IRAN
QUESTIONS ET CONNAISSANCES

ACTES DU IV° CONGRÈS EUROPÉEN DES ÉTUDES IRANIENNES
ORGANISÉ PAR LA SOCIÉTAS IRANOLÓGICA EUROPÉA
PARIS, 6-10 SEPTEMBRE 1999

VOL. III
CULTURES ET SOCIÉTÉS CONTEMPORAINES

TEXTES RÉUNIS PAR BERNARD HOURCADE

PEETERS
ASSOCIATION POUR L'AVANCEMENT DES ÉTUDES IRANIENNES
PARIS 2003

TABLE DES MATIÈRES

Avant-propos ................................................................. 5
Table des matières .......................................................... 7 - 8

Langues et littératures contemporaines
John PERRY : Persian as a HomогLOSSIC Language .................. 11-28
Alain CHRISTOL : Les deux pluriels de l'ossète ....................... 29-44
Thamar E. GINDIN : A Unique Plural form in Judeo-Yazdi ........... 45-56
Vladimir IVANOV : Traces of Animacy in Persian ..................... 57-63
Agnes KORN : The Outcome of Proto-Iranian *f in Balochi .......... 65-75
Halkawt HAKIM : De l'influence de la poésie persane sur la poésie kurde .... 77-85
Sabir BADALKHAN : Tawkali Mast, Lover, Poet and Saint
of Eastern Baloch Tribes .................................................. 87-106
Ameneh YOUSSEFZADEH : Le dastan amoureux : l'épopée lyrique,
source des romances populaires ...................................... 107-118
Isabel STÜMPEL : Aspects de l'intertextualité dans la littérature
persane contemporaine ................................................... 119-128
Rivanne Sandler : Words of Attachment : Literary Antecedents
of Expatriate Iranian Women ........................................... 129-139
Anna Krasnowolska : Rediscovery of Nature in Iranian
Post-Revolutionary Literature .......................................... 141-150

Cultures, villes et territoires
Jean-Pierre DIGARD : Pour une ethnologie du « superflu » en Iran .................. 153-160
Philippe ROCHARD : Les représentations du « beau geste » dans le sport
traditionnel iranien ....................................................... 161-170
Victoria ARAKELOVA : The Styäh-Gâlesh and Deity Patrons
of Cattle-Breeding Iranian Peoples .................................... 171-176
Chirine MOHSENI : Les réfugiés kurdes iraniens en France.Modes de vie
et intégration ............................................................ 177-183
Seyyed-Mohsen HABIBI : L'école d'Ispahan en urbanisme. Une approche
contemporaine du passé ................................................ 185-195
Masserrat AMIR-EBRAHIMI : Vers une nouvelle géographie
culturelle de Téhéran? .................................................. 197-209
Djamal ATHARI : Banlieue de Téhéran et développement durable .... 211-216
Hasan-Ali GHAYOOR : Geographical Distribution of Qanats in Iran .......... 217-223
PEETERS
PEETERS - BONDGENOTENLAAN 153 - B-3000 LEUVEN

Histoire, sociétés et politiques
Nugzar TER-OGANOV : The Persian Cossack Brigade - An Outpost of Russian Tsarism in Iran: 1879-1921 ........................................... 227-239
Olga JIGALINA : The Constitutional Revolution in Kurdistan Northern Khorasan (1905-1911) .................................................. 241-244
Hadi SULTAN-QURRAIE : Problems of Identity of Azerbaijani Authors of Post-Constitutional Iran ........................................... 247-249
Jane TREMBLAY : Mythes, histoire et impressions : Mohammad-Reza Chah Pahlavi décrit par ses biographes ........................................... 251-269
Morteza MONADI : La crise d'identité des jeunes Iraniens, futurs acteurs politiques .......................................................... 271-280
Misagh PARSA : Iranian Students and the Revolutionary Struggles .......................................................... 281-292
Farah GILANSHAH : Experience of Iranian Immigrant women in the USA The case of the Twin cities ........................................... 297-306
Mohammad-Reza DJALILI : La République islamique d'Iran et le monde (1979-1999) : Analyse critique de vingt ans de politique étrangère ........................................... 307-318

John R. PERRY
UNIVERSITY OF CHICAGO

PERSIAN AS A HOMOGLOSSIC LANGUAGE\footnote{1}

RÉSUMÉ
Depuis qu'il a été présenté en 1959 par Charles Ferguson, le concept de diglossia, a été largement appliqué aux langues dont les locuteurs, (contrairement au grec et arabien), ne reconnaissent pas de dichotomie fondamentale dans leurs niveaux de langue. Grâce à une comparaison statistique des contrastes phonologiques, morphologiques, syntaxiques et lexicaux entre, d'une part, l'arabe classique/l'arabe dialectal égyptien, et d'autre part, le persan standard et celui de Téhéran, cet article montre que le persan d'Iran est 28% moins diglossique que l'arabe classique/égyptien. Il est en outre démontré, à partir de données socio-linguistiques, qu'en fait, le persan n'est pas du tout diglossique.

Mots clés : Diglossia, arabe, persan, registres de langage, quantification des données.

