

The Etymological-Comparative Dictionary of the Balochi Language: Progress Report of an International Project ¹

Language and Literature

Adriano V. Rossi ²

1 *Descriptive lexicography and traditional lexicography*

1.1 Unlike the case of the other major Iranian languages, nineteenth century linguists at the time of the large-scale 'classifications' of world languages at the turn of the century never compiled a medium-sized Balochi dictionary comparable to those already produced by Steingass (1892) for Persian, Bellew (1867) or Raverty (1860) for Pashto and Jaba-Justi (1879) for Kurdish.

Even judging that the time was not yet ripe for an etymological-comparative dictionary of the Iranian languages be compiled, as early as 1890 Wilhelm Geiger (1890-1891:111, 1891:402) was of the opinion that the compilation of a general Balochi dictionary was an urgent desideratum¹.

At that time, the following preliminary materials were available for Balochi lexicography to European scholars:

(a) extremely general ethnographic descriptions accompanied by short *word-lists* (Leech [1838], Masson [1843], Floyer [1882], etc.);

(b) single words and phrases scattered throughout the early grammatical classifications, mainly gleaned at second hand from material recorded by previous travellers and scholars (Bruce [1869], Gladstone [1873], Mockler [1877]);

(c) *word-lists* varying in size between, but generally of no more than some hundred words in the form of appendixes to first-hand descriptions of

¹ This paper was presented and read at the 2nd *International Conference on Balochi Language, Literature and Culture*, organized by Balochi Academy on its Golden Jubilee, held in Quetta, July 31st - August 1st, 2011, but subsequently not included in the Conference Proceedings.

² "L'Orientale" University of Naples, Italy.

varieties of coastal dialects (Pierce [1874] and Marston [1877]) and of north-eastern dialects (Dames [1881], further developed in Dames [1891] - the longest (3,500 *headwords*) of the *word-lists* compiled by this author).

These *word-lists* were merely collections of *glosses* (ie. explanations of word or phrases) collected in the course of simplified oral surveys of the type "how do you say X in Y?" and grounded only on the need to publish them. This has a textual and linguistic effect on its own (cf. Rossi 1987:486): "before using for the history of the language elements drawn from a corpus consisting of phono-graphic recordings of the speech acts of individuals it is necessary to use specific means, *which are philological in the broad sense*, in order to identify the variables flowing from the repertoire of the speaker in question into the speech event, from the distortions produced by the channel used (i.e. the specific transcription selected and applied by the describer). [...] it is always necessary to bear in mind the peculiar nature [...] of the informant-describer relationship which restricts the broader use of the speech event in question, thus preventing its verification by the linguistic community [...] they share the feature that the communicative event ends in the moment in which the written text is created in order to project into the future a past speech act *which is intentionally considered not to be repeatable or variable*, while literary texts in the classical manuscript tradition are open to the future in that they allow of, if collective and oral, editing, variants, special performances, etc.; if individual and written, republication, correction, rejections, etc."

Only Mansel Longworth Dames, who had already begun redeveloping the materials, part of which were to be included in the textual sections of his *Text-book* and in his *Popular Poetry of the Baloches* published in 1907, was engaged in those years on texts transcribed from oral accounts or recitation/performance of poems, although it is significant that the organization of his glossaries (1881, later incorporated in Dames 1891) was explicitly presented as having been elicited from oral survey by listing of glosses, not from texts or speeches (Dames 1881:2).

In view of the conditions of the materials available to Geiger (1890-1891, 1891), whose work (446 *entries* of his *Etymologie* [1890-1891] and the *Lehnwörter* contained in the appendix to his *Lautlehre* [1891]) still remains the only major etymological-comparative attempt ever carried out for Balochi (cf. below on Korn 2005), it is a happy circumstance that Geiger followed the "wise practice" of listing words from his sources without taking sides. He

himself was working on Bal. texts, and was fully aware of the need of a dictionary based on texts.

However careful Geiger may have been, any etymological collection of the kind of Geiger's lists is condemned to contain aberrant lexical material for the following reasons:

(a) incorrect identification of the phonetic shape of the forms selected as the *headwords* and their *lexicographical variants* due to the multivariety *idiolects* of informants; recording errors caused by contrastive phonemics problems (English-speaking describers are notoriously incapable of fixing the vowel nuances in any linguistic variety they are describing); errors of transcription from fieldnotes to the printed text or, as in frequent cases of scholarly bequests, errors of interpretation by the posthumous editor; misprints (which are lakhs in the case of diacritic marks, especially in the publications printed in the Middle East and British India, where e.g. vowels, single vs. double consonants, etc. were also confused);

(b) ambiguous *sense discrimination*, in the case of erroneous collocation of (a set of) item(s) within the **vocabulary system** or of semantic interference in informants' idiolects (which, in view of the complex, multilingual repertoires, are at the same time intradialectal, intralinguistic and interlinguistic); undue inference by the describer, etc.;

(c) erroneous attribution to a dialectal area, when the describer accepts the informant's statement: "in the area X they say Y" without checking it; or because of intra- or interdialectal interference in the (extremely frequent) case of internal (and/or nomadic) migrations, and consequent incoherent selection of the *main form* in the *lemmatisation* process;

(d) erroneous sociolinguistic evaluation, when the describer's competence is insufficient to ensure the identification of technical and other jargons etc., or to recognize the peculiar specifications of a particular type of text (religious, sententious, ironic, etc.).

1.2 The next thirty years saw the publication of Mayer and Gilbertson's dictionaries, the only small to medium size (10-15 thousand *headwords*) published so far which record Balochi *equivalents* in a widely spoken language (English); however, for practical purposes resulting from the needs of the civilian and military personnel posted to British India, the entries were

ordered according to their English *equivalents* ('English to Balochi') and are thus hardly usable for descriptive or comparative linguistic research.

Mayer's work consists of a host of miscellaneous records, perhaps made by more than one collaborator on different occasions (cf. Mayer 1910:3: "Mistakes and superfluities there must be in the first issue of a work like this, with so many dialects, and where Munshis totally untrained, and with no knowledge of grammar, will either not acknowledge the words of another, or purposely try to show themselves better men, or come too late to help one with proof sheets") and based mainly on the north-eastern varieties, much of his material being taken from Dames, and all of it being badly edited and very badly printed. Mayer, in spite of the declared intention to «leave out any dialect difference» and of having had such a proof-reader as M. L. Dames himself, is in fact an impressive mixture of different dialectal material often hopelessly misprinted" (cf. Mayer (1910:3): "my warmest acknowledgements are due [...] to M. L. Dames, Esquire, of the Indian Civil Service, for much help, and for looking over the sheets before going to press, as well as sound advice from time to time"). In those cases in which it is possible to identify the area of origin of a single Balochi word, Mayer's collection can anyhow be considered of some *onomasiological* utility, since it enters various Balochi *equivalents*, linked through various types of (*para*)*synonymic relationships*, and ordered under the relevant English *headwords*. In any case, it is impossible to reconstruct how Dames and Mayer interacted, and therefore to determine how Dames' publications and his oral advice could have influenced the lexical records on which Mayer based his *English-Biluchi Dictionary*. "To some extent Mayer reversed Dames's Vocabulary, using Dames improvements to his original texts" (Elfenbein 1985b:161). One should emphasize that if both Dames and Mayer "extracted vocabularies from their literary collections", both mixed this material with many other records of different origin; if it is correct to state that "from about 1897 T. J. L. Mayer also began to interest himself in the collection of Balochi ballads and he published, with Dames' help, a fair collection of them in India in the years 1900-1903", his work on the translation of parts of the Bible in Balochi should have started many years before, as it is proved by the following statement: "The Auxiliary Bible Society also published during the past year (1900) the four Gospels, the Acts, several of the Epistles, and the First Book of the Psalter, in Baluchi, translated by the Rev. T. J. L. Mayer, formerly of the CMS", *Proceedings of the Church Missionary Society for Africa and the East*, p. 277 (I owe this information to an archive research carried out in London in the 1980s by Drs Simonetta Calderini and Delia Cortese, Academic Research Assistance, London; see also Grierson 1921:334-335).

