Linguistic Convergence and Areal Diffusion
Case studies from Iranian, Semitic and Turkic

Edited by
Éva Ágnes Csató, Bo Isaksson and Carina Jahani
# Table of Contents

## List of Contributors
Preface

## Introduction
Lars Johanson  Converging Codes in Iranian, Semitic and Turkic

## Part 1: Iranian Languages

### Donald L. Stilo

Iranian as Buffer Zone Between the Universal Typologies of Turkic and Semitic

### Bo Utas

Semitic in Iranian: Written, Read and Spoken Language

### Carina Jahani

The Glottal Plosive: A Phoneme in Spoken Modern Persian or Not?

### John R. Perry

Lexical Areas and Semantic Fields of Arabic Loanwords in Persian and Beyond

### Gernot L. Windfuhr

Central Asian Arabic: The Irano-Arabic Dynamics of a New Perfect

## Part 2: Semitic Languages

### Guram Chikovani

Linguistic Contacts in Central Asia

### Otto Jastrow

Uzbekistan Arabic: A Language Created by Semitic-Iranian-Turkic Linguistic Convergence

### Robert R. Ratcliffe

Bukhara Arabic: A Metatypized Dialect of Arabic in Central Asia

### Sven-Olof Dahlgren

On the Arabic of Arabkhane in Eastern Iran

### Bruce Ingham

Persian and Turkish Loans in the Arabic Dialects of North Eastern Arabia

### Bo Isaksson

New Linguistic Data from the Sason Area in Anatolia

### Stephan Procházka

The Turkish Contribution to the Arabic Lexicon
Lexical Areas and Semantic Fields of Arabic Loanwords in Persian and Beyond

John R. Perry

Toward a systematic study

The most striking, and culturally pervasive, example of lexical convergence in our area is undoubtedly the massive incorporation of Arabic vocabulary into Persian, Turkish, etc., that followed swiftly upon the adoption of Islam and of the Arabic writing system after the 9th century AD. A dictionary-based sample yields an inventory of approximately 8,000 Arabic loanwords in current standard Persian or about forty percent of an everyday literary vocabulary of 20,000 words, not counting compounds and derivatives (Rāzi 1987). Corpus-based inventories and frequency counts vary with historical period, style and topic from less than 9 percent and 2.4 percent frequency (Shāhnāmah) to more than 50 percent with almost 25 percent frequency in both Classical and Modern Persian literature.1

It is generally understood that the bulk of the Arabic vocabulary in the central, contiguous Iranian, Turkic and Indic languages was originally borrowed into literary Persian between the ninth and thirteenth centuries CE as mots savants, and transmitted thence into the written registers, subsequently the vernaculars, of Ottoman and Chaghatay Turkic, Kurdish, Pashto, Hindi, Bengali, etc. (Lazard 1965; Perry 1984). Although there is some variation in its distribution among languages and dialects; although switches of script and lexical reforms have accelerated change within living memory; and although Arabic is no longer a living lexical fount for any of these languages today—they still share a broad, readily-identifiable stratum of Arabic loanwords, mainly nouns and adjectives, both as individual words and in transparent combination with Persian and other native morphs.

Existing studies have listed such loanwords by the thousands and, in individual cases, have recorded their early use or recent demise or their difference in meaning from their etymons in Arabic. A few have attempted to progress beyond ad hoc etymology to statistical corpus-based studies (notably Lazard

---

1 See Perry (1991b: 203–206), additionally citing studies by Koppe, Moīn̄far, Osmanov and Utas. Throughout this study, transcription of the vowels of loanwords in Persian partially reflects Persian pronunciation, and they are thus intentionally distinguished from their Arabic etymons and other references in Arabic.
ARABIC LOANWORKS IN PERSIAN AND BEYOND

John R. Perry

Preconditions for the adoption of Arabic elements into late Middle Persian were highly favorable. The Arabic spelling, or at least the pronunciation, of Arabic words, was crucial for the development of new Persian words. The phonetic form of the Arabic words was adopted, and the pronunciation of the Arabic words was adapted to the Persian phonetic system. This process was facilitated by the Arabic loanwords being used in the same contexts as the corresponding Persian words. The Arabic loanwords were also integrated into the Persian grammar and syntax, and the Persian speakers were familiar with the Arabic script and pronunciation.