ABSTRACT
Since Charles Ferguson introduced the concept of diglossia in 1959, it has been widely applied to languages (unlike, e.g., Greek and Arabic) whose speakers do not recognize such a fundamental dichotomy in their stylistic registers. This paper will show, by statistical comparison of phonological, morphological, syntactic and lexical contrasts in Classical Arabic and Egyptian colloquial vs. Standard Persian and Tehran colloquial, that Persian of Iran is less than 28% as diglossic as Classical/ Egyptian Arabic. It will also be argued on sociolinguistic principles that Persian is not in effect diglossic at all.

Keywords: Diglossia, Arabic, Persian, colloquial registers, quantification of features.

\footnote{1 I am grateful to John Eisele for his valuable comments on an earlier version of this paper. Any use I may have made of such comments is, of course, my own responsibility.}
1. The range of registers and scope of interaction between Persian as the repository of the Classical literary canon, as a modern international literary standard, metropolitan formal and informal standards, and several regional varieties, is dauntingly broad and intricate. It invites sweeping and contradictory generalizations of the sort “Modern Persian is essentially the same as Classical Persian” or “Tehran colloquial is entirely different from literary Persian” or “Farsi and Tajik are the same language/separate languages.” Moreover, the popular replacement of “Persian” by “Farsi” in recent Anglo-American (and, increasingly, European) usage implies a distinct contemporary vernacular and written standard divorced from any linguistic and cultural antecedents. This centrifugal trend is unfortunately reflected in some modern pedagogy and scholarship, which postulates separate and radically different standards of written and spoken, or formal and informal, Persian. In teaching Persian, the Arabic writing system is too often presented to students as an advanced skill to be avoided until absolutely necessary. There is a European-American tradition of quite elaborate primers of colloquial Persian entirely in Roman transcription, to a degree that has never been tolerated for Greek or Russian. In sociolinguistic studies of Persian, emphasis on subsystems such as the range of phonological and morphophonemic divergences from the written word, or the special pronominal and verbal forms used in polite interaction (za’araf), have highlighted differences at the expense of similarities. Dr. Eva Jeremiás, in an important and provocative article under the title “Diglossia in Persian”, would apparently reinforce this trend, claiming that contemporary Persian offers “a striking example of diglossia.” (Jeremiás, 286).

Diglossia, as defined by Charles Ferguson in his classic treatment of the phenomenon, is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.

Thus Classical literary Arabic, popularly imagined to have been modeled on the language of the Koran and of the pre-Islamic Arabian poets, has been invested with an aura of sanctity that qualifies it as a model for the modern literary standard, to the exclusion of any of the regional vernaculars. The situation has been compared to that of Latin vis-à-vis the nascent Romance languages during the early medieval period in Europe — Ibid, 337). Modern diglossic situations postulated by Ferguson are that of Greek (Katharevousa or “puristic” vs. Dhimotiki, demotic, from approximately the 1820s until the 1970s), High German vs. Swiss German, and French vs. Haitian Creole. One might add to these the language situation in Ottoman Turkey, setting a highly Persianised Turkish of divan literature against everyday spoken Turkish.

Accepting Ferguson’s definition and examples, Jeremiás objects perceptively that “it follows — and this fact remains unassailed by Ferguson — that it is actually the degrees of these structural differences and not their mere existence that results in a diglossia situation.” (Jeremiás, 272 original emphasis). This failure to specify how far apart the two sociolects must be for this to constitute diglossia was admitted by Ferguson in his thirty-five-year retrospective (Ferguson 1993, 223-2); and indeed it appears that no attempt had been made in the meantime (nor subsequently, to my knowledge) to quantify degrees of sociolectal polarity, either for a single language or comparatively. Jeremiás undertakes in what follows to “examine whether the differences between the formal and informal varieties of a language [sc. Persian] are large enough to warrant the term diglossia. Can these differences be measured and if so, how?” After a stimulating excursus into the history of New Persian, however, she completely neglects to try to measure the pertinent synchronic differences, contenting herself with five disparate features and concluding that “there are fundamental differences between formal and informal Persian,” which are “as marked as those between two independent languages” (Jeremiás, 286, 287).

In order to reach such a radical conclusion, it is surely a prerequisite to establish a scale of measurement and apply it comparatively to the target language and at least one other diglossic syndrome. Let us therefore continue the task where Jeremiás (and Ferguson) left off, and try to measure the degrees of difference in phonology, morphology, syntax and lexis between the stylistic registers usually chosen for informal or spoken interaction, and for formal or written communication, in Persian as against Arabic, one of the agreed exemplars of diglossia.

2. Diglossia, under Ferguson’s terms, admits of a broad definition and a narrow one. According to the broad definition, there are systematic stylistic differences between “high” (H) and “low” (L) language varieties. This is true of most, if not all, natural languages, including English and Persian; indeed it is hard to imagine a language for which this statement in its broadest sense would not be valid. Even languages that have undergone radical reforms explicitly aimed at eradicating diglossia, such as Turkish
cognate, and of (b) equivalent morphosyntactic processes that are not derivatives or transformations of one of the other.