Gilbertson's *Dictionary*, although apparently four times the size (819 pp.) of Mayer's *Dictionary*, actually contains fewer *entries* than Mayer's, but it does have the advantage of being more accurately printed overall and of being (often but not always; independence from Mayer and Dames is lesser than hinted by Elfenbein [1985a:226]) based on the *idiolect* of the north-eastern informant Ghāno Khān Haddīānī (on whom cf. Gilbertson 1925:v: "our fellow-worker was "called to his fathers" before the work was half finished. He has left us with a number of unsolved problems, such as the use of the agentive case with an intransitive verb !").

It is a unique case in Balochi lexicography, in that for each Balochi word it offers an utterance (even if sometimes constructed rather ad hoc) contextualizing the relevant lexeme as corresponding to the English headword, often accompanied by author's comments on (para)synonymic or antonymic relationships with other Balochi words; references to Persian, Pashto, Siraiki, Sindhi and Urdu parallels are also frequently included, and are very useful in ascertaining the origin of a particular sense of a Balochi lexeme in the last century (better, the existence and usage in the 1920s of that particular word or phrase in the neighbouring languages).

1.3 It is significant that, although a considerable number of texts were transcribed after the publication of Geiger's etymological summaries (cf. Lewis 1885; Hittu Ram 1881; Dames 1891; Mayer 1900-1903; Geiger 1889-1893; Dames 1907), practically no need was felt by those who continued working in the field of Balochi lexicography to progress from compiling *word-lists* based on single idiolects to making systematic records from published texts (the following comment by Gilbertson [1925:xI] on Dames' *Textbook* is remarkable: "Above all, a complete vocabulary of all the words occurring in the text is needed").

The inadequacy of Dames' *Glossary of rare and obsolete words* vis-à-vis the huge mass of poetic material preceding it is commented by the following words by Elfenbein (1985b:163-64): "Perhaps the 'Glossary of Rare and Obsolete Words' is the most puzzling of all. In its nine pages of words, all cited without loci, we find in the very first columns *ad□ay*, *ad□iθa* 'to erect' (also in Mayer *āday*, *adiθa*; both are wrong for *ād□ay*, *ād□iθa*), *āhū* 'a deer', *akbat* 'somehow' ... *badh* 'an enemy', *badh* 'evil' ... *bēm* 'fear', and so on. Of the 49 words in our first specimen [...] which can reasonably be described as uncommon, only seven are in Dames' list, and of them one with a wrong meaning and two do not correspond to his text". On

the circumstance that Dames' *Glossary* is rendered unserviceable by the lack of references to the text, see e. g. Émile Benveniste's personal experience in his research on Bal. *dushkīsh* (1966:49 fn. 4): "Il est inconceivable que, dressant une liste de mots rares, l'auteur n'ait pas jugé utile d'en donner la référence. Nous avons dû lire *tout* le recueil de Dames pour découvrir cet exemple de *dushkīsh*. N'en ayant trouvé aucun de la form « *dushk* », nous ne pouvons même en garantir l'existence". Since the digital text of The popular poetry of the Baloches is now available at the *Archives of the Balochi Dictionary Project*, Naples, we are now certain by electronic interrogation that the form *dušk* is not present in Dames's book (nor any other published book as far as we know).

It is therefore all the more astonishing Dames' statement (1907.2:192) that his *Glossary* should be considered "as supplementary to the vocabularies of Balochī already published, such as those contained in [...] *Textbook* (1891), and Douie's translation of Hētū Rām's *Biluchī-nama* (1885)". Statements of this kind reveal lack of awareness of how unwise it is to introduce into a single book the results of lexicographical practices so far divergent as glossing from texts which represent literary and/or archaic sublexica, and producing a series of concordances "Balochi X = English Y" by means of questioning bilingual informants.

1.4 The first generation of Balochi lexicographers thus came to an end in the 1920s without the need clearly being felt to produce as quickly as possible a 'modern' dictionary from any Balochi dialects into any European language, a dictionary with a clearly defined linguistic basis (e.g. north-eastern, coastal or central) and enhanced by covering different specialized sublexica (technical, naturalistic, literary, etc.). The next generation, that of Georg Morgenstierne and Ivan I. Zarubin, experienced a greater availability of folklore material (Zarubin 1932 and 1949) in the 'central' dialect of Marw (then in Soviet, now in independent Turkmenistan), but no further lexical work was planned.

The credit for making a preliminary critical updating of Geiger's *Etymologie* goes to Morgenstierne (1932; 1946-48), who also was the first scholar with an etymological interest in Balochi lexicography to make use of the new material from Soviet Turkmenistan and to introduce a systematic comparison with Brahui, the main non-Iranian language in contact with Balochi, which had in the meantime become more deeply known through Bray's (1909-1934) masterly synthesis (cf. Rossi 1979a on the importance of contact linguistics in Bal. lexicography). Zarubin in his turn introduced a complex transcription of Marw Balochi, first phonemic (1932) and subsequently phonetic (1949), but his interests never convinced him to focus

his research on Balochi lexicology, though his descriptions of East Iranian languages (Bartangi, Roshani, Shughni; cf. especially Zarubin [1960], where the *Dictionary* occupies the bulk of the book [pp. 83-288 with a Russian-Shughni finder index at pp. 289-386] demonstrate that he had a real bent for lexicographical description.

1.5 The birth of modern lexicography by Baloch intellectuals dates from the late 1960s. The different trends since then prevailing among Baloch lexicographers in Pakistan ('national' Pakistani Balochi lexicography) showed two main aims:

(a) to co-operate in the production of a vast array of centrally developed tools (from the Markazi Urdu Board founded and directed by Ashfaq Ahmad with a large staff of usually well qualified collaborators, cf. Rahman 1999:263 ff.) including folklore, *onomasiological dictionaries*, *word-counts*, ready-to-use basic *word-lists* for Urdu speakers (Urdu - Balochi), dictionaries to enable Balochi speakers to read and speak Urdu, etc., for the principal purpose of relating Balochi culture to the cultures of Pakistan through Urdu;

(b) to compile (handbooks and) dictionaries for the specific purpose of 'protecting' the Balochi lexical heritage from the encroachment by Urdu (cf. Rahman 1999:281-84).

The first group includes the following:

(1) Ata Shad (1968, 1,000 *entries*), short *onomasiological dictionary* (social terminology, ecosystem, production modes, etc.) based on Balochi *folk taxonomies*, with Urdu *definitions* for Urdu speaking users with a particular interest in folklore;

(2) Mitha Khan - Ata Shad (1972), medium-large dictionary (45,000 *entries* and *sub-entries*) designed for a Balochi speaker wanting to master the *basic vocabulary* of Urdu and based on a model (*Larousse Dictionary*, cf. Rossi 1982:161 fn. 28) devised for all regional languages of Pakistan; a mix of dialects is intentionally (*Pišlafz*, i) pursued, but no dialect distinction is marked; Balochi *definitions* and *explanatory equivalents* explain to Balochi-speakers all Urdu words whose cultural content was considered of difficult understanding in the *target language* because of cultural and linguistic *anisomorphism*;

(3) Ashfaq Ahmad (1974, 2,500 *headwords*) listing the Rakhshani Balochi *equivalents* (words and phrases) of some thousands of Urdu *core words* (in many cases for different Urdu *headwords* the same Balochi *equivalents* are repeated with a considerable quantity of semantic oversimplification) based on *word frequency counts*, and aimed at offering to Urdu speakers the equivalents of their basic lexicon in the six major regional languages of Pakistan. Mumtaz Ahmad (1985) gives a Balochi-English index of Ashfaq Ahmad (1974). "The glossary consists of 2,500 entries of Balochi words and phrases most frequently used in everyday discourse and in the contemporary Balochi newspapers, periodicals, and publications of general interest in Pakistan and Iran" (Mumtaz Ahmad (1985: iv).

