamiz* ‘clean’ | GTS: <A; KEWT: <(P<) A; NS: <A; TDES: missing | $P(A)=0.43$, $P(F)=0.8$, $BF'_{AF}=-6.78$ | Here: <P⁻ | A single deletion of *j* is not strong evidence, but it is a rare adaptation which the Arabic etymology does entail while the Persian one does not. Although beyond the scope of the present paper, semantics should also be mentioned here. The Arabic hypothesis entails a particularly non-trivial evolution; the Persian etymon has a virtually identical meaning to T *temiz* but, as the anonymous reviewer points out, it may be a semantic copy from Turkish.
- teşrin-i evvel** ‘October’ | A *tişrin al-awwal* id. ~ P *taşrin-i avval* id. | GTS: <A; KEWT: <P<A; NS: missing; TDES: missing | $P(A)=0.31$, $P(F)=0.74$, $BF'_{AF}=-18.3$ | Here: <P⁺ | There are three main arguments against the Arabic etymology here, and two of them count towards the score. The adaptation of the Arabic article in such a way that it sounds exactly like the Persian *izafet* would be a highly unlikely phenomenon, even from a purely phonetic point of view as *a > i* is a very rare adaptation, and deletion of *l* is entirely unattested in the “foundation” dataset (and as a consequence, omitted here because the frequencies of Turkish reflexes are unknown). The third quite unusual rendering required by the Arabic etymology is *i > e* in *teşrin*. Cf. *teşrin-i sani* below.
- teşrin-i sani** ‘November’ | A *tişrin at-tānī* id. ~ P *taşrin-i tānī* id. | GTS: <A; KEWT: <P<A; NS: missing; TDES: missing | $P(A)=0.27$, $P(F)=0.65$, $BF'_{AF}=-18.59$ | Here: <P⁺ | Mutatis mutandis, see *teşrin-i evvel* above.
- vezir** ‘vizier’ | A *wazir* ‘minister’ ~ P *vazir* id. | GTS: <A; KEWT: <(P<) A; NS: <A ~ P; TDES: missing | $P(A)=0.59$, $P(F)=0.48$, $BF'_{AF}=2.11$ | Here: <A⁻ | The main contributor here is the adaptation of *ī*. In Arabisms, when the colour of the vowel is not changed, length is also preserved in 51% of cases, whereas in Farsisms it is lost in 65% of words. Especially positional preservation, such as in *vezir*, *-īri*, is more frequent in Arabisms.