3. In order to assess the comparative degree of “broad” diglossia at work in Arabic and Persian, and if possible to determine whether one or both are “narrowly” diglossic, I suggest the following measurable criteria as summarised in the two tables — one for modern Persian of Iran (Formal Standard and Informal Standard) and one for Arabic (Modern Standard, or MSA, and Egyptian Colloquial); each variety dubbed respectively "H" and "L" for short. Spoken samples of the H[igh] style would be, typically, a prepared public speech or a newsread or serious essay read aloud; written samples would be almost any text not intended to reproduce dialogue or humorous monologue. Spoken samples of the L[ow] style would typically be casual, non-technical dialogue; written samples would be cartoon captions, speech bubbles and representation of dialogue in fiction. The identification of features in Arabic is drawn principally from three sources: an article by an Egyptian linguist contrasting the grammars of Classical and Egyptian Arabic (Selim 1967); a 900-word contrastive list included in a study of the same diglossic registers (Schmidt 1974); and a grammar of Egyptian Colloquial (Mitchell 1962). The equivalent Persian points are drawn from an unpublished teaching handout which I synthesized from several published studies of the relation between written or formal and spoken or informal Persian (Boyle 1952, Hodge 1957, Deyhim 1368/1989, Ghobadi 1996); and, for the lexicon, from an unpublished 1500-word Persian vocabulary taken from word-frequency lists and textbook vocabularies that I compiled for student use. In each case these materials were supplemented by my own observations or analyses.

The features selected are strictly those identified as diagnostic for diglossia. They are categorized under four sections: Phonology and Morphophonology; Morphology; Syntax and Semantics; and three degrees of Lexis. The system of scoring (which is bound to be somewhat idiosyncratic and controversial) works as follows.

One point is awarded to across-the-board phonological features distinguishing H and L, and one to cognate lexical units, i.e., words which show a considerable difference of articulation as between H and L — beyond the standard phonological correlations — yet are transparently cognate.

Two points are awarded for (a) morphophonological alternations that characterise H and L: such features as the assimilation of the Egyptian definite article to non-šamī consonants, or the intrusive consonants before pronominal enclitics in Persian, appear to be more distinctively diglossic than individual phonemic correlations; and (b) non-cognate lexical units, i.e., words exclusively (or almost exclusively) distributed between H and L that are etymologically different, as Arabic ḥidā' ~ gazma ‘shoe’, Persian sar ~ kalle ‘head’.

Three points go to (a) morphological differences: this is mainly a matter of recording grammatical features of H that are missing in L, and (b) lexical items that are grammatical function-words, or parts of a system or series (numerals, interrogative particles) or otherwise arguably more frequent in everyday conversation (basic verbs and adverbs).

Four points each are awarded to syntactic structures differing in kind as between H and L, or absent in one register (such as the loss of adjective agreement in some noun phrases of Arabic L, and the addition of the progressive construction with dāštān to Persian L); and to distinctive semantic features, such as the productive passive senses acquired by some derived stems of verbs in Arabic L.

Wherever possible, I have attempted to generalise in order to reduce the number of separate features: thus in Arabic, reflexes of the glottal stop as a glide and as zero, and of the diphthong aw as ə and α, are counted as one feature, as are the alternations ən ~ in and əm ~ um in Persian. In a handful of cases in each language where there is a phonological differentiation other than one of the regular sound correlations already counted, but it seems to be minimally different in morphological or lexical terms, I have ignored it: thus in Arabic, the pronouns huwa ~ hawaya, hiya ~ hiya, ənta ~ ənta, ənti ~ ənti, əntum ~ əntu and hum ~ humma (whereas naHnu ~ iHnu appears distinctive enough to be counted under Morphology); in Persian, the adversative interjections əxar ~ ãxe and magar ~ mage, agar ~ age ‘if’ and ãtowr ~ ãtow ‘how?’ (where asseverative dige (< digar) additionally involves a lexical innovation and, like asseverative ke, is counted under Functional/Frequent lexical features).

Let us now flesh out the specific points noted in the tables, section by section for each language, the better to compare their diglossic indices. Examples in both H and L will be transcribed phonemically as if from written sources; pausal form is generally cited for Arabic, whether H or L. Uncontrasted examples in the tables are reflexes of L.

PHONOLOGY AND MORPHOPHONOLOGY.

In this section, Persian slightly outscores Arabic in phonology, but has marginally fewer distinctive morphophonological alternations.

Table 1. Arabic.

The eight automatic consonant alternations between H and L include each of the two vernacular reflexes of literary ə (as ənwa ‘revolution’, talāta ‘three’) and literary ə (as ən ‘this’ and ɪzə ‘if’), and the salient na-
tional shibboleth, *jīm* as g. Strictly speaking, this latter is a dialectal feature; however, it is subjectively accepted as a distinguishing feature within registers of spoken Egyptian Arabic (Kaye, 51; Selim, 135). The alternations between diphthongs and vowels are seen in *lawn* ~ *lān* 'color', *mawlid* ~ *mālid* 'holiday', *bāyti* ~ *bāti* 'house'. Under morphophonology, the progressive assimilation of *l* of the definite article to an initial velar in colloquial Egyptian (*ik-kābri* 'the bridge', *ig-gawāb* 'the letter') is not obligatory, but is a feature absent from MSA (Selim, 13). The four alternations affecting verbal prefixes are exemplified in *yaktabu* ~ *yiktab* 'he writes', *taktabu* ~ *tiktab* 'she writes', *tafaDDal* ~ *itaDDal* 'please', *mutawassiṭ* ~ *mitwassiṭ* 'average'.

Table 2. Persian.

