



BĀLAYBALAN LANGUAGE

BĀLAYBALAN LANGUAGE (or Bālaïbalan, Bāl-a i-Balan, Bāleybelen), an *a priori* constructed language combining elements of the grammar of Arabic, Persian, and Turkish, and known principally from a single text, a dictionary. It is the earliest attested constructed language, and one of a very few that are not of European origin. On the basis of its heterodox components and apparent sectarian usage, C. Colpe (p. 47) has compared it to Sant Bhāṣā, the northwest Indian *Mischsprache* in which the Guru Granth Sahib, the central religious text of Sikhism, was composed. Unlike Sant Bhāṣā, which is comprehensible to Hindi, Urdu, and Punjabi readers, Bālaybalan is completely incomprehensible to speakers of Arabic, Persian, and Turkish.

Origin of the language and history of scholarship. Although the dictionary gives no indication of its authorship or origins, a tentative hypothesis was first advanced by [Edgard Blochet](#), who attributed it to Sufis of the [Bektāšiya](#) order, who preserved the esoteric doctrines of the Ḥorufis (see [Ḥorufism](#)) concerning language (Blochet, p. 246). This hypothesis was subsequently adopted by nearly all of the scholars who discussed the language. [Alessandro Bausani](#) further proposed [Faẓlallāh Astarābādi](#), the founder of the Ḥorufi movement who died at the end of the 14th century, as a candidate for its inventor (Bausani, 1954, p. 236).

This hypothesis remained unchallenged until 1966, when Midhat Sertoğlu published a short notice about the jurist and poet [Moḥyi Moḥammad Golšani](#), who in 1580 composed a dictionary and grammar, *Loḡa wa qawā'ed-e bālaybalan*. In this work, an edition of which was published in 2005 by



Mustafa Koç, Moḫyi Moḫammad identifies the language as his own creation (Sertoğlu, p. 67). [Tahsin Yazıcı](#) notes that another work of the same author, a quadrilingual dictionary entitled *Maṣāder-e alsena-ye arba'a*, contains a column in Bālaybalan in addition to those in Arabic, Persian, and Turkish. Both manuscripts are to be found in the Egyptian National Library and Archives (see [Moḫyi Moḫammad Golšani](#)).

Regardless of who initially invented the language, its dictionary indicates that the development of its vocabulary had been a collective effort for years prior to its eventual publication in manuscript form (fol. 6v, lines 5-15).

Description of the dictionary. The dictionary, *Ketāb aṣl al-maqāṣed wa faṣl al-marāṣed* or (in Bālaybalan) *Dātayvakšā vahātaybakšā* “The Origin of Goals and the Division of Observations,” is represented by at least two manuscript copies, one found in the Bibliothèque nationale de France, and the other in the Manuscript Collection of the Princeton University Library. The French orientalist Louis Jacques Rousseau reported to Silvestre de Sacy that he had consulted yet another copy of the same work in an unspecified Baghdad library in 1805 (de Sacy, p. 365).

The Paris manuscript is a quarto manuscript containing 334 folios, on 331 of which is written the text (332 according to de Sacy, but the number 258 was omitted when the manuscript was paginated). It was originally accessioned as Persan 188 (de Sacy, p. 365), subsequently renumbered Supplément persan 1030 (Blochet, p. 246), and eventually restored to its original designation (Richard, 1989). It is a composite of different fragments, evidently written at different times and by different hands (de Sacy, p. 396). The text on the majority of the manuscript (fols. 1v–69v, 98r–255v) is neatly written and framed by borders of two red lines and black and gold lines, respectively. The remaining portions (fols. 70r–97v, 256r ff.) appear to be a work of restoration by a later copyist, as they appear in a different hand and are not framed with a border, like the remainder of the book.

The other copy of this manuscript, found in the Manuscripts Collection of the Princeton University Library, Islamic Manuscripts Third Series no. 265, bears a date in the second of the two preliminary sections of this work, which indicates that it was completed in 988 A.H. (1580-81 CE).