The second group consists of a series of works conceived and printed in Quetta and Karachi in the last four decades:

(1) Mitha Khan - Surat Khan (1970, 30,000 *headwords*) contains Balochi entries followed by a series of synonyms or parasyonyms, sometimes mixed with (areal) variants (no dialect distinction is marked); Balochi words with peculiar cultural meanings are glossed by *encyclopedic definitions* (in Balochi language), but its being based on an Urdu word-list and the lack of a clear dialectal profile reduces the utility of the Dictionary ("no attempt is made to distinguish between dialects, so that we have here again a 'string of vocables' style, in which important variants [...] are mixed unmercifully and without explanation. Underlying it was an Urdu word-list for which simple Balochi equivalents were sought; thus most of the vocabulary of classical poetry is missing, as well as e.g. most of Mayer", Elfenbein 1985a:226).

(2) Nasir Khan Ahmadzai (1975) is a list of 3,000 Balochi simple and compound verbs, with Brahui, Urdu and English equivalents;

(3) Mir Aqil Khan Mengal (1990, first issue of an unfinished work, containing only letters "alif" to "xe") lists more than 3,000 Balochi equivalents of Persian words of Middle Iranian antiquity ("It may help those who want to introduce new words in the Balochi writings as counterparts for various terminologies", *Preface*, p. 18);

(4) Abdul Qayyum Baloch (1997) lists 3,500 words and phrases divided into main semantic fields, mainly Rakhshani, with Urdu and English translations;

(5) Mir Ahmad Dehani (2000) lists 7,800 headwords (Coastal and Rakhshani, dialectal variants sometime provided), with Urdu translations;

(6) Sayid Hashimi, *Sayid Ganj*, Bal. monolingual dictionary mostly based on literary Kechi, contains some 20,000 headwords (words and phrases, with examples, remarks and scattered etymologies, unfinished, the first part was revised by the author); the printed (and abridged) version of some 900 pp., contains many editorial interventions, and represents the best standard of modern Bal. lexicography (copies of the original, unedited ms. are kept at the Sayyad Accademy, Karachi and at the Bal. Etym. Dict. Archives, Naples);

(7) Bozdar (2003, English to Balochi) and Bozdar (2004, Balochi to Urdu) are dictionary listing some thousands of Bal. neologisms.

To these one should add Rzehak, L. and B. Naruyi (2007), a revision of A.L. Grjunberg's unpublished notes for a Balochi-Dari-Pashto-Russian based on the Afghan Rakhshani dialect as spoken by Abdurrahman Pahwal. The files were compiled at the beginning of the 1990s within the framework of a formal agreement between the Institute of Linguistics of the Soviet Academy of Sciences and the Afghan Academy of Sciences. Lutz Rzehak has worked (together with B. Naruyi) on the original ms. kept at Humboldt University, and produced a completely revised and updated edition, reversing the original Russian definitions into English; words and phrases are ca. 10,000.

1.6 In addition to the above publications the following lexical collections deriving from research made in the field over the past thirty years deserve some lexicographical, sociolinguistic and/or etymological relevance:

(a) Spooner (1967): 400 words, coastal and Sarawani varieties of Iran, sociolinguistically commented;

(b) Barker-Mengal (1969.2:403-603): 3,500 words often sociolinguistically commented (with an English-Balochi index), Rakhshani variety of Noshke;

(c) Foxton (1980): 2,100 words, Kechi and coastal varieties;

(d) Collett (1983:68-116): 2,300 words, Kechi varieties (with an English-Balochi index);

(e) Coletti (1981:35-56): 700 words, Sarawani variety of Iran, sociolinguistically commented;

(f) Farrell (1990:77-90): 650 words and phrases, Karachi Balochi.

The work of indexing (groups of) texts was begun in recent times: Elfenbein 1963 (1,800 words, with English-Balochi index), a glossary of words taken from the collection of Marw Balochi stories published in USSR (for which cf. Rossi 1979a:xvi n. 47; Elfenbein 1985a:227); Elfenbein 1983:123-56, containing 1,700 words found in ms Cod. Or. add 24048 of BL and referring to literary forms of coastal varieties; Buddruss 1989 (approx. 500 words, complete indexing of Mohammad Hanif's Life-history in Afghani Rakhshani), and Elfenbein 1990.2, containing more than 6,000 headwords (without counting dummy entries, with some 3,000 sub-headwords) from texts in all dialects mainly occurring in the Anthology ("All words which occur in the texts [...] are in principle included", Elfenbein 1990.2:xix; "the largest collection ever assembled up to now", Elfenbein 1990.1:2).

1.7 Lastly, the overall dialectological framework drafted by Elfenbein (1966) was updated and more accurately defined (Elfenbein 1989a, 1989b, 1990.2:vii-xviii). For the relevant problems of phonological typology, cf. Rossi 1979b; for the influence of the dialectological assumptions on the standardisation practices, cf. Jahani 1989.97 ff.; historical phonology is treated exhaustively in Korn 2005.

2 Problems of etymological approach

2.1 I have discussed elsewhere (cf. Rossi 1979b for the phonology and Rossi 1979a for the areal lexicography) the difficulties raised in the field of comparative Iranian linguistics by a situation such as that described above. Suffice it to recall how the etymological reconstruction of any Iranian word can be complicated by the difficulty involved in checking whether the Balochi equivalents of any determined Iranian word actually exist in the two meanings of the term: i.e. whether they have been recorded in the small-scale works described above (many of which are arranged in the alphabetical order of the English equivalents), or whether their graphic fixing (archaic, incorrect or ambiguous) makes their identification impossible; or else how in areal lexicographical research a given Indo-Aryan word may be considered as a loanword, e.g. from Brahui, not being recorded in Balochi, etc.

2.2 The first point to stress is that there are no large comparative Iranian dictionaries like the Indo-Aryan ones compiled, respectively, by Turner (1966, more comparative in nature) or by Mayrhofer (1956-80, cf. also 1992-2001, more etymological). Indeed, the latter scholar explicitly declares, in his stimulating essay *Zur Gestaltung des etymologischen Wörterbuches einer*

Grosscorpussprache (1980), that he was tempted to list "surreptitiously" (*heimlich*), under the Old Indian headwords of his *Kurzgefasstes Etymologisches Wörterbuch des Altindischen*, precisely that material which was lacking in the Iranian etymological research. Three volumes of a Russian language etymological dictionary have recently appeared (Rastorgueva-Èdel'man (2000, 2003, 2007), but they contain only reconstructed Iranian roots starting with letters 'a' to 'h'; even if many criticisms have been advanced in their regard, I consider them anyhow as a useful contribution as far as the etymological material covered is concerned. Another useful tool is Cheung's work on Iranian verbs (2007), on a preliminary version of which Shahbakhsh (2004) is based, even if his decision to introduce the laryngeal symbol in proto-Iranian reconstructions is debatable.

Of course, this does not mean that no materials of use as an etymological starting point for modern Iranian languages actually exist: for instance, there is Morgenstierne's (1927) etymological dictionary of Pashto, and its updating published posthumously (2003); Horn's (1893) *Grundriss* with a critical appendix by Hübschmann (1895) for Persian, Geiger's (1890-1891) etymological lists, partially updated by Morgenstierne (1932) and recently by Korn (2005) for Balochi, Justi's 1879 edition of Jaba's lexical notes, and the first volume of a modern etymological dictionary in Russian (Tsabолоv 2001) for Kurdish, Abaev's (1958-89) monumental work on Ossetic etymology.

The works described above follow roughly (with some notable exception) the same approach: i.e. to separate the original lexical core (*Erbwörter*) representing the earliest lexical origin - the only one deemed of interest to linguists -, from the subsequent overlays of *Lehnwörter* (structured loanwords) and *Fremdwörter* (foreign words) according to the degree of gradual assimilation to the 'original body' of language. The etymological lists mentioned above are actually lists of words which, for mainly phonetic reasons, are believed to have been part of the respective languages ever since the earliest recorded times, and even before, in the reconstructed phases (*Proto-phases*) of the language. On the other hand, words in any way suspected of having been imported through another language, whether Iranian or not, are rigorously excluded. Thus Geiger (1890-91) lists for Balochi only 450 words which he considers to form the original core of the lexical system, although he suspects that a hundred or so of them actually came through Persian; to these Geiger (1891) adds a list of about 300 loanwords, the vast majority of which from Persian, but he never thought to compile a real etymological dictionary for Balochi as a whole.