The phonological correlations between Standard and Colloquial Persian (omitting the two exemplified in the table) may be illustrated thus: *ān* ~ *an* 'that', *āmad* ~ *umad* 'he came'; *mardā* ~ *mardā* 'men', *bedeh* ~ *bede* 'give'; *ferēstå* ~ *ferēstå* 'he sent', *mesl-e* ~ *mess-e* 'like'; *dodzide* ~ *dozdide* 'stolen'; *goff* ~ *goff* (goof) 'he said'; *qadar* ~ *quad* 'quantity', *fekr* ~ *fe:k* 'thought'; *vaqii(le)* ~ *vaqiti* 'when', *naqī* ~ *naxī* 'design'; *saxt* ~ *sax* 'tough'; *very', *boland* ~ *bolan* 'high'. In morphophonology, the reflexes of the postposition *-rā* and the intrusive *h* and *θ* before pronominal enclitics are significant (the use in L of these enclitics after prepositions and other non-verbal forms is conducted separately under Syntaxis). The colloquial reflexes of third person singular and second plural verb endings are counted, but not that of the third plural (*and* ~ *-an*), since this is included above as a general phonological rule. The contrast *nami*- ~ *nemi*- in verbal prefixes, like the Egyptian Arabic *j* ~ *g*, is primarily a dialectal feature of modern Standard Persian, independent of stylistic register (raised *a* > *e* in many environments is now typical of Tehran enunciation); but again, this feature has been co-opted as distinctive between H and L registers of the Standard to the extent that it is often preferred in declaiming Classical literature and solemn utterances. It therefore seems legitimate to include it as a sociostylistic variable. (We may discount both the Arabic and the Persian features without affecting the comparative score).

### Table 1. Diglossic Features in Arabic

<table>
<thead>
<tr>
<th>Features</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHONOLOGY: 1 pt. each</td>
<td></td>
</tr>
<tr>
<td><em>q-</em>; <em>f-</em>; <em>g-</em>; <em>θ-</em>; <em>θ-</em>; <em>θ-</em>; <em>θ-</em>; <em>D-</em>; <em>-w</em> glide, <em>aw</em>- <em>av</em>; <em>ay</em>- <em>ey</em></td>
<td>10 10</td>
</tr>
<tr>
<td>MORPHOPHONOMY: 2 pts. each</td>
<td></td>
</tr>
<tr>
<td>Nom.: <em>il-k-k</em>- <em>il-g</em>-gram. : <em>hu-</em> <em>ū</em>-</td>
<td>4</td>
</tr>
<tr>
<td>Verb: Prefix types *ya-*pi-, *ta-*ti-, *ta-*ti- <em>mu</em>-mi-; assim. yiggawīb</td>
<td>5</td>
</tr>
<tr>
<td>L stem types <em>tfim</em>, <em>nis</em>- <em>kātib</em></td>
<td>3 24</td>
</tr>
<tr>
<td>MORPHOLOGY: 3 pts.</td>
<td></td>
</tr>
<tr>
<td>Noun: in H, 41 inflected &amp; pausal form types; in L, 9 (32 fewer)</td>
<td></td>
</tr>
<tr>
<td>(types <em>xadd</em>- <em>a</em>- <em>-it</em>- <em>in</em>- <em>-i</em>- <em>-āt</em>- <em>kiθābn</em>- <em>-ē</em></td>
<td>32</td>
</tr>
<tr>
<td>Pers. pn.: in L, 7 fewer forms; 1 differs (*ifna)</td>
<td>9</td>
</tr>
<tr>
<td>Dem. pn.: in H, 14 forms; in L, 6 (da, di, dīl, dīkha &amp;c.)</td>
<td>8</td>
</tr>
<tr>
<td>Rel. pn.: in H, 8 forms; in L, one (<em>iθt</em>)</td>
<td>7</td>
</tr>
<tr>
<td>Verb: Perfect. forms, suffix: <em>H</em> 13, <em>L</em> 7 (6 fewer)</td>
<td></td>
</tr>
<tr>
<td>Imperf. forms, suffix: H 23, L 8 (15 fewer)</td>
<td>15 231</td>
</tr>
<tr>
<td>SYNTAX, SEMANTICS: 4 pts.</td>
<td></td>
</tr>
<tr>
<td>Adj. agreement: in H, 3 discrepancies (<em>έp</em>, <em>m</em>- &amp; <em>f</em> dual)</td>
<td>3</td>
</tr>
<tr>
<td>Verb: L bi-pres., Ha-fut., <em>kiθ</em> tense; themes V, VI, VII as passives</td>
<td>6</td>
</tr>
<tr>
<td>In L, <em>b</em> <em>θ</em> neg., <em>θθ</em>- <em>θ</em>- neg.; <em>θiθ</em> construction</td>
<td>3</td>
</tr>
<tr>
<td>VS-SV word order; L postposed demonstr. (<em>il</em>-Hāga <em>di</em>)</td>
<td>2 56</td>
</tr>
<tr>
<td>COGNATE LEXICON: 1 pt.</td>
<td></td>
</tr>
<tr>
<td>23 pairs, inc. <em>awθ</em>- <em>awθ</em>- <em>awθ</em>- <em>awθ</em>-; <em>θlθ</em>- <em>θlθ</em>-;</td>
<td>23 23</td>
</tr>
<tr>
<td>NON-COGNATE LEXICON: 2 pts.</td>
<td></td>
</tr>
<tr>
<td>76 pairs, inc. <em>fam</em>- <em>θ</em>- <em>θ</em>- <em>θ</em>-; <em>θ</em>- <em>θ</em>-</td>
<td>76 152</td>
</tr>
<tr>
<td>FUNCTIONAL/FREQUENT LEXICON: 3 pts.</td>
<td></td>
</tr>
<tr>
<td>44 pairs, i.e. teens 11-19; 7 interrogs.; 7 basic verbs; 4 time advbs.; <em>fiddan</em>- <em>awθ</em>- <em>awθ</em>- <em>awθ</em>-; <em>θlθ</em>- <em>θlθ</em>-; <em>θlθ</em>- <em>θlθ</em>-</td>
<td>44 132</td>
</tr>
<tr>
<td>Total features/points</td>
<td>256 628</td>
</tr>
</tbody>
</table>
TABLE 2. DIGLOSSIC FEATURES IN PERSIAN