Social modalities of borrowing

Morpho-phonological strata

Two other sound shifts of late Middle Persian exerted arguably more important effects on the assimilation of Arabic loanwords. These were:
Versatility of the feminine ending

The dichotomy -at/-ah in this class of borrowings from Arabic is not random, and affords other significant clues to the sorting and processing of Arabic vocabulary in Persian. The grammatically feminine marker in Arabic is realized phonetically as either -at/ (in pre-juncture position) or -al (pausal form), according to the contextual syntax of Arabic, but written with a single hybrid graph (the tā’ marbūta). This syntactically determined variation in spoken Arabic was irrelevant to Persian, which is uninflected for grammatical gender. It requires, on the contrary, that loanwords be lexicalized definitively in one form—in this case, either with or without final i. Why, then, of some 1450 such loans in current Persian, do 640 end in -at, and 810 in -ah, including some 40 pairs of doubles (80 items) lexicalized with both endings? Moreover, in earlier centuries, this proportion was virtually reversed, with a preponderance of -at affiliates in the inventory: which means that more than 200 original -at affiliates have since dropped final i, and thereby shifted to the -ah category.

The short answer (for the long answer, see Perry 1991b), lies in a phenomenon called “exaptation” (a term taken from biology), whereby a feature that functioned in one system, and is made redundant on transfer to a different system, may be adapted to function in a quite different capacity within the new system. Just as the flotation-bladder that had served a marine organism was able to evolve into lungs for a land animal, so this Semitic syntactomorph has been unconsciously reinvented as a Persiane semanteme. Distribution between -at and -ah of the modern Persian inventory appears to be determined primarily by semantic features, and additionally by factors of syntactic and stylistic environment or historical evolution of the words. These processes are particularly apparent in the doubles. These may represent either disambiguated homonyms of Arabic: resālat ‘mission’ vs. resālah ‘message, letter, dissertation’; sarārat-ah (Fig. 1, No. 8), or semantic specializations in Persian: erādat ‘wish, goodwill’, erādah ‘resolution, edict’, or in the languages beyond: harakat-ah in Turkish (Fig. 1, No. 5; for further examples, see Perry 1995). An early instance, which exhibits a semantically motivated shift dating from the Persian works of philosphers such as al-Ghazali (d. 1111 AD), is govvat/govvah:

govvat ‘strength, power’ (general, intangible; mass noun).

govvah ‘a (physiological or mental) faculty’ (count noun: govvah-ye bāsherah ‘the sense of sight’); 11th century.

‘a (military) force’ (pl. govvah-hā, govā); 19th century.

‘(industrial) energy’ (govvah-ye hārāq ‘electrical power’); 20th century.

In the course of the next several centuries, hundreds of the -at class shifted to the -ah class, some leaving behind traces as doubles in -at. In general, the

---

2 This long i will not be shown with a macron in the transliterations from New Persian and Arabic below, for the sake of a unified transliteration system.
phonetic and orthographic change recapitulates the original rationale, as can be seen by reference to Fig. 1: the resulting -ah words are semantically more particularized (Nos. 1, 2, 3, 6), specialized (5, 7) concrete and/or more firmly established in the vernacular (Nos. 6 and 7).

The “semantic spectrogram” (Fig. 1) illustrates the typical distribution by -at and -ah across this semantic and sociolinguistic spectrum (see Perry 1991b: 195–224). Occurrences of doublets of each type (denoted by t and h respectively) represent ranges of connotation of the word in this form as found in Classical and Modern Persian (the default language, though the borrowing is usually available in at least some senses in the other languages). Thus nouns with more abstract and intangible, or less imageable, referents tend to end in -at. Typical examples not occurring as doublets are xošnat ‘aspiration, roughness’ (quality noun), xošsat ‘permission, leave’ (action noun, “+N”), xoš̄at dādan ‘to dismiss, give leave’ (“+V”); they are more often mass nouns than count nouns, and learned or literary rather than everyday words. Nouns with more concrete, tangible, imageable and countable referents (more likely to appear in the plural, and more frequent in use) tend to end in -ah: molāhezh ‘note, remark’ (instance noun), nox̄ah ‘prescription, text’ (instance and product noun), mahallah ‘place, neighborhood’ (countable entity noun, since as a loanword it has lost any transparent relation to its verbal derivation in Arabic).