2.3 The question of Persian loanwords is one of the dilemmas facing Iranian etymology both in the case of languages with a comparatively extensive written tradition and in those with a very recent and/or practically non-existent written tradition (for the perceptions of Baloch intellectuals of the loanwords question in the planning practices for Balochi, cf. Jahani 1989.124-28). Persian has acted as a superstratum for other Iranian languages on at least three occasions: during the Sasanian expansion, 3rd-4th century A.D.; at the time of the establishment of Arabo-Persian bureaucracy, 7th-9th century A.D.; during the introduction of a pan-Islamic scientific prose by Arabo-Persian bilinguals (e. g. Avicenna) in the 10th-12th century A.D. To make a comparison with the Romance languages in Europe, also Latin had a superstratum role in at least two different periods, that of the early Romanization, which varied from area to area, and that of the rise of modern European scientific prose, starting in the late Renaissance. Of course, the Latin of the second wave was quite different from that of the first, having in the meantime become not only a technical language but also a dead one. Moreover, also modern Persian was quite different from early Persian, as this was a technical and elitarian language and, among other things, lexically much closer to Arabic than to Persian. What makes the case of Romance languages in Europe totally different from that of the Islamicized Iranian world is that, as a result of contemporary observations, indirect evidence and direct reading of texts, our knowledge of the two Latin periods is fairly substantial, while in the latter case there is practically no evidence at hand concerning the first two waves of Persianization, even in the Persian documents themselves. Furthermore, modern Persian, to which most of the available evidence refers, is known to us through documents written from 9th-10th century A.D. on, Kurdish from 15th-16th century and, with the same reservations as for Kurdish, Balochi for just over two centuries. This gives a better idea of just how desperate the situation may appear at present.

2.4 To come back to the question of Persian loanwords in Balochi, it must be admitted that there is no way of establishing their relative chronology except in the few cases in which their semantics refers to meanings than can luckily be dated as entering or leaving usage at a given time in the history of the Persian language. However, here begins another tale of woe in that, just as no general Iranian etymological dictionaries exist, so of course the majority of Iranian languages also lack descriptive dictionaries compiled on historical principles. In any case, again, for highly multilingual linguistic traditions which - as is the case of Balochi - basically lack fixation in written form, in view of the fact that the socio-cultural structure prevailing in the country prevents a single standard from being accepted by the entire community (cf.

Jahani 2000 passim, especially Farrell 2000 and Axenov 2000), it may be sometime misleading to speak in terms of 'loanwords': they are more likely to be *Fremdwörter*, ie. current foreign words (such as the German word *Länder* is in an English newspaper article), than *Lehnwörter*, ie. structured loanwords.

2.5 Whether or not to include *Fremdwörter* in etymological dictionaries is probably the theoretical problem implying the highest practical implications for Balochi lexicography (cf. Korn 2005:19-20). If the most extensive option is chosen one might well have to include into a Balochi dictionary the whole Persian and Urdu lexica, since the two major cultural languages in Iranian and Pakistani Balochistan are those in which most everyday linguistic creation occurs. On the other hand, can it be claimed that all the particular sub-lexica associated with changes in the environment, which are of course based on Persian and Urdu as the changes, contributions and new ideas come from the outside, are extraneous to the language? In other words, is it true that every time a traditional animal transport is replaced by a mechanical means of transport, or a traditional dwelling is replaced by a new building structure, a part of the *echt* (=original) Balochi lexicon disappears and the speakers acquire an equivalent number of *Fremdwörter*? This does not seem to be the case: indeed psycholinguistic experiments carried out in Balochistan by the present writer and other scholars seem to point to something similar to the linguistic situation that had been convincingly described many years ago by Gumperz (1964) with the commutation of Panjabi/Hindi code in a group of Panjabi speakers in Delhi, namely that the only way of explaining what takes place in the speaker's mind is to argue in terms of repertoire and domain: the speaker has several languages and/or sub-lexica in his repertoire which he uses in accordance with the prevailing rules of linguistic etiquette, and according to the communicative situation; in the 'home domain' he will use a language (or a mix of languages, occasionally differentiated according to the interlocutor), another in the 'school domain', another in that of 'politics domain', etc. (cf. Rossi 1982:167-69; on this matter cf. also Jahani 1989:76-79, 124-28, Jahani 2000, Tan 2000).

2.6 However, arguing in terms of repertoire and domain (to use Fishman's sociolinguistic terminology) leads to trouble, or rather to such grave doubts as to produce definitive systematic silence. What exactly is being listed in the wordlists contained in native glossaries (Persian with Persian and Arabic glosses, Kurdish with Persian and Arabic glosses, Pashto with Pashto, Persian and Urdu glosses, Balochi with Persian and Urdu glosses) which were drawn upon when compiling the first western bilingual dictionaries, and repeated in

their successors; or in the bilingual wordlists made by the early travellers, missionaries, officers and so on, and again utilized in the early western bilingual dictionaries (and their successors)? Paradoxically, in the latter case, viz. of lists based on relatively recent recordings of individual idiolects (generally those of the guides in the case of travellers, of the attendants in the case of officers, and of seminarists in the case of missionaries), they may be said to be listing themselves; in other words they record a situation, or domain, in which two persons speak to each other simply with the purpose that one of them may write down (sections of) the lexicon of a language X known by the other.

To be consistent, it is therefore necessary to eliminate completely, or at least list separately, all words taken directly or indirectly from such glossaries. According to Malkiel (1976:52) no self-respecting lexicographer would ever take into consideration words transmitted through glosses alone. On the other hand, however, it is and will remain impossible for the historical and comparative linguist to have direct access to firsthand materials for all the languages he is working with: just to give only one example, all 38 Balochi words contained in Pokorny's *Indogermanisches Etymologisches Wörterbuch* have been taken, without exception, from such glosses repertoires. One can imagine the number of further etymological dictionaries of different languages in which they now appear reproduced as they were printed in Pokorny's dictionary, including the (not few) misprints added by the editors and/or printers of the dictionaries.

3 The Italian etymological project

3.1 A research team led by the author and funded by the Italian National Council for Scientific Research (CNR) has been working since 1977 at L'Orientale University, Naples on "Preliminary Work to the Compiling of an Etymological-Comparative Balochi Dictionary". In this project, priority has been given to setting up a unified reference filing system of all the lexicographical material, ordered by Balochi headwords and with English as the output language of the explanations (to be added in a subsequent phase).

Since 1982 the research team has also been receiving funds from the Italian Ministry of Education for a project at L'Orientale University, Naples, also directed by the author, and parallel to the CNR-financed project, as well as funds for individual sub-projects of a more specifically ethnolinguistic nature. The latter have been part of the Italian National Project on "Ethnolinguistics of the Iranian area" originally directed by Prof. Gherardo Gnoli (until 1983) and by the author since 1984.

After the decision in June 1983 to incorporate the above research in the activities of the IsIAO Center for Asian Lexicography, at the Italian Institute for African and Oriental Studies, Rome, within the framework of a five-year CNR-IsIAO agreement, a scientific committee was formed which included among others the author and the late Prof. Alessandro Bausani ("La Sapienza" University of Rome); since then the Project denominated "Etymological-comparative Dictionary of the Balochi language" has developed as a project of the Centre for Lexicography, IsIAO managed in joint venture with L'Orientale University, Naples (see details in Rossi 1986).