<table>
<thead>
<tr>
<th>Feature/Point</th>
<th>Features Points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PHONOLOGY</strong></td>
<td>1 pt.</td>
</tr>
<tr>
<td>an-un, h-ei, Vh/V~V: (ba: id, o: de), γ-γ (mityād), xC-xz, 2C-2z, fi-ff, Cr-Ce, qC-xC. C/d-Ca, nul-n</td>
<td>11 11</td>
</tr>
<tr>
<td><strong>MORPHOPHONOLOGY</strong></td>
<td>2 pts.</td>
</tr>
<tr>
<td>Noun: -īf~ro, -o; in L: be-h-ei, bā-hā types; borda-t-e types, omission of eṣfī particle</td>
<td>6</td>
</tr>
<tr>
<td>Verb: nāmī~nēmī; kardē-kardē; ast-e; 3sg. -od-e; 2pl. -id-in</td>
<td>5 22</td>
</tr>
<tr>
<td><strong>MORPHOLOGY</strong></td>
<td>3 pts.</td>
</tr>
<tr>
<td>Lack of periphrastic future; mā-hā, šomā-hā</td>
<td>2 6</td>
</tr>
<tr>
<td><strong>SYNTAX, SEMANTICS</strong></td>
<td>4 pts.</td>
</tr>
<tr>
<td>In L: dīštēn progressive pres., past; def. -e; postposed advb./NP; rafat-e type; be-e type; če-e type; omission of prep.</td>
<td>8 32</td>
</tr>
<tr>
<td><strong>COGNATE LEXICON</strong></td>
<td>1 pt.</td>
</tr>
<tr>
<td>10 pairs: nesh=šin, šuy=šur;</td>
<td>10 10</td>
</tr>
<tr>
<td>L: deqeq, gole, gošne, hanu, martike, golp, sāhāb, sula</td>
<td></td>
</tr>
<tr>
<td><strong>NON-COGNATE LEXICON</strong></td>
<td>2 pts.</td>
</tr>
<tr>
<td><strong>FUNCTIONAL/FREQUENT LEXICON</strong></td>
<td>3 pts.</td>
</tr>
<tr>
<td>23 pairs: raff-r and 7 other verb stems; bāyest=bāst/bāhās; az-a, ham-am, šes-sē, yek-ye, bar-ye-vās(ey), bāz-vāz, bar-var, sepas-unvax; namudan-kordan</td>
<td>23 69</td>
</tr>
<tr>
<td>L: dīqe, ke, hākī, čīn, hamči</td>
<td></td>
</tr>
<tr>
<td>Total features/points</td>
<td>77 174</td>
</tr>
</tbody>
</table>

MORPHOLOGY.

At 77 features against 2 (231 points to 6), this section is by far the most indicative of Arabic's greater degree of diglossia, at least in the mechanical sense of a reduction of grammatical forms. It might be objected that, since standard Persian already enjoys a much reduced inflectional morphology (an old legacy, from Middle Persian times), fully-inflected literary Arabic is unfairly handicapped. The fact remains that comparative grammatical complexity vs. simplicity is a salient feature of the distinction between H and L and must be considered as an index of diglossia.

Table 1, Arabic.

Colloquial Egyptian is here characterised by a considerable reduction of inflectional forms as against Classical or literary Arabic. The loss of case endings (i'rab) and collapse of a variety of nominal inflections is summarised in the nine remaining nominal forms of the prototypical nouns xaddām 'servant' and kitāb 'book', and the third-weak adjective Gāli 'expensive'. Pronoun reductions are likewise summarised in the list of remaining L forms, and one alternation judged more than merely phonological is also counted here. Loss of distinctive personal endings in the feminine plural and dual of verb paradigms, and the collapse of indicative, subjunctive and jussive into one form (Ibid., 140-4) complete the count.

Table 2, Persian.

There is only one comparable reduction, that of the periphrastic future (type xalām raft) in L, and one pronoun variant, the addition in L of the plural morpheme -hā to mā 'we' and šomā 'you' in order to distinguish true plural from polite singular. These features are not absolute, merely typical of informal Persian.

SYNTAX AND SEMANTICS.

At 56 points to 32, this comparison leaves Arabic considerably more diglossic than Persian. This section is admittedly the least homogeneous; features so disparate (within both languages) may not merit a similar score, especially in the absence of testing for frequency.

Table 1, Arabic.