- vezne** ‘checkout; *arch.* weight’ | A *wazna* ‘weight’ ~ P *vazne* id. | GTS: <A; KEWT: <(P < A) A; NS: <A; TDES: missing | $P(A)=0.79$, $P(F)=0.87$, $BF'_{AF}=-0.94$ | Here: ? | The renderings here are practically neutral, except *a* > *e* which is slightly more typical for borrowings from Persian. It occurs twice in *vezne*, and pushes BF'_{AF} a little below 0, but it is weak evidence.
- yeşim** ‘jade; jasper’ | A ~ P *jaşm* id. | GTS: <P; KEWT: <A ~ P; NS: <P; TDES: missing | $P(A)=0.45$, $P(F)=0.72$, $BF'_{AF}=-4.81$ | Here: <P⁻ | It was mentioned in *sihir* above that “positional” epenthetic vowels are more frequent in Arabisms than in Farsisms. The same is true for “full” epenthetic vowels such as the one in *yeşim*, *-imi*, but not for specifically *i*. There are 42 examples of “positional” *i* (as in *sihir*, *-hri*), but only three of “full” *i*: A *laşm* > T *lehim*, *rahn* > *rehin*, and *sağl* > *sitil*. (In total, twelve Arabic etyma end in *-Cm*, and in all the remaining eleven the epenthetic vowel is “positional”.)
- zembil** ‘basket’ | A ~ P *zambıl* id. | GTS: <A; KEWT: <A ~ P; NS: <P; TDES: <A ~ P | $P(A)=0.7$, $P(F)=0.77$, $BF'_{AF}=-1.07$ | Here: ? | While in combination, the renderings here lean a little towards Persian mediation, the differences between their frequencies in the two groups are slight, and do not constitute reliable evidence.
- zencefil** ‘ginger’ | A ~ P *zanğabil* id. | GTS: <A; KEWT: <A ~ P; NS: <A ~ P; TDES: missing | $P(A)=0.16$, $P(F)=0.45$, $BF'_{AF}=-16.44$ | Here: <P[±] | By far the strongest argument against the Arabic etymology is the rendering of *b* as *f*. Such adaptation is entirely unattested among Arabisms; however, it is simultaneously only attested in just one Farsism (*çirkāb* > *çirkef*), so the strength of this evidence is less than BF'_{AF} makes it seem.
- zeytin** ‘olive’ | A ~ P *zajtün* id. | GTS: <A; KEWT: <A ~ P; NS: <A; TDES: <A ~ P | $P(A)=0.13$, $P(F)=0.61$, $BF'_{AF}=-18.23$ | Here: <P⁺ | There are no examples among Arabisms for the adaptation of *ū* as *i*, while among Farsisms there are three (*şaltūk* > *çeltik*, *lablabū* > *leblebi*, and *tarāzū* > *terazi*). The other renderings are neutral.
- zümrüt** ‘emerald’ | A ~ P *zumurrud* id. | GTS: <A; KEWT: <A ~ P; NS: <A ~ P; TDES: missing | $P(A)=0.1$, $P(F)=0.11$, $BF'_{AF}=-2.03$ | Here: ? | Deletion of *u* is unattested in Arabisms, and in only one Farsism (*şirrauğan* > *şırılağan*). On the other hand, shortening of geminated *rr* is attested in four borrowings from Arabic and none from Persian – it is however also true that *rr* only occurs in two Farsisms in total, so the available pool was rather limited to begin with. Lastly, positional preservation of voicedness in auslaut (*zümrüt*, *-dü*) is more frequent in Arabisms than it is in Farsisms.

3. Conclusions

The paper presents the results of the application of a quantitative method that estimates the phonetic probability of etymologies, to a group of 69 Turkish words which may have been borrowed directly from Arabic, or passed through Persian mediation.

Table 2 presents an overall comparison of these results, with the opinions of GTS, KEWT, NS, TDES, TETTL, and my own. They appear to be divided into

two groups. One comprises KEWT, the quantitative results, and my propositions (labelled “Here”), whereas GTS, NS, TDES, and TETTL represent the opposing camp. (Note that cases where both Arabic and Persian are given as a possible source have been excluded from the comparison as inconclusive answers.)

As for the quantitative results (BF'_{AF}) and my propositions, the two only diverge with regard to *lal* and *leğen*. The differences between my propositions and KEWT are in *ait* and *erbap*, and those between KEWT and BF'_{AF} in *ait*, *erbap*, *his*, and *leğen*.

KEWT did not give a definite answer for 48 out of 69 words in the “unclear” dataset. The method presented here has provided sufficient base for conclusions to be made for 30 out of those cases, and for 49 cases in total.

Overall, BF'_{AF} indicated Persian mediation in 58% of cases. It ranged from -39.63 (*gavur*) to 14.27 (*acur*), but 55% of words fell between -3 and 3 , and 71% between -5 and 5 (cf. Table 1). Levels of confidence of my own propositions, which were based primarily but not automatically on quantitative indices, were distributed relatively evenly. I reached no conclusion in 20 cases, a hardly certain conclusion also in 20, a fairly certain one in 13, and a quite certain one in 16.

	GTS	KEWT	NS	TDES	TETTL	Here	BF'_{AF}
GTS	—	44%	93%	71%	77%	64%	58%
KEWT	8/18	—	33%	75%	45%	89%	81%
NS	43/46	4/12	—	60%	77%	59%	56%
TDES	5/7	3/4	3/5	—	100%	80%	50%
TETTL	23/30	5/11	20/26	2/2	—	50%	50%
Here	29/45	17/19	19/32	4/5	13/26	—	96%
BF'_{AF}	37/64	17/21	28/50	4/8	16/32	47/49	—

Table 2. Level of agreement between various sources. In the lower triangle, the number of cases where both sources give the same definite answer, by the number of cases where they both give any definite answer. In the upper triangle, the result of the division as a percentage.