In Fig. 1, it can be seen (No. 1) that the shifted form ešārah in modern Persian came to be used as a verbal component (ešārah kardan ‘to point, indicate’) and as a countable noun of instance and product (‘an indication, sign, token’). The specialized nisba forms and meanings of Nos. 2 and 4 are examples of dozens of derivatives from simpler substantives which enhanced the technical and intellectual vocabularies of Persian, Turkish and Urdu in the nineteenth century especially (see Perry 1991b: 23–40, 176–78): the pair mānāviyat-ah are universally available, whereas ešāriyat-ah are confined to Urdu. In No. 5 the specialized sense ‘vowel diacritic’ is common to Persian, Turkish and indeed all languages that use, or have used, Arabic script; but only in Turkish is it differentiated from the general sense ‘movement’ by a shift to -ah. The shifted reflex mas‘alah ‘question, matter, problem’ (No. 6) is widely distributed in the pertinent languages; in modern Persian, it has an additional use as a euphemism for the male sex organ (Jamālzādah 1962). No. 8 illustrates a disambiguation of an original Arabic homonym (the coincidental surface representation of two different concepts encoded in the same consonantal root). In No. 9 we see the ingenious disambiguation of three words generated from two distinct (but formally identical) morphological patterns of Arabic: the augmented action noun (mašdar mīni), which produces maymana ‘good fortune, etc.’ and the locus noun (ismu‘l-makān), which gives both maymana ‘right wing (of an army)’ and the name of a city in Afghanistan (see Perry 1991b: 45–47).
The role of early lexicographers

The awesome ingenuity of the root-and-pattern system of Arabic lexical morphology evident in the bilingual traditions of eastern Classical Islamic grammarians and lexicographers. More to the point, a Persian dictionary, which, despite being as Pettid's (1993) argued, produced from the 11th century onwards, never left home except to go to the mosque. Students instead went to Nishapur to memorize his or her parts of the script.

Varno by language and style

Not only were personal loans incorporated into Persian then passed on to Turkish, Hind-Irani, and Urdu languages. Even more, the category of non-Persian loans in the following dialects of Afghanistan and Tajik of Central Asia and Hind-Irani, Pashto, Persian, Turkish, and Urdu.

3) These reflexes are not without a language affiliation in parenthesis exist in this form and meaning also in Persian. A total of fifty examples are given in Perry (1991b, 120).
Assimilation of form classes

As an example of what might nevertheless be done with the help of dictionaries, I turn finally to a case study of the modern Persian inventory of 166 loans of the pattern mofā'elat-ah and mofā'āt (one of the maṣādir of Theme III of the Arabic triliteral verb), set against a comprehensive Arabic inventory of 883 words of this class. Bu Ja'farak listed 749 of them, an impressive 85 percent.

This pattern notably encodes, among other semantic modalities, those of competition, co-operation and reciprocify which are realized lexically in three archetypical human activities above all else: love, war and trade (or sex, conflict and business; and ‘business’ I expand to mean all forms of interaction and negotiation in the contexts of law, politics and gainful work). Thus in the following epitome of the scheme, Persian moqārebah ‘sexual congress’, mojādelah ‘quarrel’, mo'āmelah ‘operation, deal’ each belong to a cluster of assonant near-synonyms which collectively define the bulk of humankind’s interactive pursuits. (The symbol (!) denotes polysemous items that qualify for more than one category.)