3.2 From the very beginning a series of guidelines were decided: (1) the work language of the Dictionary should be English, so that the maximum number of scholars could joint efforts during the elaboration phases of the Project, and have access to its final printed form; (2) Balochi should be written in Roman transcription, in order both to bypass any problems of standard of script, and to enlarge the number of users outside Balochistan; (3) words and examples should be selected out of a substantial corpus of real language build on the basis of all published texts and a significant amount of new material; (4) an effort should be made to include references to past lexicographical works, but without depending on them; (5) three main varieties (Coastal, Rakhshani and Eastern) should be included, within the same headword when possible or with separate headwords when a particular dialect form is absent from the other dialect(s), with a preference to base the Dictionary on the Coastal (Kechi) literary standard (last feature being open to further discussion).

3.2.1.1 **The texts**

The team at L'Orientale University, Naples and their international collaborators have already edited in electronic form (for general problems about extracting dictionaries from electronic corpora cf. Sinclair 1987) the following texts, for an approximate total of one million words of running texts:

- (1) all texts in Roman transcriptions published from late XIX century to Elfenbein's *Anthology* (running texts of at least some lines, except the examples in the grammars and the texts contained in Barker-Mengal 1969.2 because of the editorial treatment stated *ibid.* at p. 171 ff.), edited by various teams (coordination and revision by E. Filippone [4.20 MB]);
- (2) translations of portions of the new version of the Bible edited by T. Farrell (0.34 MB);
- (3) unpublished short stories in the Marw dialect edited by A. L. Grjunberg (0.25 MB);

- (4) unpublished short stories and proverbs in the Iranshar dialect edited by Rahim Bandawi (0.08 MB);
- (5) unpublished short stories in the Karachi dialect edited by T. Farrell (0.15 MB);
- (6) unpublished interviews with pahlawans and poetry (mainly literary Kechi) edited by Badal Khan (0.5 MB);
- (7) articles from Balochi cultural periodicals (mainly literary Kechi) edited by Badal Khan (0.85 MB).

3.2.1.2 In order to enlarge the corpus to dimensions at least three times the present ones (presently 1,500 pages, 1 million words; minimum extension required for a balanced corpus: 5,000 pages, 3 million words), teams located in different centres shall input data from homogeneous collections of texts, either already published or to be elicited on the basis of ethnolinguistic questionnaires (a selection of semantic fields/domains was discussed with Balochi scholars at the Balochi Academy, Quetta and has to be further defined in agreement with the scholars involved in the international framework) according to the following outline:

- (1) classic poetry: team originally directed by Abdullajan Jamaldini, University of Balochistan, Quetta, presently by Badal Khan, L'Orientale University (preliminary research on the lexicon of the oral poetry is contained in Badal Khan (1994) doctoral dissertation);
- (2) modern prose: team originally directed by Carina Jahani, University of Uppsala, in association with doctoral students at Uppsala University; scholars from the Balochi Academy, Quetta, and the University of Balochistan, Quetta;
- (3) Marw folklore: team originally led by the late Aleksandr L. Grjunberg, Institute of Linguistics, Russian Academy of Sciences, St-Petersburg, who has been preparing for years a re-edition of all Marw material with the addition of recent recordings from Turkmenistan; Grjunberg's mss. are now in possess of L. Rzehak, Humboldt University, Berlin, who has been acting (together with Bedollah Naruyi) as editor in the Balochi-Pashto-Dari-English Dictionary project outlined in 1.5 point (7) above; Serge Axenov collected further folkloric material in seven trips to Turkmenistan (six hours of recorded texts, ca. 150 printed pages, cf. Axenov 2006: 27)

(4) geomorphological lexicon and lexicon of the body parts: Ela Filippone, University of Tuscia, Viterbo (more than 1,700 lexical items recorded in many interviews with informants from the main dialect areas);

3.2.2.1 Indexing of existing dictionaries

At the same time, card-indexing (in earlier phases with paper cards, then electronically) began for a separate lexical database, putting all entries in the alphabetical order of the Balochi words and transferring from different languages into English the lexical explanations wherever necessary. The present status is as follows:

(a) Mayer 1910: fully card-indexed (15,000 cards); revision partially completed, the files can be consulted in paper form at IsIAO-L'Orientale;

(b) Gilbertson 1925: fully card-indexed (10,000 items); revision completed and text available on CD-ROM (*An Index to Gilbertson, English-Baluchi Colloquial Dictionary*: sample page in Appendix 1);

(c) Mitha Khan - Ata Shad 1972: fully card-indexed (50,000 cards); the revision will be carried out as soon as fitting personnel will be available at IsIAO-L'Orientale;

(d) Mitha Khan - Surat Marri 1970: fully card-indexed (30,000 items); revision has been carried out for 80% of the cards (draft – not in alphabetical order but electronically searchable, including all parasyonyms listed in the second column which are not easily retrievable in the book form - available at IsIAO-L'Orientale).

The smaller lists described above (1.6) will be incorporated in the general index in a later phase.

In the framework of the preliminary research mentioned above, the following sublexica have been to date processed by the IsIAO-L'Orientale team, and are already incorporated into the Dictionary database:

(1) kinship lexicon and human settlements: A. V. Rossi (preliminary research on this lexicon is contained in Orywal (1982) and Ferraro (1990) doctoral dissertations; two hundred lexical items with Iranian etymological cognates already available);

(2) geomorphological lexicon and lexicon of the body parts: E. Filippone, La Tuscia University, Viterbo (more than 1,700 lexical items with Iranian etymological cognates already available);

(3) colour lexicon: A. V. Rossi, L'Orientale University (100 lexical items with Iranian etymological cognates already available);

(4) botanical lexicon: very preliminary work was carried out by P. Gionta (IsIAO-L'Orientale Archives) jointly with A. H. Bajoi, PhD, Agricultural Research Institute, Quetta (cf. Gionta 1982-83).

3.2.2.2 Unpublished lexical files

In a subsequent phase of the indexing, data extracted from the following collections of unpublished cards and notes will be systematically added to the main lexical database:

(1) Tim Farrell, *Balochi Ganj*. Balochi-English Dictionary, more than 19,600 words and phrases with many examples, originally based on the Kechi/Karachi dialect, and subsequently expanded to many words taken from existing dictionaries (on electronic support; the printout kept at L'Orientale University Archives is dated "4th draft, 2001");

(2) Agha Nasir Khan Ahmadzai, Rakhshani dialect (including dialectal variants), unpublished notes for a Balochi-Brahui-Urdu-English Dictionary, 2,000 pp., lists of more than 15,000 words and phrases, copy of the ms. at L'Orientale University Archives (compiled in the 1970s and the 1980s; on the first page of the ms. is written: "checked by the late Mir Gul Khan Naseer 7/9/83");

(3) Hans Strasser, 40,000 handwritten cards mainly based on Rakhshani, containing words and phrases (with German and/or English definitions), many examples (part of which with translations) taken from literary texts printed and/or elicited in the 1960s, and further annotations of sociolinguistic, comparative etc. relevance; original file (compiled in Pakistan in the 1960s and 1970s) at the Archives of the Austrian Academy of Sciences (Wien); a preliminary agreement for its computerization has been drafted between the Iranische Kommission and L'Orientale University (for a summary description cf. Rossi 2004-2006)

(4) Balochi-Persian Dictionary, some 10,000 words and phrases, produced by a school teacher from Iranshahr;

(5) *Naguman*. English-Urdu-Balochi Dictionary, some 3,000 words with many neologisms, mostly Coastal (digital searchable copy, unfinished, available at L'Orientale University).

3.3 The etymologies

Lastly, the central team will prepare electronic files aimed at recording all etymologies appeared in the scientific literature after Geiger 1890-91. The compilers will have at their disposal a list of more than 500 bibliographic references collected since the start of the Project (etymologies for Balochi words scattered in scientific literature of difficult access, 30% not included at all in Geiger 1890-91, 1891, with a further 20-25% handled in a wrong or incomplete way), a database containing ca. 1,700 Balochi body part terms and terms of the geomorphological lexicon with their Iranian cognates (partly used by Filippone 2006, 2010), and the thematic files mentioned in 3.2.2.1 above. Elfenbein 1985a and Rossi 1998 provide critical revisions (176 and 255 items respectively) of the Balochi etymologies contained in Abaev 1958-89 and Bailey 1979; a project funded by the Italian Ministry of University has permitted the revision of the draft articles for the 1,500 principal lexical bases of the central core of the three main Bal. dialect areas.