The intention behind the method used here is not for its results to be followed blindly. Not only the final estimated phonetic probability needs to be analyzed by a human, but also its component elements, as has been done in Section 2. When approached with such caution, such indices constitute substantial, objective phonetic evidence that cannot be obtained through intuition alone.

Abbreviations

A = Arabic; *arch.* = archaic; *esp.* = especially; P = Persian; *pej.* = pejorative; T = Turkish

References

- Eren H. 1999. *Türk Dilinin Etimolojik Sözlüğü*. [2nd edition]. Ankara: Bizim Büro Basım Evi.
- GTS = Türk Dil Kurumu. *Güncel Türkçe Sözlük*. [available at: <https://sozluk.gov.tr>; accessed in: May 2020].
- Kass R.E., Raftery A.E. 1995. Bayes Factors. – *Journal of American Statistical Association* 90.430: 773–795.
- KEWT = M. Stachowski 2019.
- Nişanyan S. [on-line]. *Nişanyan Sözlük*. [available at: <https://www.nisanyansozluk.com>; accessed in: May 2020].
- NS = Nişanyan [on-line].
- Paraskiewicz K. 2017. In search of *giaoour*. Notes on the New Persian *gabr* ‘a Zoroastrian; infidel’. – Németh M., Podolak B., Urban M. (eds.). *Essays in the history of languages and linguistics. Dedicated to Marek Stachowski on the occasion of his 60th birthday*. Kraków: Księgarnia Akademicka: 473–481.
- R = R Core Team. 2020. *R: A language and environment for statistical computing*. [R Foundation for Statistical Computing, Vienna, Austria; available at: <https://www.R-project.org>].
- Stachowski K. [forthcoming]. Estimating phonetic probability in etymology.
- Stachowski K. 2020a. Phonetic renderings in Turkish Arabisms and Farsisms. – *Türkbilig* 40: 23–47.
- Stachowski K. 2020b. Tools for semi-automatic analysis of sound correspondences: The *soundcorrs* package for R. – *Glottometrics* 49: 66–86.
- Stachowski M. 2019. *Kurzgefaßtes etymologisches Wörterbuch der türkischen Sprache*. Kraków: Księgarnia Akademicka.
- TDES = Eren 1999.
- TETTL = Tietze 2002–2009.
- Tietze A. 2002–2009. *Tarihi ve Etimolojik Türkiye Türkçesi Lugatı. Sprachgeschichtliches und etymologisches Wörterbuch des Türkei-Türkischen*. [vols. 1–2: A–J]. İstanbul: Simiurg, Wien: Österreichische Akademie der Wissenschaften.
- TRT = *TRT Telaffuz Sözlüğü: Genel Sözlük*. [available at: <http://trttelaffuz.com>; accessed in: May 2020].

Appendix

Some borrowings from Arabic display adaptations that are more typical for loanwords from Persian, and vice versa. In the case of 5% of Arabisms and 6% of Farsisms, the irregularity is sufficient to produce a BF'_{AF} below -2 or above $+2$, respectively, i.e. what constitutes “positive evidence” in favour of the alternative hypothesis (cf. Table 1). The reasons for this are probably varied. Some words may have been borrowed from a dialect, or conversely, to a dialect and only from it penetrated into the literary language; others may have passed through Arabic or Persian mediation; others still may have been contaminated with a different word on the Turkish ground, etc. They are, at any rate, cases that require further investigation.

Below is a list with the BF'_{AF} ’s but without any analysis. The probabilities of individual adaptations can be found in K. Stachowski (2020a), but it is likely that few of those words can be solved with phonetics alone, without the help of philological and historical evidence.

