The SARVA (South Asia Residual Vocabulary Assemblage) Project¹

F. C. Southworth, University of Pennsylvania (Emeritus)
(last updated 21 September 2005)

1. Goals and background. Etymological research in South Asian languages has turned up large numbers of words in languages of all families which are of unknown origin, many of which are probably remnants of otherwise unknown ancient languages. In addition, there are language remnants (such as Nahali/Nihali) and isolated languages (such as Burushaski) which may possibly belong to formerly more widespread language families. The goal of the Project is to assemble all words showing early language contact among the (known and unknown) languages of the subcontinent, in order to provide data for the reconstruction of the history of language contact in the region, from the time of the earliest knowable South Asian linguistic strata, including inferences regarding the locations of these strata in time and space.

Though the SARVA Dictionary takes the form of an etymological dictionary, the Project is not concerned only with the origins of individual words. Our ultimate concern is the discovery of patterns of linguistic interaction that will lead to reconstructions of the times, places, and cultural circumstances under which prehistoric language contact took place in the subcontinent.

Recent work in archaeobotany by Dorian Fuller (e.g. Fuller 2002, 2003) has made clear the possibility of relating the history of plants to linguistic history, including correlations between the domestication and spread of cultivated plants and the linguistic origins and distribution of plant names. There is good reason to hope that in the not too distant future it will be possible to relate the history of Fuller five early agricultural regions to specific linguistic groups, and to get a much more detailed understanding of the interactions of those groups from at least the 3rd millennium BCE onward.

The SARVA Project is directed by Michael Witzel of Harvard University and Franklin C. Southworth of the University of Pennsylvania (Emeritus). The Project activity is centered on the SARVA website, which is in the process of creating an online, ongoing comparative etymological dictionary of the language families of South Asia. The Project was first discussed at the Fourth Harvard Round Table on the Ethnogenesis of South and Central Asia in May 2002. In March of 2003, the Research Institute for Languages and Cultures of Asia and Africa (ILCAA) at the Tokyo University of Foreign Studies (TUFS) agreed to host the website and provide technical and administrative support. The website went on line in June of 2005, and is still under development.

2. Design of the Project. The project consists of two related parts: the SARVA online Database and the SARVA Dictionary. The database consists of a series of dictionaries and word-lists, which constitute the data from which entries in the SARVA Dictionary are being constructed. The components of the Database (see list below) are, or will be, searchable on a number of categories, including aspects of phonetic shape, grammatical features, semantic categories, and English glosses. While the Database is primarily designed as an aid for the compilers of the Dictionary, it is also available to other scholars.

Partial list of Database items:

¹ Based on a paper presented at SALA 25, UIUC, Champaign-Urbana, IL, September 2005
3. Structure of entries in the SARVA Dictionary. SARVA Dictionary entries are composite entries containing material from two or more language families, along with comments by the compilers and others who wish to contribute. Each entry will be signed and dated by its compiler, and all comments will also be dated and signed with their authors’ initials. All entries are subject to discussion and emendation for an indefinite period of time, as there are no plans to publish the dictionary in the immediate future.

4. Current state of the Project. At present the SARVA website is partially in operation, though still under development. The home page can be viewed at www.aa.tufs.ac.jp/sarva/, and the active windows can be entered directly at www.aa.tufs.ac.jp/sarva/entrance/html. To view the SARVA Dictionary (upper right window) it is necessary to have a Unicode (UTF-8) font installed, and to set your browser for Unicode.

In addition to a few sample entries in various stages of development, the Dictionary window contains the entire list of starred words from Turner’s Indo-Aryan dictionary. Some of these are words of non-Aryan or unknown origin, such as the following:

179 *a — 2 ‘obstruct, stop’. [Cf. *a ha—2, *a —, *a —: prob. —
Drav. see *a —] S. a aru ‘obstinate’; L. a m. ‘sediment in a well,
silt’; Or. a āḍī ‘enclosure for cattle’; G. āḍ m. ‘pounding’ (?), a v āṇa
f. ‘obstacle’; M. a f. ‘obstruction’, āḍ f. ‘obstinacy’, a e ‘to thicken,
inspissate’. (Turner, CDIAL, p. 10)

while others are words which may have existed in OIA but are unattested, such as
25 *ak aḷā— ... m. ‘lynch—pin’ lex. [āk a—1, kīla—] Bi. (Patna)

akhailā ‘axle of brick—crusher’. (Turner, CDIAL, p. 2)

(This is obviously a word that could have occurred in Sanskrit, but does not happen to be attested.) The latter type of entries will be deleted as time permits. The numbering system of the CDIAL entries will provide a framework for numbering the new entries.

As noted, SARVA entries will be initialed and dated by the concerned compiler. Comments of other scholars will also be initialed and dated. We invite interested and knowledgeable colleagues to participate in this process. For the moment, we ask you to send comments to us by email (see below). If non-standard characters are involved, the email version of Skt. transcription can be used², or the comment can be sent by .pdf.

The format of SARVA entries is not yet fixed, and suggestions are welcome. This is very much a dictionary-in-progress, and we look to our scholarly colleagues to provide assistance from their own knowledge and expertise.

Addresses for comments and suggestions: fsouth@mail.sas.upenn.edu, witzel@fas.harvard.edu. Questions regarding Munda may be sent to stampe@hawaii.edu.

REFERENCES


-----------. 1962. Review of Pinnow 1959. JAOS 82.10.107-9

Fuller, Dorian Q. 2003. n agricultural perspective on Dravidian historical linguistics: archaeological crop packages, livestock and Dravidian crop vocabulary Ch. 16 (191-213) in Peter Bellwood & Colin Renfrew (eds.), Examining the farming/language dispersal hypothesis. Cambridge: McDonald Institute for Archaeological Research.


² a A i I u U R R e a i o au aM aH k kh g gh G c ch j jh J T Th D Dh N t th d dh n y r l v $ S s h (Skt alphabet)
Masica, C. 1979. ryan and non-Aryan elements in North Indian agriculture in Deshpande & Hook (eds.), *Aryan and non-Aryan in India*. Ann Arbor: University of Michigan Center for South and Southeast Asian studies 55-152


Pfeiffer, Martin. 1972. *Elements of Ku u x historical phonology*. Leiden


------------------------. 1979.


Witzel, Michael. 1999a. ubstrate languages in Old Indo-Aryan *EJVS* 5.1.1-67

------------------------. 1999b. arly sources for South Asian substrate languages *Mother Tongue* (extra number) 1-70


**Abbreviations:**

AG = Arlo Griffiths

BhK = Bh. Krishnamurti

CD = Central Dravidian (v. BhK2003)

CN = attested in CD & ND

CDIAL = Turner 1966

DED = Burrow & Emeneau 1984

FCS = Franklin C. Southworth

MW = Michael Witzel

ND = North Dravidian (v. BhK2003)

SD = South Dravidian (v. BhK2003)

S1 = South Dravidian I (v. BhK2003)

S2 = South Dravidian II (v. BhK2003)