The reduction in the range of adjectival agreement (cfem. to masc. ending) is illustrated in the alternations mudarrisatīn mašriyātīn ~ mudarrisāt mašriyān 'Egyptian teachers' (cfem. pl.), maktabān mukhimān ~ maktubān mukhimīn 'two important offices' (masc. dual), maktabātān mukhimmatān ~ maktubātān mukhimīn 'two important libraries' (cfem. dual). This is balanced by innovation in the verb system:
the analytical present and future tenses, as (hūwwa) bi-yiktib ‘he is writing’, (biyya) Ha-tīgī ‘she will come’; expanded use of the active participle, as ‘anā kātid-ī ‘I’ve written it’; and productive generation of derived stems as passives, yitbā’ yīnba ‘it is sold’. Verbal and nominal negation in L involves the enclitic -š in two distinct constructions: ma gat-š ‘she didn’t come’, miš fāhīm ‘doesn’t understand’. The possessive pronominal suffix, as in il-bāb bītā ‘il-bēt ‘the door of the house’; two contrasts in word order complete the tally of 14 points.

Table 2. Persian.

Seven of these points, unsurprisingly, constitute an expansion of Jeremiā’s five distinguishing features of Persian “diglossia”: the innovation of progressive constructions of the type dārām kār mi-konam, dāstām kār mi-kardam ‘I am / was working’ (since this two-clause structure has not been grammaticalised into a “tense”, it is counted under Syntax rather than Morphology); the stressed suffix marking a defined noun, as pesar-ē jīm šod ‘the boy made himself scarce’; expanded and versatile use of pronominal enclitics as compared with the formal language; omission of the preposition, e.g., Tehran rāft ‘he went to Tehran’; and violations of the Standard verb-final word order in L utterances of the type rāft bāzār ‘he went to market’ and na-didam bāzār-t-o ‘I didn’t see your dad’ (Jeremiā, 281-8). I have counted as three separate features the colloquial use of the enclitic as a subject marker (rāf-ē ‘he went’), in combination with a preposition (bē-ūt-ī) goft ‘he said to him’, and as attached to a sentence constituent other than these or a verb (ē-s-e ‘what’s up with you?’, be-Irān-at borde ‘took you to Iran’). It should be noted, however, that these usages are all to be found in earlier styles of Persian, notably Classical poetry, as well as in modern informal colloquial. I have added in this section the aversive use of the particle ke as in man ke nemīdānām ‘I don’t know’ (i.e., ‘don’t ask me!’): the similar use of digē (< digar) is included under the rubric of Frequent and Functional Lexicon below.

LEXICON.

Contrastive pairs in the lexical sections of our Arabic diagnosis total 143, and the weighted score is 307 points; in Persian, the totals are 44 contrastive features and 100 points.

Table 1. Arabic.

The category of Cognate Lexical variants as between H and L is relatively straightforward. The 23 pairs identified include abn ~ wīn ‘ear’, waqū ‘waqū ‘face’, axāda ~ xad ‘to take’, āxīd ~ wāxīd ‘taking’. Typical Non-Cognate pairs are anf ~ mnāxīr ‘nose’, dajāda ~ farxà ‘chicken’, Hodiqa ~ gīnēna ‘garden’. The selection of Non-Cognate pairs is more problematic. In Schmidt’s list of 900 words from his samples, he claims a maximum of 560 diglossically distinct pairs (i.e., 62.2% of his list, including examples of regular phonological correlation, not counted in my Lexical section); of these he identifies a maximal non-cognate vocabulary (i.e., both different roots and different morphology) of some 212 pairs, and a more strictly-defined set (non-cognate roots) of 179 pairs (Schmidt, 55-5). My own restriction of this to non-cognate pairs that seem to be distributed exclusively yields only 114, i.e., 70 so designated and 18 that are included in my Functional /Frequent Lexicon. I have further discounted as non-cognate vocabulary many words that appeared not to be exclusive to either H or L, e.g. faqīr ~ maskin, fa ‘ir ‘poor’, since the L forms are minimally differentiated from H reflexes maskin and faqīr. My list is therefore quite conservative, tending to undervalue contrasts.

In the last sub-section, Functional and Frequent Lexical items, “teens” refers to the alternation al-Ḥadā al-‘asāra ~ Ḥidāāṣ(ar) ‘eleven’, etc. The “hundreds” series (not counted) shows similar alternations, but not to a much greater extent than accords with general phonological rules. The basic verbs referred to are fa’ala ~ ‘amāl ‘to do’, ra’ā ~ šāf ‘to see’, ḍahaba ~ rāḥ ‘to go’, atā bi- ~ gāb ‘to bring’, ẓāhara bi- ~ sāl ‘to take, remove’, arāda ~ ẓawār, etc., ‘to want’, wajāda ~ lā ‘to find’; the time-adverbs are al-‘ān ~ dilwa ‘now’, ḍāmā, ‘alā ‘l-ti’r ‘be-dān then, next’, amān ~ imbārā ‘yesterday’, Gād(ān) ~ būlār ‘tomorrow’; interrogatives are má ‘mādā ‘is / are what’s’, kāfya ~ izz ‘how’, bimā ~ addā ‘how much?’, man ‘mīn ‘who?’, ‘ayna ~ fīn ‘where?’, matū ‘timā ‘when?’, limā(bā) ~ lēh ‘why?’. Other words, as listed in Table 1, include ma’a ~ wayya ‘with’, dāxīl ~ gūwwa ‘inside’, līka ~ ‘asān ‘in order that’, ḍāmāta ~ āf ‘there is /are’ and its negative counterpart layṣa ~ mā fīs.

