PERSIAN-ARABIC BILINGUALISM IN THE EVOLUTION OF NEW PERSIAN

By JOHN R. PERRY, Chicago

I

Bilingualism is obviously an important factor in the transmission of lexical and other influences from one language to another. But it is not the only factor, and bilingualism itself may be of different kinds and degrees. During the formative period of New Persian (ca. 8th-11th centuries) we have little direct evidence for the operation of any of these factors: we must begin any speculations with assumptions about the relations between pre-literate Islamic Persian and Arabic during the first two centuries of caliphal rule in Iran, bolstered by observations of established literary features in manuscripts from the twelfth century onward.

We may assume that a store of basic Arabic loanwords pertinent to the imposition of Arab rule and conversion to Islam was acquired by Persian speakers during the first few decades of the conquest (ca. 650-700C.E.). This vocabulary would have been transmitted orally to Persian speakers who were not necessarily literate or bilingual. The phenomenon is well documented in more recent colonial processes, such as the expansion of British dominion in India or Russian rule in Central Asia. Vocables from this early stage tend to be personal and cultic names, military and administrative terms, luxuries and curiosities imported by the dominant community. Such experiential fields are not in themselves reliable indicators of an early date for a loanword: in some fields (notably religion) the very success of the colonizing or evangelizing process favors translation by analogy (xodā, namāz, ferešta, etc.) over borrowing--loans appear later, with bilingual native clerics. A better index is phonological and morphological assimilation to the recipient language, such as can be seen in P. mosalmān (< ?muslim[īn]), mollā (< mawlid[nā]), bu (< abu), mir (< amīr) and its compounds such as mirzā, mirāb, mirāxor--all, incidentally, from the appropriate experiential fields. This first stratum of Arabic vocabulary in New Persian, dating from the monolingual and preliterate period (i.e., before Persian was commonly written in Arabic characters), we may refer to as the "vernacular stage."

Our second assumption is that the bulk of the immense Arabic vocabulary in Persian was introduced progressively by literate bilingual Iranian intellectuals from the late ninth to the twelfth century. Since direct Arab rule no longer extended into Iran, this vocabulary entered in the form of mots savants in all fields of Persian literature and trickled down into vernacular usage.1 In contradistinction to the words of the vernacular stage, these loans were mined not from a living Arabic vernacular but from classical literary models, and moreover consciously avoided any suggestion of assimilation to norms of Persian phonotactics or morphophonemics. This stratum of borrowing from Arabic may be designated the "literary stage."

There were thus two potentially antithetical processes at work during the formative period of New Persian: a natural tendency on the part of illiterate and/or monolingual

Iranians to assimilate foreign loanwords to the phonology of the Persian vernacular, and an artificial effort by Iranian Muslim scholars to preserve the orthography of Arabic etyma. The latter process, which we might call "scripturalism," was presumably motivated not only by professional pride and pedantry but by the sacred status of Arabic characters and words in themselves, as the substance of holy writ. Each tendency can be said to have prevailed, for the modus vivendi that soon evolved (and which survives to this day) stipulated a pronunciation of Arabic in accordance with the Persian sound system, but a spelling of loan vocabulary in accordance with the Arabic originals.

The few partial exceptions--fossilized forms in which a process of scholarly normalization was arrested or one of vernacularization was reversed--serve to illustrate the tensions and compromises between these rival strategies. An example of successful retro-vernacularization is P. ḥalā 'now', originating in an Arabic tanwin accusative such as would normally be pronounced /-an/ (as in the loans taqriban, nesbatan, etc.), but assimilated to the vernacular to the extent of a spelling- pronunciation of the anomalous alef (cf. Afghan and Tajik P. ḥālā, a more emphatic--or perhaps an original?--vernacularization in which the Arabic orthography is ignored). As an instance of scripturalism triumphant one may cite the well-known case of A. xidma[t] > early N.P. xezmat > P. xedmat 'service'. The spelling with dotted dāl as found in earlier Persian manuscripts represents an analogy with the spelling of what is now postvocalic /d/ in native Persian words, and which at that time presumably had a spirantized allophone /ð/ that was optionally written with the Arabic gāl. This widespread orthographic vernacularization of an Arabic loanword is a unique episode, and was not permitted to continue beyond the twelfth century: literary Persian usage (followed by Ottoman Turkish and Urdu in turn) reverted to the canonical Arabic spelling with dāl. Spoken Persian has since accepted the pronunciation suggested by this normative spelling. Nevertheless, the vernacular reflex has survived as a variety of /xzmat/ in several Turkic and Iranian vernaculars of the region (e.g., modern Turkish, showing a rejection of the early Ottoman literary norm; Tajik, extending the analogy to the derivative /xizmatgār/ 'servant'; and Baluchi, where the latter word is usually /xzmatgār/, exhibiting the acrolectal reflex through literary Urdu). This normalizing tendency in favor of Arabic orthography probably explains the dearth of surviving vernacularized loans: all but the earliest and best established were "reformed" during the eleventh century and (with a few exceptions such as abu and amir) left no trace of their lower-class antecedents.