The text, which is written in a very high register of Ottoman Turkish (Peter Golden, personal communication), is divided into two parts, of which the first



(fols. 2–70), subdivided into six parts, is dedicated to the verbs of the language, arranged according to the infinitive form or *dāt*, and the second (fols. 72–332) is dedicated to its substantives.

Example of entry. *Sanam* öğmek ki sepâsiden ve sütuden ve medh ve hamddir, *nasam* ve *savam* ma'nâsına. *Nâs* ve *sâm* ve *sânah* ve *savah* öğüş ki sepâstır. *Nanas* ve *nasam* ve *nasav* 'ahmeddir. *Nasaş* ve *sanaş* ve *savaş* maħmud ve memduħdur. *Nanneş* ve *sanneş* teşdid nunlu ve *savnaş* muħammeddir müta'adiden. *Nasan* ve *sanam* ve *savan* hâmed ve mâdehtir. *Gensnam* kendi-yi öğmek ki hodrâ sütudendir *geni sanam*"dan muhaffeftir ve sevmek ki *hamam*"dı beyân oldu.

Translation. “[The verb] *sanam* is [Turkish] *öğmek* [“to praise,” modern Turkish *övmek*], which is [Persian] *sepâsīdan* and *sotudan*, and [Arabic] *madh* and *hamd*, with the same meaning as [the verbs] *nasam* and *savam*. *Nâs*, *sâm*, *sânah*, and *savah* are “praise,” [Turkish] *öğüş* [modern Turkish *övinç*] which is [Persian] *sepâs*. *Nanas*, *nasam*, and *nasav* are [Arabic] *aħmad* [an elative form meaning “most praiseworthy”]. *Nasaş*, *sanaş*, and *savaş* are [Arabic] *maħmud* and *mamduħ* [passive participles meaning “praised”]. *Nanneş* and *sanneş*, with doubling in the /n/, and *savnaş* are [Arabic] *moħammad* [a passive participle meaning “intensively praised”] from the transitive [form II]. *Nasan*, *sanam*, and *savan* are [Arabic] *hâmed* and *mâdeh* [active participles meaning “praising”]. *Gensnam* is [Turkish] *kendiyi öğmek* [“to praise oneself,” modern Turkish *övmek* or *kendi kendini övmek*], which is [Persian] *k'od-râ sotudan*, contracted from *geni sanam*, and [Turkish] *sevmek* [“to love”] which was indicated in [the entry for the verb] *hamam*.”

Description of the language. As a dictionary, *Dātayvakšā vaħātaybakšā* does not include any verbal paradigms or indeed any substantial text samples apart from a short bilingual (Arabic and Bālaybalan) passage that opens the volume. Nonetheless, it is possible to glean some information about the language from the entries and the sample.

Phonology. The writing system of Bālaybalan reflects a system of six discrete vowels, comprising three tense (or possibly long) vowels, <ā>, , and , which contrast with three lax (or possibly short), , and , represented by the pointings fatha, kesra, and zamma, respectively, and several diphthongs, including , , and (two additional diphthongs, and , which are graphically indistinguishable from the vowels and). The manuscript offers no direct indication of the pronunciation of these vowels, but the presence of diphthongs such as in the word gewzā



“sources,” which would be impermissible according to the rules of Arabic phonology, suggests that vocalic repertoire is more akin to that of Persian or Turkish.

The writing system of Bālaybalan distinguishes between 33 consonants, employing the full set of 32 letters of the Perso-Arabic script as well as the so-called Indian *gāf* (Bālaybalan *gi*). These include the homophone letters *t* and *ṭ* (both of which are pronounced /t/ in Persian and Turkish), *ṣ*, *s*, and *ṣ̣* (all pronounced /s/), *ḥ* and *h* (both pronounced /h/), and *z*, *ẓ* and *ẓ̣* (all pronounced /z/). Again, the manuscript offers no direct indication as to how these letters were intended to be pronounced. While these homophone letters are largely an artifact of the wholesale incorporation of Arabic vocabulary into these other two languages, the role that they play within Bālaybalan is less easily explained, as its lexicon is largely original, and there are therefore few, if any, genuine historical or etymological spellings to be met. Either the letters are to be pronounced as in Persian or Turkish (and were incorporated purely for their numerological significance), or they are to be pronounced as in Arabic (and each therefore represents a distinct phoneme).