Table 2. Persian.

The greatest contrast between Arabic and Persian lexical “diglossia” is seen in the everyday non-cognate vocabulary, where Arabic shows a six- or sevenfold superiority; in the other two sub-categories, Arabic has just over twice as many diglossic features. There are relatively few non-cognate word-pairs in Persian distributed strictly between H and L (āhdeste ~ yāvās ‘slow(fly), gentle, gently’ is the clearest example). As in Arabic, I have usually discounted pairs for which, while each item denotes either the colloquial or the formal, neither is categorically excluded from one stylistic register or the other: such is Persian farzand, kudak ~ bāche ‘child’, since bāche is also admissible in H. On the other hand, I have counted the systemic use of namudan as an elegant variant of the auxiliary kudan, a
systematic marker of written or formal Persian, even though this does not exclude *kardan* from the lexicon of H.

Among the Functional-Frequency listings are common verbs showing stem alterations in colloquial, as *miš-am* 'I become', *mi-g-am* 'I say', *mi-d-am* 'I give', *mi-x-in* 'you want', *mi-tun-am* 'I can', *be-zár* 'let' and *na-zāšt* 'he didn't let'.

4. On the basis of this mechanical quantification, Arabic would appear to be about three and a half times as diglossic as Persian. Of course, variables other than the features specific to Arabic of Egypt could also affect this generalization. Features of Arabic vernaculars other than Cairoene vary widely, some contributing to, and some mitigating, a contrast with Standard Arabic; however, it is my impression that Syrian or Iraqi spoken Arabic do not differ significantly in the degree of their overall lexical and grammatical diglossia with MSA from that of Egyptian colloquial.

As noted, in phonological alternations the language situations are evenly matched, and in syntactic structures Persian is less than 40% as "diglossic" as Arabic. The overwhelming number of diglossic morphological features in Arabic and the negligible number in Persian is attributable chiefly to the reduction in Arabic L of grammatical complexity in H. As Jeremiás notes, innovation by L plays a distinctive role in the "diglossia" of Persian (see under Syntax, Table 2); in the corresponding section of Table 1, however, it can be seen that Egyptian colloquial has also been innovative, in its generation of three new tenses and an alternative to the status construeus. The most salient contrast, however, is in the contrastive lexicon of the H and L registers of the two languages, and especially in the non-cognate alternants: counting frequent /functional lexemes, a total of 114 pairs in Arabic to 13 pairs in Persian.

Quantification alone is not enough, especially if it does not include a rigorous frequency test. I return to Dr. Jeremiás' arguments in principal for diglossia in Persian, some of which appear rather to support a theory of homoglossia.

First, she dismisses the existing studies of systematic variation between literary and colloquial Persian as "simple transformational rules to indicate how to 'classicise' or 'colloquialise' the style." (Jeremiás, 276). But of course they are; and the very case with which many of these variations can be expressed as "transformational rules" — an operation much more difficult, if not impossible in the case of Arabic diglossic variants! — shows that we are dealing with closely related stylistic variants of the same language.

Again, "[e]vident as they are for a linguist, the differences between the two languages, Classical and Modern Persian, are far less conspicuous for a layman." (Ibid., 278). Precisely: cognate vocabulary and modified structures as between H and L (however defined) in Persian make it easy for the native and naive foreign user to intuit the same "transformational rules" elaborated by the scholar, whereas the characteristic non-cognates and generically different rules that distinguish Classical Arabic from a vernacular are intuited as belonging to a different grammar even by native speakers.

Dr. Jeremiás' determination to foist diglossia on Persian leads to some exaggerated criticism of pedagogic approaches. The progressive aspect of the verb in present and past tenses, using *dāšt* as an auxiliary, is a feature of the Informal Standard in Persian. According to Jeremiás, "present-day grammars do not usually include [these forms] in the paradigms of verbs, but only mention them as a colloquialism," in support of which she cites a 1963 coursebook by S. Obolensky (Ibid., 281). In fact most well-known modern grammars (Lambton 1953, Lazar 1957, Boyle 1966, Windfuhr and Tehrani 1981, Thackston 1993) include adequate descriptions of the forms, even as they categorise them as colloquial in usage (which Jeremiás does not dispute); it is hardly necessary to give a full twenty-four-part paradigm for a construction comprising standard verbs that are conjugated elsewhere. Modern grammarians also note "classicisms," such as the dative use of the postposition *-rād*, in similar offhand ways, with warnings that they are no longer a regular part of the grammar of Standard Persian. Could they be right in thinking they are describing a single synchronic language with stylistic variations adhering from past and nascent usage, rather than two or three different languages with the same name?