In what follows I shall attempt to look beyond the evidence of individual loanwords, and survey whole phonological and morphological classes of Arabic vocabulary in Persian where I believe we can discern these conflicting tendencies at work, together with other systematic influences of Arabic on Persian and Persian on borrowed Arabic. In particular I hope to characterize the effects of Persian-Arabic literary bilingualism, and more especially the limits of its role, in these processes.

\[^2\] Meier 1981, 103-5, 111; Pisowicz 1985, 107-8.
\[^3\] Rossi 1979, 225.
I begin with a phonological development of late Middle Persian that resulted in a morphological coincidence with two important classes of Arabic loans. This demonstrably produced several Perso-Arabic morphological blends, and arguably facilitated and encouraged the incorporation of Arabic words of these classes at an early date. By the seventh century C.E., postvocalic final /k/ had tended to become voiced and ultimately to be dropped (/-Vg/ > /-V0/). Thus, in the case of the important adjective and type-noun formative /-ik/, MP. parik > NP. pari 'fairy', and tāzik > tāzi 'Arab'. Arabic borrowings such as zindiq < MP. zandiq 'heretic' attest that this process was not yet completed on the eve of the Arab invasion. The NP. variant tāzik 'Muslim' Iranian', and several common words such as bārik, tārik, nazdik have resisted both changes. This final -ik was in most cases an adjectival suffix, semantically equivalent to--and, after the sound change, practically identical with--the Arabic nisba adjective in -i; the etymology of an early NP. adjective such as Kermāni would thus be indeterminate and, indeed, irrelevant. This process of morphophonological blending can be captured in the pattern

\[ pārsik > pārsi \quad [ > fārsi \quad \Rightarrow A, \ fārsi 'Persian' \]

--ratified in this case by gratuitous Persian adoption of Arabic /f/ for Persian /p/. On both chronological and semantic grounds, there seems every reason to suppose that the sound change in Persian facilitated the oral transmission of Arabic nisba adjectives and nouns, and that in turn the massive influx of nisba loans accelerated the sound change and confirmed this new ending as a highly productive formative in New Persian. As the principal adjective suffix, it is responsible for the derivation of hundreds of words in an open and expanding lexical class.

In the other principal form-class of Middle Persian nouns/adjectives subject to this same loss of the final velar, the argument for a morphological blend is perhaps not so obvious or convincing; yet the suggestion is strong that besides random fossils such as mir there is a substantial number of a whole class of Arabic loanwords whose morphophonological adaptation in Persian provides a clue to their origin as early vernacular forms transmitted orally without bilingual input. The class is that of loanwords ending in /el/ (modern standard Persian), /el/ or /al/ (eastern dialects and Classical Persian), written with the "silent k" (hā-ye moxtaš) and originating in Arabic etymology ending in the tā' marbūta. The kind of words I have in mind are nouns and adjectives denoting basic tangibility or cultic and administrative concepts imported with Islam, such as madrasa 'school', manāra 'tower', am[m]āna 'urban', sekkā 'coin[age]', atiya 'gift', ariza 'petition', xatna 'circumcision', tawba 'repentance', sefla 'vile'. The semantic and experiential fields of these words (tangibility, deverbal nouns of instance, product, agent and patient, unitary nouns, diminutives, etc.) correspond to a considerable degree with those of the class of native Middle Persian words ending in /ag/, which around the period of the Arab invasion dropped the final velar (in accordance with the same rule mentioned above) to produce, e.g., xāna, nāma, bacca, busa, nāvestā, zenda, tāza, bahāna, etc. That the two classes were immediately identified morphologically in Persian is shown by frequent early