Bālaybalan syllables fall into five basic patterns: CV ‘a “and,” CVV *fā.jā* “as a light,” CVC *jah* “upon,” CVVC *ra.yān* “to Allāh,” and CVCC *gens.nam*, “to praise oneself.” A CVVCC syllable might be predicted, but no examples of this pattern were attested. Consonants alone may serve as the onset of a syllable, but no syllable may begin with a cluster of two or more consonants. If a prefix is appended to a syllable beginning with the glottal stop <’>, the stop is elided; e.g., *y- def + ’ān* “god” becomes *yān* “God” rather than the expected ***y’ān*. In all other instances, an anaptyctic intervenes to break up potential clusters, particularly in word-initial position; e.g., *y- def + ’ašanā* “intermediaries” becomes *ya’šanā* “the intermediaries,” and *r- “to, for” + y- def + karfanā* “well-disposed ones” becomes *raykarfanā* “for the well-disposed ones.”

As indicated by the former examples, a series of two open syllables containing lax (or short) vowels are regularly consolidated into closed syllables whenever prefixes or suffixes are added. As initial consonant clusters are not tolerated, the deleted vowel is always the former nucleus of the second syllable, the onset of which becomes the coda of the preceding syllable, e.g. *y- def + sa.nam* “praise” becomes *yas.nam* “the praise” instead of the expected ***ya.sa.nam*, and *ma.kan* “lord” + *-ad* 1pl becomes *maknad* “our lord” instead of the expected ***ma.ka.nad*. The vowels <ā>, , and are never deleted in this context, and the rule does not appear to operate following a clitic or series of clitics, e.g.,



'a=fa.jas “and he rose” instead of the expected **'af.jas, or 'a=ja=maq.ri “upon his family (maqar)” rather than the expected **'aj.maq.ri.

Morphology. Bālaybalan shares with Arabic and other Semitic languages a “root and pattern” system of morphology, at least with regard to basic lexical stems. The Bālaybalan “root” consists of a set of consonants (here called “radicals”), which are arranged in a specific sequence. This sequence of consonants imparts the general meaning of the word; any additional information, such as the part of speech it represents, is reflected by the pattern of vowels, e.g., *ad* “one” but *ād* “single,” *taf* “cry!” but *tāf* “crying,” and *babam* “to look,” but *bubam* “to look carefully, watch attentively.” Even though polysyllabic words are common due to agglutination and compounding, no basic stem contains more than three radicals or consists of more than two syllables (fol. 72v, line 14: *yohsa bâleybelen'di üç harfdan ziyâde aşl kelime bulunmaz*).

While the root and pattern system expresses most of the intrinsic semantics of a given word, most of its grammatical information is indicated by preposed and postposed inflectional morphemes. Bālaybalan appears to be “synthetic” in every sense of the word, being not only manmade rather than naturally evolved, but also characterized by an agglutinative morphology, which has frequently been compared to Turkish (e.g. Colpe, p. 47). Much like Persian and Turkish, but unlike Arabic, Bālaybalan dispenses completely with grammatical gender.

Noun phrase. As noted above, the noun can be inflected to indicate its role within the sentence, number, and definiteness. With regard to the last, subjects are unmarked, and objects are marked by a number of prepositions. The direct object of a transitive verb is marked with the preposition *r-*, corresponding to the Persian postposition *-rā*. Indirect objects are indicated with prepositions such as *b-* “in, with, through”; *f-* “from”; *m-* “like”; and once again *r-*, which in this context can mean “to, for.”

When a noun is modified by another noun (substantive or adjective) in a genitive or attributive relationship, this relationship is indicated by the suffixed morpheme *-(v)a*, which follows the head noun directly, e.g., *şān-a yān* “the name of God,” and *gewzā-va ynaşā* “the origins of things.” As can be seen from the preceding example, this morpheme assumes the form *-va* after a vowel. By means of this morpheme, “chains” of three or more nouns can be built, e.g., *dāt-a jām-a ynanşanā 'a yakşanā* “the source of all things elementary



and derived.”