The progressive construction is recent; the fact that it is not yet established in literature, and indeed not yet even fully grammaticalised (it comprises two independent clauses, and is not therefore morphologically established as a tense) shows that it may well be in a process of evolution. Rather than a fixed marker of diglossia (in Jeremiás' words, one of the "fundamental differences between formal and informal Persian, which are deeply rooted in the history of the language" (Ibid., 285), it is a bridge between formal and informal, frequently met in journalistic writing if not yet established in "serious" literature. It evidently meets a need and will probably soon insinuate itself into the literary language (as similar constructions did in Spanish, English and Tajik). Jeremiás' claim could more justifiably be made in respect of truly diglossic constructions in Arabic, such as *lā 'adīr* vs. *muš /muš 'ārij* 'I don't know', where not a single component of the colloquial idioms (except the phonology) behaves in a way remotely analogous to the formal standard.
Since Ferguson’s popularisation of the term in 1959 “diglossia” has become something of a fad, and has spawned a considerable industry: a recent bibliography of the topic lists at least 2,500 books, articles and dissertations (Fernández 1993). It is therefore hardly surprising that the term has sometimes been misapplied. Sociolinguists have expressed serious doubts whether it is still a useful concept; e.g., “le terme ‘diglossie’ a été mêlé à toutes les sauces – à tel point qu’on peut bien se demander quelle est son utilité distincte” (Mackey, 11). But it is precisely the asking of such questions that advances linguistic studies; Dr. Jeremiáš would no longer repeat her claim in its original terms, any more than Charles Ferguson would (Ferguson 1993, 214ff). Despite the problems involved (in particular the “fuzzy interfaces” between diglossia, bilingualism, and sociolectal or dialectal variation) I believe (currently, at least) that diglossia can be a measurable variable. My intention in reopening this case was to attempt what seems so far not to have been attempted, a comparative quantification of diagnostic features for this elusive phenomenon.

However crude and debatable the methods, I think the evidence here is compelling that Persian is not nearly as dichotomous as Arabic in terms of a broad definition of diglossia. If, moreover, one accepts radical differences in morphology and a plethora of non-cognate lexical substitutes as salient diagnostics for diglossia, I believe I have shown that Arabic is, and Persian is not, diglossic in principle, i.e., in terms of the narrow definition. The real question to be answered is not “How diglossic is Persian?” but “How did Persian avoid diglossia?”

5. A full examination of the question posed above is beyond the scope of this paper. Different diglossic situations have arisen under different historical and socio-cultural conditions, and are subject to modification in course of time. Latin-Romance diglossia was resolved in the course of western Europe’s political and cultural evolution, by way of a diminishing bilingualism. Turkish diglossia was engineered into oblivion during the 1930s-40s, and Greek diglossia was legislated out of existence in 1974. Modern French, in which a marked conservatism and puristic tendencies are opposed by a vernacular phonologically and syntactically out of sync with the written standard, might be on its way via a “widening structural schism” to becoming measurably diglossic (Joseph, 24, 36). Among the varieties of Persian, Tajik was arguably reduced from a state of diglossia (vis-à-vis classisizing madrasa-Persian) to the state of a standard-with-dialects under Soviet linguistic engineering; however, it may become diglossic again if the current tendency of the written and formal standards toward convergence with international Persian (barely distinguishable from Standard Persian of Iran) increases its divergence from most of the Tajik dialects.

One popular perception of the genesis of diglossia — that the spoken tongue of the uneducated falls away from the established standard of the written classics — does not usually reflect immediate historical reality. Insofar as concerns Greek, and modern Sanskritised Hindi vs. the spoken Hindi-Urdu of northern India, the converse has been the case: a politico-culturally motivated purist movement artificially elevated a retrospective formal style above the spoken standard. Allowing for the different problem of the rise of the Arabic vernaculars, this is largely true of modern Arabic. Certainly there was no lack of purists among writers and scholars of Persian from the later eighteenth century up until the 1940s, and it is conceivable that a movement for de-Arabicised Persian might have produced a diglossic situation in modern Iran. In the event, these efforts never resulted in either a popular movement or an official endorsement of purism in Persian; the Farhangestân, or Persian Academy, was a moderate institution (Perry 1985).

There appear also to be sound historical and linguistic reasons for Persian’s remaining homoglossic. Islamic New Persian evolved in symbiosis and competition with Arabic. Content for the first three or four centuries to resign to Arabic the status of “H” (the high language of scripture, liturgy and scholarship), written and spoken Persian thus developed side by side: sociolinguistic dichotomy there was, but by way of bilingualism, not diglossia. The Iranian intelligentsia wrote in Arabic while speaking in Persian (and writing popular essays and poetry in Persian). This state of hierarchical bilingualism was reversed after the Mongol invasions: Arabic throughout the Eastern Islamic world suffered a reduction in its intellectual domain, while Persian expanded into philosophy and the religious and secular sciences, written not only by Iranians but Turks, Indians, etc. Conversely and contemporaneously, spoken Persian was challenged for hegemony as the lingua franca of the same region by other vernaculars, the most important of which were several varieties of Turkish. Once again, Persian was in a hierarchical bilingual (not diglossic) partnership, this time as H to Turkish’s L — a situation that has lasted in many parts of the Iranian world up until the present time.

Obviously, bilingualism with Arabic and/or Turkish is a sufficiently polarised sociolinguistic dilemma, without complicating things by going diglossic. When native speakers of Turkish or Baluchi needed to reach a wider audience, they would use written or spoken Persian; and their style, in so far as they were educated, would tend toward the universal, or more
formal rather than less formal, standard. When the native Persian elite wished to express their national identity vis-à-vis the Turks or Arabs in their midst, they would use a Persian comprehensible to, and valued by, their congers of whatever socioeconomic station. The pressures for a fissure of this language of solidarity along diglossic lines were simply not there.

BIBLIOGRAPHY


John R. PERRY
University of Chicago
j-perry@uchicago.edu