\[ 4 Geiger & Kuhn 1895, I/2:13, 63-4; Piaowicz 1985, 139-41. \]
"restoration" of the velar to Persian derivatives of such Arabic loans, by analogy with the same derivatives of Persian etyma (sefjagān, nazzāragān, naz[z]āragi, bi-salīqāgi, xezragān, etc.). Their subsequent phonological history was also shared, most recently the raising of the final vowel /æ/ > /e/ in Tehran Persian. It seems reasonable to postulate that, as in the case of the nisba adjective, a sound change in this Persian form-class contemporaneous with the influx of numerous semantically and morphologically similar Arabic words (marked, moreover, by a lexically unstable ending that was resolvable as the same open vowel as was favored for the Persian form-class) resulted in a degree of morphological blending; this accelerated the processes of fusion and assimilation, and created a lexical form-class hospitable to the incorporation of all subsequent tāʾ marbūta loans in the Persian /-a/ inventory (a modern total of some 800).

III

The question of the incorporation of feminine-ending Arabic nouns in Persian, i.e., the realization of tāʾ marbūta as either final t (tāʾ mamdūda) or vocalic h, representing final /at/ or /a/, is complex and many-faceted. Insofar as it concerns the vernacular and literary stages of the incorporation of Arabic vocabulary, the following observations are relevant.

The /at/ ending is originally inspired by the contextual realization of the Arabic feminine ending (before a syntactic juncture, primarily the first element of a status constructus) in, e.g., sunnata-l-ʾawwalīn 'the custom of the ancients'; such phrases form lexical sources for literate bilingual scholars who know how to interpret the ambivalent hybrid tāʾ marbūta graph, and thus for a generally later and more sophisticated stage of borrowing. The /-a/ ending is heard in pre-pausal position, e.g., in isolated words, noun-plus-adjective phrases, and words at the end of utterances--less structured environments of the sort in which a foreigner may learn single words by ostensive definition, imitation, rote, or other oral means. (These are of course tendencies, not mutually exclusive processes.) More important is, once more, a sound change affecting the corresponding Persian side of the lexicon. Evidence from book Pahlavi as well as from the subsequent state of New Persian indicates that, by the fifth century C.E., Persian had lost final postvocalic /t/. This dental was voiced, first as /ð/ (the motivation for xezmat as noted above), then as modern Persian /d/, e.g. in /magdāt/ > /mʊbad/ > /mʊbad/. The change included existing Greek and Aramaic loanwords, but did not affect a handful of subsequent loans of miscellaneous provenance--nor the hundreds of Arabic feminine-ending words about to be incorporated in final /at/. This meant, however, that the whole class of /-at/ loans was phonotactically marked as aberrant or alien. Contrary to the loans in /-a/, where native phonology helped to camouflage their difference, the /-at/ inventory was predisposed to be seen initially as a category of Fremdwörter.

Added to this--and, I would argue, in part because of it--is that, semantically, the /at/ inventory (in modern Persian totalling some 640 words) contains a much higher proportion of abstracta and intangibilia (quality nouns, nomina actionis used as abstract

---

6Geiger & Kuhn 1895, I/1:256-57.²
nouns rather than verbs, mass as distinct from count nouns); whereas the referents of the /a/ inventory are more frequently concrete and tangible count nouns, and *nomina actionis* are used as verbs (in combination with auxiliaries) or as count-noun derivatives. Sociolinguistically, the /at/ inventory contains a larger proportion of high-register classical and literary loans, the /a/ inventory more vernacular and everyday words: e.g., *hekdyat* and *qessa* both mean 'story, tale', but the former is reserved for literary anecdotes and parables, while the latter is applicable to any (generally, oral) tale. These and other factors characterize the /a/ inventory as, generally speaking, a lexical category more receptive to conceptually fundamental, syntactically simple, and lexically frequent Arabic feminine-ending loans; and the /at/ inventory as the more hospitable to semantically or structurally complex, learned, and lexically less frequent loans from the same broad source. The basis of the /a/ inventory was most likely established during the earlier, essentially oral and monolingual stage of transmission, and that of the /at/ vocabulary during the efflorescence of Persian literature under the Samanids, Ghaznavids and Seljuks from the late ninth century—the heyday of the literate bilinguals. These writers, the scribes of the imperial chancelleries who were also historians and literary and social critics (e.g., Abu'l-Fazl Bayhaqi and Nezāmi ‘Aruzī of Samarqand), and philosophers such as Ibn Sinā and al-Ghazzālī, who wrote technical treatises in Arabic and works of *haute vulgarisation* in Persian, were prolific importers of Arabic loanwords, coiners of loan translations, and orthographic normalizers of earlier, oral borrowings. Emblematic of this latter propensity is Ibn Sinā's very name: Bu ‘Ali as the sorcerer of Persian folktales, Abu ‘Ali as the polymath of high Islamic culture.