Number is indicated through the absence or presence of a postposed plural morpheme. Singular nouns are unmarked, e.g., *far* “head,” but plurals are indicated with the inflectional morpheme *-ā*, e.g., *farā* “heads.”

Finally, nouns with a definite referent (one that is assumed to be known to the audience) or a generic referent (in which the reference is to all members of a class) are marked with the prefix *y-*, while indefinite nouns (those that do not have a specific referent) are unmarked. Nouns are inflected to indicate definiteness regardless of the role they play in the sentence, with one exception. As in Arabic, the noun governing a genitive relationship is never explicitly marked as definite, even if it is semantically definite, e.g., *Bāl-a ybalan* “the language of the life giver,” rather than ***Yabāl-a ybalan*.

Pronouns. Bālaybalan personal pronouns are marked for singular and plural number, and first, second, and third persons. While no independent personal pronouns are attested in the brief sample at the beginning of the manuscript, there are several suffixed possessive/objective forms of the pronoun, e.g., *-(b)i* “his, him”; *-ad* “our, us”; and *-yā* “their, them,” which are appended directly to nouns and prepositions. Bālaybalan has a single relative pronoun *čonā* “that,” which is inflected for number when its referent is plural (hence *čonāyā*).

Verb phrase. The finite forms of the verb distinguish between three tenses (past, simple future, and continuous present), two voices (active and passive), and an imperative. There are also non-finite forms, including an active participle, a passive participle, and a verbal noun or *ḍāt*, which serves as the lemma for each of the entries in the first part of the book.

All forms of the verb are derived from the stem of the verb through the addition of a series of prefixes and suffixes. On its own, the bare stem yields the imperative, which is inflected for singular or plural with the same inflectional morpheme *-ā* that is found on the nouns, e.g., *bar* “know (sg.)!”, *barā* “know (pl.)!” The *ḍāt* is created by adding the suffix *-m*. Thus the stem *bar* becomes the verbal noun *baram* “to know.” An adjective may also be derived from a verbal stem with the suffix *-ān*, e.g., *barān* “knowledgeable.” Other derived forms include the active participle, which is formed with the suffix *-n*, e.g., *kašan* “extending (sg.)”, *kašnā* “extending (pl.)”, and the passive participle, which is formed with the suffix *-š*, e.g., *baraš* “known (sg.)” and *barsā* “known (pl.)”. These participles are often used substantively, in contexts which



occasionally demand that they be paraphrased as present tense forms, e.g., *mafñā* “admirers; those who praise.”

The stem of the past tense is built with the suffix *-as*, e.g., *baras* “he/she knew.” This may be further inflected for person, e.g., *barasā* “they knew.” The simple present/simple future is likewise formed with the suffix *-ar*, e.g., *barar* “he knows; he will know,” and the continuous present is formed from this by means of the prefix *ma-*, e.g., *mabarar* “he knows.” Additionally, the suffix *-r* also serves as a copula, e.g., *’ādir* “it is a single (letter).” See Table 1 for a complete paradigm of finite forms. These forms may be negated with the prefix *la-*.

The valence of a verb may also likewise be increased or decreased through the familiar process of affixing a series of morphemes to its stem. An intransitive verb, like *vazam* “to begin,” can be made transitive with the inclusion of the infix *-n-*, becoming *vaznam* “to introduce (something)”; a transitive verb such as *sanam* “to praise” can be intensified, becoming *sannam* “to praise highly.” The opposite process is achieved through the infix *-z-*, which converts an originally transitive verb such as *balam* “to revive” into an intransitive one, *balzam* “to live.” Transitive verbs can also be made reflexive with the morpheme *geni* “-self,” e.g., *geni sanam* “to praise oneself,” which contracts to *gensnam*.