The results of this section of the Project will be published progressively in a cumulative interim form.

3.4 The structure of the etymological Dictionary

Each entry should take into account the following main factors [cf. sample articles in Appendix 2]:

(a) areal variation in lexical items, both with regard to different meanings in particular dialects and phonetic and/or morphological differentiation;

(b) phraseology (it may sometime help to reconstruct an old meaning, and therefore the true etymology, of a word no more analysable because of major changes in its phonetic structure);

(c) etymological-comparative material (with full references to previous Bal. dictionaries and glossaries when needed). In the example at Appendix 2, ie. the present draft of the article treating *gwūt* 'wind', not all modern Iranian cognates of the word are given, but only those of the modern languages more directly connected to Balochi, ie. Persian, Kurdish and Pashto, plus those from Old and Middle Iranian languages necessary to illustrate that: (1) the meaning 'wind' is the core meaning of the word already in the oldest forms of

Iranian (in this case Gathic Avestan); (2) a divine/magic meaning is also of very old date within Iranian (and therefore is not a Bal. semantic development, but an inherited one); (3) in some modern Iranian languages its core meaning is associated to the concept of ‘arrogance’; (4) metaphorical projections in the botanical lexicon may be borrowed from surrounding languages or be autonomous developments (cf. Bal. *gvāt-rōf* name of a plant, and NPrs. *bādrō* name of a plant) etc.

Just how far to go in offering the materials for the areal variants is still an open question. Obviously, as no recognized standard do exist for Balochi yet, the lexemes supplied must necessarily be of a comparative and interdialectal nature, at least to the minimal degree contained in the model used by Morgenstierne (1927, 2003) for Pashto. Hopefully also the phraseology - an essential characteristic for a modern small to medium-sized dictionary which would not repeat the nineteenth century structure of mere wordlists should refer to all important regional variants, even if this could be considered as a luxury for an etymological dictionary.

It is not even possible to follow Abaev's model which, although it can be considered as the best etymological dictionary of any Iranian language ever produced, cannot be strictly reproduced because it is based on a nationally widespread written standard (the same could be said even for Tsabolov 2001, even if the relationships between Kurdish Kurmanji and Sorani may recall the Bal. situation). Variants of a social, situational and stylistic nature should be treated, as the occasion arises and if relevant to historical reconstructions, within the mentioned areal sections, in view of the fact that socio-linguistic rules cannot apply to Balochi as a whole.

Two main problems concerning the phonetic aspects have also been left open at the present stage of the Project:

a) which diacritics should be used in the transcriptions of Balochi (a working semi-phonemic transcription has been worked out for the preliminary phases, even if all published texts in Roman characters have been provisionally input in their original transcriptions);

b) how to interpret forms taken from publications where the Urdu script does not allow any distinction between vowels, diphthongs, various diaphragmatic characters, etc.

The first problem will probably be solved by means of a semi-phonemic transcription, with retention of the original transcriptions in cases in which no

reinterpretation is possible (these transcriptions will explicitly be marked with "..."); the second by the interpretation of native speakers originarious from similar dialect areas (further details in Rossi 1979a:xxiv-xxviii).

3.5 The present project is based on the assumption that the international edition will use English as its medium language and have transcriptions of the Balochi entries into Roman characters. A special edition aimed at Balochistan people could have the Balochi (and Urdu) language(s) as medium, and the Balochi entries could be in Urdu script along with transcriptions into Roman characters.

References

- Abaev, V. I., 1958, *Istoriko-ètimologičeskij slovar' osetinskogo jazyka*, 1, Moskva-Leningrad; 2, Leningrad 1973; 3, ibid. 1979; 4, ibid. 1989
- ‘Abdul Qayyum Baloch [‘Abdul Qayyūm Baloč], 1997, *Baločī būmīā*, Quetta
- Ahmad, Ashfaq et al. [Ašfaq Ahmad], 1974, *Haftzabāni lughat*, Lahor
- Ahmad, Mumtaz, 1985, *Baluchi Glossary. A Baluchi-English Glossary: Elementary Level*, Kensington, Md.
- Ahmadzai Baloch, Agha Nasir Khan [Āyā Mīr Nasīr Xān Ahmadza’ī Baloč], 1975, *Baločī kārgonag*, Quetta.
- Ata Shad [‘Atā Šād], 1968, *Baločī-nāma*, Lahor
- Axenov, S., 2000, “Balochi Orthography in Turkmenistan”, in Jahani (ed.), 71-78
- Axenov, S., 2006, *The Balochi Language of Turkmenistan. A Corpus-based Grammatical Description*, Uppsala
- Badalkhan, S., 1994, *Poesia epica e tradizioni orali balochi: i menestrelli pahlawān del Makran*, PhD Diss., Napoli
- Bailey, H. W., 1979, *Dictionary of Khotan Saka*, Cambridge

Barker M. A. R. and A. K. Mengal, 1969, *A Course in Baluchi*, 1-2, Montréal

Bellew, H. W., 1867, *A Dictionary of the Pukkhto or Pukshto Language*, London

Benveniste, E., 1966, *Titres et noms propres en iranien ancien*, Paris

Bray, D., 1909, *The Brahui Language*, 1, Calcutta 1934, *The Brahui Language*, 2, 3, Delhi

Bruce, R. I., 1869, *Manual and Vocabulary of the Biluchi Dialect*, Lahore

Buddruss, G., 1989, *Aus dem Leben eines jungen Balutschen von ihm selbst erzählt*, Stuttgart

Cheung, J. 2007, *Etymological Dictionary of the Iranian Verb*, Leiden

Coletti, A., 1981, *Baluchi of Mirjave (Iran). Liku Couplets*, Roma (unpubl.)

Collett, N. A., 1983, *A Grammar, Phrase Book and Vocabulary of Baluchi*, s. 1. (1986²)

Dames, M. L.,

1881, *A Sketch of the Northern Balochi Language*, London

1891, *A Text-Book of the Balochi Language*, Lahore

1907, *Popular Poetry of the Baloches*, 1-2, London

Dihani, Mir Ahmad [Mīr Ahmad Dihānī], 2000, *Mīr Ganj*, Karachi

Elfenbein, J.,

1963, *A Vocabulary of Marw Baluchi*, Naples

1966, *The Baluchi Language. A dialectology with texts*, London

1983, *A Baluchi Miscellanea of Erotica and Poetry: Codex Oriental Additional 24048 of the British Library*, Napoli

1985a, "Balochi from Khotan", *Studia Iranica*, 14:223-238

1985b, " 'Popular poetry of the Baloches' ", in *Papers in honour of Professor Mary Boyce*, I, Leiden, 159-178

1989a, "Baluchi Language and Literature", *Encyclopaedia Iranica*, III, London – New York, 633-644

1989b, "Balōčī", in R. Schmitt (ed.), *Compendium linguarum iranicarum*, Wiesbaden, 350-362
1990, *An Anthology of Classical and Modern Balochi Literature*, I: *Anthology*, II: *Glossary*, Wiesbaden

Emeneau, M. B., 1962, *Brahui and Dravidian Comparative Grammar*, Berkeley - Los Angeles

Farrell, T.,

1990. *Basic Balochi. An Introductory Course*, Naples.
2000, "Mother Tongue Education and the Health and Survival of the Balochi Language", in Jahani (ed.), 19-32

Ferraro, F., 1990, *Il lessico della parentela e dei gruppi di parentela in Baluchi*, PhD Diss., Napoli

Filippone, E.

2006, "The Body and the Landscape. Metaphorical Strategies in the Lexicon of the Iranian Languages", in *Proceedings of the 5th Conference of the Societas Iranologica Europæa*, ed. by A. Panaino & R. Zipoli, II: *Classical and Contemporary Iranian Studies*, Milano 2006, 365-389
2010, *The Fingers and their Names in the Iranian Languages*, Wien

Floyer, E. A., 1882, *Unexplored Baluchistan. A Survey ... of a Route through Mekran, Bashkurd, Persia, Kurdistan and Turkey*, London

Foxton, W., 1980, *Dictionary of English Baluchi and Omani Arabic*, s. 1.