These generalizations must not be pushed too far. Oral and literary processes are not mutually exclusive: New Persian poetry, incorporating some Arabic loanwords, was being recited up to a century before it was first written consistently in Arabic script. Verse is primarily an oral-aural phenomenon, and poets are concerned with the phonetics and phonotactics of their own language, not the morphology or orthography of an alien one—i.e., they are representatives of the "vernacular stage." They too, as well as the scholars, were arbiters of the /at/ ~ /a/ dichotomy: the earliest Persian verses show unstable reflexes of feminine-ending loans, such as *hilat* varying with *hila*. This variation depended upon the metre, the poet's whim, or perhaps on whether the word was intended as the mass noun 'trickery, deception' or the count noun 'trick, ruse.' Significant, too, is that the poetician Shams-e Qays of Rayy explicitly treats the final syllable in /a/ (written with the silent *h*) as phonetically and prosodically the same whether it reflects etymologically a Persian or an Arabic word.

Once this intuitive system of lexical organization was fully established, semantic criteria seem to have taken precedence over those of social register and lexical frequency. Thus the mass noun *govvat* remained the general word for 'strength, power', while the doublet *govva* (a count noun, with plurals *govvahā* and *govā*) soon split off from it in the specialized senses '(physical or mental) faculty, (military) force, (industrial) energy'. Several hundred words originally assimilated in the /at/ inventory shifted to /a/, generally as a result and reflection of their acquiring specialized, count-noun or vernacular senses. Thus *ta'robāt* 'experience' began to drop the /t/ in the eleventh century to become "ta'roba"
'experience; a test, experiment', and ketābat 'writing, penmanship' spawned a doublet ketāba 'inscription, motto, chit'. The system had evolved into an ingenious sub-routine for lexical acquisition and filing, independent of its speakers' relationship with the source language.

IV

The "literary stage" thus saw not merely the massive incorporation of single loanwords but the evolution of systematic lexical strategies and semantic and syntactic calques on Arabic—the true sign of bilingual interference. Scores of new verbs were coined from Arabic verbal nouns, using two strategies already available in Middle Persian: suffixation of -idan (yaratiidan, fahmidan, etc.), and combination with Persian auxiliaries, notably kardan. The earlier, synthetic method, requiring sensitivity to both Arabic and Persian morphology, was arguably the strategy of choice for conservative literate bilinguals. The Persian verb gošidan 'to open', used in the sense 'to conquer' (an opaque metaphor in Persian) from the very earliest works of New Persian literature, is an obvious calque on the metaphorical extension of Arabic fath. Two overt imitations of Arabic usage came to be widely practised from the twelfth century on: in lexical morphology, the use of "broken plurals," which was even applied to phonotactically susceptible Persian words (dahāqin, banāder, akrād), and in noun phrase syntax, the application of pseudo-concord to "feminines" or plurals by addition of a feminine ending (final h) to adjectives of Arabic provenance (manālek-e mahrusa, xānom-e mohtarama). These redundancies are precisely the sort of games indulged in by literate bilinguals.