Borrowings from Arabic:

- qīrāt* > *kırat*: -2.04
manqabaī > *menkibe*: -2.04
majdān > *meydan*: -2.07
kursī > *kürsü*: -2.08
ḥandaq > *hendek*: -2.09
walḥāşil > *velhasil*: -2.12
murtad > *mürtet*: -2.13
ḥarūn > *harın*: -2.15
malik > *melik*: -2.21
qanāʿat > *kanaat*: -2.32
ḥalwā > *helva*: -2.40
taḍkiraī > *tezkere*: -2.41
ijālā > *eyalet*: -2.44
iğtimāʿ > *içtima*: -2.55
falak > *felek*: -2.56
maslak > *meslek*: -2.57
dimāğ > *dimağ*: -2.65
manğaniq > *mancınık*: -2.84
anāi > *enayi*: -2.85
muştarak > *müşterek*: -2.89
iṭāʿat > *itaat*: -2.92
surʿat > *sürat*: -2.95
fatil > *fıtil*: -2.99
qabāḥat > *kabahat*: -2.99
ḥamāʿil > *hamail*: -3.05
faşāḥat > *fesahat*: -3.06
dallāl > *tellal*: -3.08
qatīfat > *kadife*: -3.09
qadr > *kadar*: -3.24
ḥalāʿiq > *halayık*: -3.25
ḥanğar > *hançer*: -3.31
farāğat > *feragat*: -3.33
ḍamm > *zem*: -3.42
ğamāʿat > *cemaat*: -3.61
sāʿat > *saat*: -3.61
hiğāʿ > *hece*: -3.74
dukkān > *dükkan*: -3.84
dallāk > *tellak*: -3.89
sātūr > *satır II*: -3.97
handāsaʿ > *hendese*: -4.20
qīmaʿ > *kıymet*: -4.33
ṭālīʿ > *talih*: -4.41
şūfi > *sofu*: -4.44
ğabran > *cebren*: -4.50
rahn > *rehin*: -4.69
mustamlak > *müstemleke*: -4.73
mumkin > *mümkün*: -4.88
qāḍi > *kadı*: -4.96
ajlül > *eylül*: -5.01
albatta > *elbette*: -5.17
māʿi > *mavi*: -5.26
muşkil > *müşkül*: -5.26
laḥm > *lehim*: -5.33
ʿajbaī > *heybe*: -5.62
kattān > *keten*: -5.63
karbatān > *kerpeten*: -5.72
sāʿis > *seyis*: -6.70
saṭl > *sıtıl*: -6.72
sijāḥat > *seyahat*: -6.90
saʿtar > *zahter*: -7.27
şamğ > *zamk*: -7.27
qurʿān > *Kur'an*: -7.40
ğallābi > *çelebi*: -9.73

Borrowings from Persian:

- siñğāb* > *sincap*: 2.02
zahr > *zehir*: 2.02
direğ > *diriğ*: 2.05
zamīn > *zemin*: 2.06
ḥurda > *hurda*: 2.07
didār > *didar*: 2.12
tirjākī > *tiryaki*: 2.15
girdāb > *girdap*: 2.19
zarrīn > *zerrin*: 2.24
ḥasta > *hasta*: 2.34
taḥta > *tahta*: 2.35
giraç > *kireç*: 2.38
zağara > *zağara*: 2.38
ğihān > *cihan*: 2.46
feğan > *figan*: 2.47
naçar > *naçar*: 2.50
jārān > *yanar*: 2.52
hafta > *hafta*: 2.53

ta'limḥāna > *talimhane*: 2.55
ḥāqān > *hakan*: 2.63
baḥir > *bakır*: 2.65
bahār > *bahar*: 2.87
ḡānān > *canan*: 2.97
šalgām > *şalgam*: 3.08
farjād > *feryat*: 3.23
sīmurḡ > *simurg*: 3.26

čōb > *cop*: 3.28
āb > *ab*: 3.54
nalband > *nalbant*: 3.57
lāḡar > *lagar*: 4.57
nawrūz > *nevruz*: 4.65
taḥtapūš > *tahtaboş*: 5.78
rovḡan > *rugan*: 10.32