Geiger, W.,

1889, "Balōčische Texte mit Übersetzung, [I]", *ZDMG*, 43:579-589
1890-91, "Etymologie des Balōčī", *ABAW*, 19:107-153
1891, "Lautlehre des Balōčī", *ABAW*, 19:399-464
1893, "Balōčische Texte mit Übersetzung, II", *ZDMG*, 47:440-449

Gilbertson, G. W., 1925, *English-Balochi Colloquial Dictionary*, Hertford

Gionta, P., 1982-83, "Working on the Baluchi Agricultural Lexicon", *Newsletter of Baluchistan Studies*, 1:67-71

Gladstone, C. E., 1873, *Biluchi Handbook*, Lahore

Grierson, G. A., 1921, "Balōchī", in *Linguistic Survey of India*, 10: *Eranian Family*, Calcutta, 327-451

Gumperz, J., 1964, "Hindi-Panjabi Code-Switching in Delhi", in Lunt, H. (ed.), *Proceedings of the Ninth International Congress of Linguists*, The Hague, 1115-1124

Hittu Ram, 1881, *Biluchi Nameh. A Text Book of the Biluchi Language*, Lahore

Horn, P., 1893, *Grundriss der neupersischen Etymologie*, Strassburg

Hübschmann, H., 1895, *Persische Studien*, Strassburg

Jaba A. and F. Justi, 1879, *Dictionnaire kurde-français*, St-Pétersbourg

Jahani, C.,

1989, *Standardization and Orthography in the Balochi Language*, Uppsala

2000 (ed.), *Language in Society. Eight Sociolinguistic Essays on Balochi*, Uppsala [if followed by page indication: Jahani, C., "Language Attitudes and Language Maintenance among Iranian Baloch in Sweden", in Jahani (ed.), 33-58]

Korn, A., 2005, *Towards a Historical Grammar of Balochi*, Wiesbaden

Leech, R.,

1838, "Epitome of the Grammars of the Brahuiky, the Balochky and the Panjābi Languages, with Vocabularies of the Baraky, the Pashi, the Laghmani, the Cashgai, the Teerhai, and the Deer Dialects", *JASB*, 7:608-620

Lewis, A.,

1885, *Bilochi Stories as Spoken by the Nomad Tribes of Sulaiman Hills*, Allahabad

Malkiel, Y.,

1976, *Etymological Dictionaries. A Tentative Typology*, Chicago - London

Marston, E. W.,
1877, *Grammar and Vocabulary of the Mekranee Beloochee Dialect*,
Bombay

Masson, C.,
1834, *Narrative of a Journey to Kalat*, London

Mayer, T. J. L.,
1900, *Baloch Classics*, 1, Fort Munro
1903, *Baloch Classics*, 2, Agra
1910, *English-Biluchi Dictionary*, Calcutta

Mayrhofer, M.
1956-80, *Kurzgefasstes etymologisches Wörterbuch des Altindischen*,
I-IV, Heidelberg
1980, *Zur Gestaltung des etymologischen Wörterbuches einer
„Großcorpus-sprache“*, Wien
1992-2001, *Etymologisches Wörterbuch des Altindoarischen*, I-III,
Heidelberg

Mengal, Mir Aqil K. K[hān],
1990, *A Persian-Pahlavi-Balochi Vocabulary*, I (A-C), Quetta

Mitha Khan and Ata Shad [Mit□hā Xān and ‘Atā Šād],
1972, *Urdū- Baločī luyāt*, Lahor

Mitha Khan and Surat Khan Marri [Mit□hā Xān and Sūrat Xān Marrī],
1970, *Baločī-Urdū luyāt*, Quetta

Mockler, E.,
1877, *A Grammar of the Baloochee as it is spoken in Makran*,
London

Morgenstierne, G.,
1927, *An Etymological Vocabulary of Pashto*, Oslo
1932, “Notes on Balochi Etymology”, *NTS*, 5:37-53
1946-48, “Balochi Miscellanea”, *AcOr*, 20:235-292
2003, *A New Etymological Vocabulary of Pashto*, Wiesbaden

Orywal, E.,

1982, *Die Balūč in Afghānisch-Sīstān: Wirtschaft und sozio-politische Organisation in Nīmrūz, SW-Afghanistan*, Berlin

Pierce, E.,

1874, "A Description of the Mekranee-Beloochee Dialect", *JBRAS*, 11:1-98

Rahman, T.,

1999, *Language, education and culture*, Karachi

V.S. Rastorgueva and D.I. Èdel'man,

2000, 2003, 2007, *Ètimologičeskij slovar' iranskix jazykov*, 1: Moskva 2000; 2: Moskva 2003; 3: Moskva 2007

Raverty, H. G.,

1860, *A Dictionary of the Puk'hto, Pus'hto or Language of the Afghāns*, London

Rossi, A. V.,

1979a, *Iranian Lexical Elements in Brāhūī*, Naples

1979b, "Phonemics in Balōčī and modern dialectology", in G. Gnoli and A. V. Rossi (eds.), *Iranica*, Napoli, 161-232

1982, "Balōčī Miscellanea", in *Monumentum Georg Morgenstierne*, II, Leiden, 157-183

1982-83, "Linguistic Inquiries in Baluchistan Towards Integrated Methodologies", *Newsletter of Baluchistan Studies*, 1:50-66

1986, "Il dizionario etimologico-comparativo della lingua baloci", in *Progetti lessicografici*, Centro di lessicografia asiatica dell'IsMEO, Roma, 159-190

1987, "Linguistica filologia e lessicografia storica iraniche", in *Atti del VII congresso internazionale di linguisti* (Milan, 12-14 september 1984), Brescia, 479-486

1998, "Ossetic and Balochi in V. I. Abaev's Slovar'", in *Studia iranica et alonica. Festschrift for Prof. Vasilij Ivanovic Abaev on the Occasion of His 95th Birthday*, Rome, 373-431

2000, *Etymological Vocabulary of Balochi*, *Encyclopaedia Iranica Online*, 2000, available at www.iranicaonline.org, 4 pp.

2004-2006, "Editorial Notes, 5", *Balochistan Studies*, 11-12:68-69

Rzehak, L. and B. Naruyi (eds.),

2007, Abdurrahman Pahwal: *Balochi Gâlband. Balochi, Pashto, Dari & English Dictionary. Completely revised and updated edition of an unpublished dictionary manuscript of the Balochi language*, Kabul

Shahbakhsh, A.

2004, *The Balochi Verb: An Etymological Study*, PhD London

Sinclair, J. M. (ed.),

1987, *Looking Up. An account of the COBUILD Project in lexical computing and the development of the Collins Cobuild English Language Dictionary*, London and Glasgow

Spooner, B.,

1967, "Notes on Baluchi spoken in Persian Baluchistan", *Iran*, 5:51-71

Steingass, F.,

1892, *A Comprehensive Persian-English Dictionary*, London

Tan, E.,

2000, "A Mother Tongue Literacy Programme among the Baloch of Singo Line, Karachi", in Jahani (ed.), 59-67

Tsabolov, R. L.,

2001, *Ėtimologičeskij slovar' kurdsogo jazyka*, 1: Moskva

Turner, R. L.,

1966, *A Comparative Dictionary of the Indo-Aryan Languages*, London

Zarubin, I. I.,

1932, *Beludžskie skazki*, 1, Leningrad

1949, *Beludžskie skazki*, 2, Moskva-Leningrad

1960, *Šugnanskije teksty i slovar'*, Moskva-Leningrad

Appendix I: Sample page from *An Index to Gilbertson, English-Baluchi Colloquial Dictionary*