In the sphere of verb phrase syntax, the very heart of a language like Persian, there is also evidence of a construction calqued on Arabic. Beginning with one of the earliest monuments of Persian prose, Bal'ami's quasi-translation of Tabari's Arabic history (c. 963 C.E.), we find examples of a dependent subjunctive after a modal verb under identity of subject (marā hamī bāyad ke da `vi-ye xodā tī konam), alongside the usual construction with an infinitive or short infinitive (marā bāyad . . . kard[an]). In the following century Naser Khosrow has, e.g., kasi natavānad ke az kasi cizi setānad, and by the end of the thirteenth century this construction had virtually replaced use of the infinitives.10 It is perhaps significant that the subjunctive construction is preferred to translate an Arabic verb governed by a modal, even where the Arabic has a dependent infinitive instead of a subjunctive, in a twelfth-century Persian translation of the Koran: thus natavānestand Ya juj va Ma juj ke bālā-ye divār bar šavad va natavānestand ke ānrā sulāx konand, 'Gog and Magog could not climb over the wall nor could they breach it' translates fāmā istaṭā`ī `an yazharuha wamā istaṭā`ī lahu naqb.11 Similarly, xwāst Fer'awn ke birun konad Musārā `Pharaoh wished to remove Moses' translates ja-`arādā `an yastafizzahum.12 Although the same-subject dependent subjunctive is by no means inherently non-Indo-European (Greek uses it), it is not a feature of Middle Persian or the earliest New Persian; and since Arabic characteristically uses this construction, regardless of identity of antecedent, it seems quite possible that this change was reinforced, if not

11Yahaqqi 1985, Sura 18 v. 97.
12Ibid., Sura 17 v. 103.
actually initiated, by analogy with Arabic in the translations and imitative compositions of bilinguals such as Balʿami. Whatever the case, an important corollary of this shift was that the residual desiderative construction with the short infinitive, of the type *xwâhad kard*, became the periphrastic future tense.\(^1\)

V

So it was that in Persian, which for almost a thousand years has been both the superordinate literary language and a major contact vernacular of a broad region stretching from Anatolia to Bengal, the opposing tendencies of vernacularization (the unconscious assimilation of foreign influences to native norms) and scripturalism (the conscious conservation of high-status foreign or classical forms) seemed constantly to vie for dominance. Some time after the zenith of Persian literature—perhaps as a consequence of the Mongol invasions, with Arabic definitively relegated to the status of a dead language, and Turkish competing with Persian in both the vernacular and the literary markets—vernacularizing tendencies reasserted themselves. They can be sensed in the definitive replacement (at least, in western Persian) of the synthetic method of verb-formation by the analytic, which imposes no phonotactic or morphological limitations. Semantically, vernacular forces are certainly reflected in the lexical assimilation of many of the Arabic broken plurals as Persian singulars, often in a lower sociolinguistic register (*arbâb* 'boss', *'amâla* 'laborer', etc.).

They can also be heard, rather than seen, when Arabic *šajāʾa[t]* 'boldness' is realized as /šojâat/, as if derived by suffixation from its cognate adjectival loan *šojâ* 'bold'; and more insistently in the pronunciation of the type /mofâela/ or /mofâelat/ of Arabic loans on the pattern *mujâla* (an inventory of at least 120), as if they were derived by suffixation from the equally common active participle loans on the pattern *mujâ il* (> P. /mošel/)\(^7\) This may be in principle an autochthonous phonological development (cf. P. /harekat/ < *haraka*[t]); however, it is widespread in all Persian dialects (Tajik *muborîza*, etc.), and in view of the clear case presented by /šojâat/ it is tempting to see in this vowel mutation a trace of "folk morphology," i.e., the strategy of a speech community no longer guided by Persian-Arabic bilinguals that is still striving to assimilate the nonsegmental derivatives of this vast loan vocabulary to familiar, segmental Persian patterns.
REFERENCES


NOTES (2011) on ‘Persian-Arabic Bilingualism…’

A. -- and dictated the modern necessity of a mixed transcription-transliteration system.

B. Not necessarily; like zindig, a Zoroastrian technical term, they might have been taken from literary Middle Persian.

C. Perhaps ‘assimilated’ is better.

D. Add to note 6 some examples from Pisowicz?

E. Add to 7: Unless vernacularized in a dialect, such as Tajik, which evolved a written reflex.

F. Perhaps examination of some Semitic cognates (Heb *ketuba*, might be instructive.

G. Other calques: bāb < dar, ?

H. This argument is elaborated in ‘Grammaticalization in Process:…xât’ (20??).

J. Any more examples of this “folk-morphological tendency to assimilate Arabicate forms to IE or Altaic continuous lex. morphology?