Syntax. In many respects, the syntax of Bālaybalan resembles that of Arabic rather than of Persian or Turkish. Unlike Persian or Turkish, Bālaybalan regularly marks both the number and definite or indefinite status of the noun in all syntactic roles. The sole exception, as noted above, is the head noun of a genitive construction, which is not marked for definiteness, as in Arabic (e.g., *aşl al-maqāşed* “the origin of goals,” rather than ***al-aşl al-maqāşed*). The relative pronoun *čonā* agrees in number with its antecedent, as in Arabic, but unlike relative pronouns in these other languages. Furthermore, Bālaybalan exclusively employs prepositions, rather than postpositions like Persian *-rā* and the varied Turkish postpositions, and much like Arabic (but unlike Persian or Turkish) the verb regularly precedes its predicate rather than following it. In this regard, the language can be described as strongly “right-branching” or head-initial.

Even so, the source languages do share some marked similarities with one another, which Bālaybalan therefore shares with them. For example, all three languages are pro-drop: the subject of a verb is indicated through inflection



rather than being explicitly indicated with an independent pronoun, e.g., *fajas fa-mim-a ymafñā* “it ascended (*fajas*) from the mouth (*mim*) of those who praise (*mafñā*).” Likewise, the language makes extensive use of suffixing, particularly in the inflection of the verb, as do all three of the source languages.

Lexicon. While at first glance, the vocabulary of Bālaybalan appears to have been generated *ex nihilo*, de Sacy (pp. 368-69) identified several common words and particles that bear an apparent resemblance to words in the source languages, such as *b-* “in” (Arabic *bi-*), *dāt* “origin” (Arabic *dāt* “essence”), *jam* “all” (Arabic *jam*), and *r-* “to” (the Arabic preposition *li-* and, serendipitously, the Persian postposition *-rā*). Bausani (1974, p. 93) points out that many other words, which do not bear any immediate resemblance to their equivalents in the source languages, nonetheless parallel them in metaphorical ways that may not be obvious without familiarity with the Sufi poetic tradition. For example, the aforementioned word *mim* “mouth” recalls the Arabic letter *mim*, which is poetically compared to a small and graceful mouth; the word *pir* “mirror” recalls Persian *pir*, a Sufi master and metaphorical “mirror” of the disciple. Like that of a natural language, the vocabulary of Bālaybalan is filled with unexpected parallels, subtle nuances of meaning, and a wealth of opportunities for wordplay.

BIBLIOGRAPHY

A. Bausani, “About a Curious “Mystical” Language BĀL-A I-BALAN,” *East and West* 4/4, 1954, pp. 234-38.

Idem, *Geheim- und Universalsprachen: Entwicklung und Typologie*, Stuttgart, 1970.

Idem, *Le lingue inventate : linguaggi artificiali, linguaggi segreti, linguaggi universali*, Rome, 1974.

E. Blochet, *Catalogue des manuscrits persans de la Bibliothèque nationale II*, Paris, 1912.



C. Colpe, “The Phenomenon of Syncretism and the Impact of Islam,” in K. Kehl-Bodrogi et al., eds., *Syncretistic Religious Communities in the Middle East: Collected Papers of the Symposium, Berlin 1995*, Leiden, 1997.

E. Drezen, *Historio de la Mondolingvo*, Moscow, 1991.

T. Jung, *De Muheddin ĝis Mundilatin: Mundlingvaj projektoj tra la jarcentoj*, Purmerend, 1937.

M. Koç, *Bāleybelen Muhyî-i Gülşenî: ilk yapma dil*, Istanbul, 2005.

A. Okrent, *In the Land of Invented Languages: Adventures in Linguistic Creativity, Madness, and Genius*, New York, 2010.

F. Richard, *Catalogue des manuscrits persans I. Ancien fonds*, Paris, 1989.

S. de Sacy, “Kitab asl al-maqasid wa fasl al marasid: Le capital des objets recherchés et le chapitre des choses attendues, ou Dictionnaire de l’idiome Balaïbalan,” *Notices et extraits des manuscrits de la bibliothèque Impériale et autres bibliothèques* 9, 1813, pp. 365-96.

M. Sertoğlu, “İlk Milletlerarası Dili Bir Türk İcat Etmışti,” *Hayat Tarih Mecmuası* 1, 1966, pp. 66-68.