Obviously it is not practicable here to list and compare all items, Arabic and Persian, in this closely related complex; I shall confine comment to the illustration of trends that appear to be supported by the data. It is evident at the outset that only a fraction of the potential Arabic vocabulary has been assimilated into—or at least, has survived in—Persian (less than 20 percent; it would be instructive to calculate similarly what proportion of other well-defined formal and semantic classes has been incorporated). From Fig. 2 it may be seen that the proportion of the smaller Persian inventory having to do with the three thematic areas is directly comparable with that of Arabic, at almost half the total mofā’ala class (49 percent—despite some differences in meanings, and in numbers of relevant meanings per lexeme, between the two language inventories).

Fig. 2. Mofā’ala (inc. mofā’āt) in Arabic and Persian

<table>
<thead>
<tr>
<th>Arabic mofā’ala</th>
<th>“Love”</th>
<th>“War”</th>
<th>“Trade”</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>7.6%</td>
<td>179</td>
<td>20.3%</td>
<td>433/883; 49%</td>
</tr>
<tr>
<td>Persian mofā’elat</td>
<td>16</td>
<td>9.6%</td>
<td>25</td>
<td>15%</td>
</tr>
</tbody>
</table>

 Obtained principally from Steingass (1884), the only comprehensive Classical Arabic (Eng-lish) dictionary arranged alphabetically by words, not roots.
In comparing classes of Persian loanwords with their Arabic etymons, it is of course not enough to identify forms and compute statistics. Meanings change, polysemy is exploited, and one must be on guard against faux amis. An example in the mofāʿala category is furnished by Arabic murābāta ‘lining up in battle order; military station’ vs. Persian morābētāh ‘connection, communication’. Though the formal identity of the Persian word is not in doubt, its meaning bears little relation to that of its Arabic “etymon”. Most probably the Persian word is not actually derived from murābāta, but is a plausible invention suggested by the genuine loanword rābeṭaḥ ‘tie, connection, liaison, relation’ < Arabic rābiṭa ‘ditto’. It is therefore counted among Persian mofāʿalāt loans, though its thematic category will be different (“trade”) from that of its Arabic cognate (“war”).

Within this scheme, then, the three thematic areas (Fig. 2) show noticeably different proportions as between the two languages: in Persian, the volume of the vocabulary of “trade” scores 2.8 percentage points more than in Arabic, that of “love” two points more, that of “war” 4.7 points less. It is doubtless simplistic to infer from this that the Persians are more businesslike than the Arabs and prefer to make love, not war. It may, however, bolster the reasonable assumption that Arabic synonyms and lexical nuances on the theme of war (as are conspicuous by their absence in the overwhelmingly Persian vocabulary of the Shāhnāma) were less in demand than in the context of love (as increasingly employed in the lyrical and mystical verse of later centuries).

Fig. 3. Persian mofāʿelat – mofāʿelah

<table>
<thead>
<tr>
<th></th>
<th>“Love”</th>
<th>“War”</th>
<th>“Trade”</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>-at</td>
<td>8</td>
<td>10</td>
<td>3</td>
<td>21/56*</td>
</tr>
<tr>
<td>-ah</td>
<td>6</td>
<td>14</td>
<td>36</td>
<td>56/93*</td>
</tr>
<tr>
<td>Shift -at &gt; -ah</td>
<td>5</td>
<td>13</td>
<td>22</td>
<td>40/66/93</td>
</tr>
</tbody>
</table>

*From a total Persian inventory of 149, i.e. including 20 doublets, but excluding 17 reflexes of mofāʿāt investigators for Sa’dī’s Golestān (1258 AD), we find a preponderance of -at by 35 to 19, or almost 65 percent.

Thus despite the near-total orthographic uniformity of Arabic etycons and Persian borrowings, there is the dichotomy of feminine-ending forms a wealth of clues to the progressive assimilation of whole classes of Islamic loans into Persian, and to their further distribution and lexical evolution in the other Iranian, Turkic and Indic languages of the central Islamic umma. There are also significant tendencies in the semantic evolution and distribution of other morpholexical classes of Arabic vocabulary that might assist in mapping and dating their incorporation into Persian, their further distribution, and their contributions to the cultural contours of Turco-Persian and Indo-Persian societies.

References