A

abhá papa
abtar (*haptar*) hyaena
achá clean <to wipe> / *achá khanagh* to wipe
adab (*adáb*) courtesy, good manners, politeness; manners; respect <respectfully> / *go adab-a* respectfully; *adab khanagh* to respect; to show respect; see also *be-adab*, *be-adabi*
adagh [*aditha*] to spread <to spread a net>
adainagh (but *adagh* in G.) [*adaintha*] to pitch, as a tent
adá payment or discharge of a debt <to pay> / *adá khanagh* to pay, to repay
adáb see *adab*
adábíl swallow
ádíl just, honest
ad¹ (*ád*) watercourse
ad² (*adh*) ledge; support <to lean> / *ad-deagh* [*ad-dítha*] to lean; to rest, to place
adagh [*aditha*] to erect; to lay down, to spread; *jáíl adagh* to set a snare; to spread a net
adh see *ad²*
adhúr / *adhúr-e bacch* a child born before the proper time <born>
afi see *afi*
afí (*afí*) viper
afim opium
afsar officer, official (in the case of an European)
afshán leather bag used for holding flour <bag>
ag rate, price
agar (*aghar*) if / *agar ki* although; though
agá in front, before
agh again; by-and-by, in the future; if; then
agh-a again
aghadí (*agha dí*) again
aghar see *agar*
agharchi although; in spite of; whereas
aghdí (*agh dí*, *agh-dí*) again; more, still
aghl (*Caql*, *'aql*, *akul*) discernment, wisdom; intellect; mind; sense; understanding / [*khas-e*] *gár biagh* to lose one's wits; see *be-Caql*
aghlab probably
aghm see *gham*
aghmá effort; endeavour / *aghmá khanagh* to strive, to make an effort
aghsar most likely, most probably
ahad see *cahd*
ahar ash <banyan-tree>
ahár the hot season; June-July; summer / *ahár-a* in the summer
cahd (*ahá*, *ahdh*, *cahdh*, *ahad*) agreement; alliance; covenant; warranty
cahdh-bandokh wire-puller, intriguer

ahd see *cahd*
ahdh see *cahd*
cahdh see *cahd*
caiv see *caiv*
caiv (*Caib*, *'aiv*) fault; flaw, blemish; vice, defect / *caiv janagh* to vituperate; *'aiv-a khanagh* to judge; [*-chakhá*] *caiv lá'imagh* to condemn
cajab (*'ajab*) curious; extraordinary; odd; queer, strange; unique; wonderful; wondrous
ajám settlement, arrangement
cajib quaint
ajnábi stranger
akastha see s.v. *caksagh*
caqd wedlock, marriage-knot, marriage-tie
akistha see s.v. *caksagh*
caql see *aghl*
akrot walnut
akrot drashk walnut tree <walnut>
caksagh (*'aksagh*) [*'akastha*, *'akastha*, *'akistha*] to repose, to rest, to sleep / *'akisthiy-e* he is asleep, in a sleeping condition, lying down <asleep>; *'akistagh* (*'akistagh*) sleeping
aksar-a generally, as a rule
akási # *akási tašwir* photograph <to repose>
akul see *aghl*
al-amán mercy on us! <mercy>
'aláhida see *'aláhida*
'aláhida (*'aláhida*) separately / *aláhida khanagh* to separate
camál doings / *camal giragh* to gain experience
albat indeed, yes, truly, undoubtedly, assuredly; see also *albatta*
albatta certainly <timid>; see also *albat*
alláh God
almás diamond
al-gharaz in short; see also *gharaz*
al-qádir God, the Almighty; see also *qádir*
amánat care, anything left in the safe-keeping of another, charge, deposit, something left in charge (of men or things), trust; custody; guardianship; keeping / *amánat deagh* to entrust, to intrust (in the case of lifeless objects)
amb mango
ambár see *hambár*
ambázi (*am-bázi*) embracing <to embrace> / *am-bázi khanagh* to clasp to one's bosom
gwar-ambázi embrace <to embrace> / *gwar-ambázi khanagh* to embrace
ambráh attendant; fellow, associate; servant, companion / *nokh-e ambráh* recruit
ambrá'i see *ambráhi*
ambráhi (*ambrá'i*) service / *ambráhi khanagh* (*ambrá'i kh^e*) to associate with; to help; to serve; to support
ambur pliers; see also *zambur*

Adriano V. Rossi, " L'Orientale" University of Naples

The Etymological-Comparative Dictionary of the Balochi Language: Progress Report of an International Project

Balochi Academy Golden Jubilee International Conference on Balochi Language and Culture, Quetta 31 July - 1 August 2011

Appendix 2: Sample article from *The Etymological-Comparative Dictionary of the Balochi Language*

gwāt wind; air; flatulence | **gwāt** id.; adj.
swollen CoBal | **gwāθ**, **gwās** id. EBal

	Term of wide diffusion in the whole Indo-Iranian area. Attested in Iranian since Old Avestan with the meaning ‘wind’ (but already in recent Av. ‘the Wind God’), and continued with this core meaning in most Iran. languages; in further modern Iranian languages its core meaning is associated to the concept of ‘arrogance’
	<ul style="list-style-type: none"> • Secondary meanings ‘weather’ CoBal, <i>āf-gwāθ</i> climate, EBal <i>gwātō</i> windy EBal (D 108), <i>gwātū</i>, <i>gwātōk</i> fan, <i>gwātī</i> exorcist, a person who tries to help people who have jinns, exorcism (gen. by a woman, <i>gwātīyen mās</i> witch, Gloss.) nom. comp. <i>gwātgir</i> chimney RaBal (fortress, Gloss.?), <i>gwāt-sarī</i> pride; proud RaBal, CoBal, cf. NPrs <i>bād-sar</i> vain, <i>bād-sarī</i> vanity; <i>gvāθ-šalwār</i> having fluttering shalwars, EBal, adj. transl. <i>gvātšalwār</i> proud, boastful, RaBal.; <i>gwāt-rōf</i> A 39^b name of a plant, with which NPrs. <i>bādrō</i> is to be compared, according to Vu, I. 162^a: herba quaedam foliis basilico similis et odore mali citrei; <i>sar-gwāt</i> windward CoBal many vb. comp., e.g. <i>gwāt deag</i> to cause to ache, swell; <i>gwāt kaššag</i> to swell (tr. and intr.); <i>gwāt gīrag</i> to swell; to get angry; <i>gwāt warag</i> to take air (all in Farrell).
	In the Iranian cultures the wind is often conceived as an extra-human agent intervening in everyday life, and therefore many words derived from this term have magic connotations.
Comparative section →	<ul style="list-style-type: none"> • OAv. <i>vāta-</i> wind, YAv. also ‘the Wind God’, MPers. <i>wād</i> id., breath (also pl. <i>wādān</i> ugly spirits), ManPth. <i>w’d</i> wind (also spirit), BudSogd. <i>w’t</i> id. (also spirit), <i>w’tō’r</i> living being, Bactr. <i>oađo</i> ‘the Wind God’ (Davary 240), Khot. <i>bāta-</i> wind; Pers. <i>bād</i> wind, air, breath, haughtiness (all meanings attested since <i>Shāhnāme</i>), Psht. <i>wo</i> wind (both WIr. and EIr. apparently < *<i>wāta-</i>)
Etymological section →	<ul style="list-style-type: none"> • EWA 2.542 (cf. Ved. <i>vāta-</i> wind, Wind God; certainly I-Ir., also in the div. mg. ‘the Wind God’ (OIr. PN *<i>vāta-frac{d}{a}ta-</i> ‘granted by the Wind God’, NPrs. <i>Wātfrac{d}{a}t</i>, Schmitt, IrNKleinas. IV, 27) and conn. to OIr. <i>Vāyu-</i> ‘the Storm God’, cf.

	Chyet 19a), CDIAL 11491, certainly IE (cf. LIV 287 <i>h₂ueh₁</i> ‘to wind’)
Bibliographical section →	<ul style="list-style-type: none">• GE148. In this example on historical phonology, Bal. differs from both Parthian and Middle Persian. The reflexes of OIr. *v- depend to some extent on the vowel which follows. In this case *va- > gwa-/gu- [Elfenbein 1989: 353-4].
	Cf. Mokri, <i>Les vents du Kurdistan</i> , 1970, 233-258.