

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's 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's 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

- (1) Turner's *Comparative Dictionary of the Indo-Aryan languages* (Oxford 1966)
 - (2) Burrow & Emeneau, *A Dravidian etymological dictionary*, revised edition (Oxford 1984)
- (These two items are available by permission of Oxford University Press and the Digital South Asia Library, University of Chicago.)

Additional items to be added as time permits include:

- (3) Kuiper's famous list of 'foreign' words in RV (Kuiper 1991, see Witzel 1999b);
 - (4) Nahali words of unidentified origin (Kuiper 1962, Mundlay 1996);
 - (5) words listed in DEDR with etyma limited to a single branch of Dravidian;
 - (6) Munda materials from David Stampe's website www.ling.hawaii.edu/Stampe/AA/ETYM/Stampe&Munda, plus other materials in preparation at the University of Hawai'i;
 - (7) Sino-Tibetan materials from James Matisoff's Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) Project (www.stedt.berkeley.edu/Matisoff/) when they become available;
 - (8) Materials from M. Witzel's articles (e.g. 1999a-c);
 - (9) Masica's list of agricultural words from "language X" (Masica 1979);
 - (10) Words of unknown origin found in OIA and Dravidian from Southworth 1979 & 2005a;
 - (11) Indo-Iranian words from Lubotsky 2001 (see also Witzel 2001, 1999b §5);
 - (12) words of unknown origin from Marathi (Kulkarni 1964), Khandeshi (Chitnis 1964), Katkari (Kulkarni 1969), to be contributed by FCS;
 - (13) Vedda words from De Silva 1972;
 - (14) Residual words in the Dravidian Nilgiri languages (Emeneau 1997, v. Zvelebil 1990:63-72, Witzel 1999b §2.4);
 - (14) Oriya vocabulary (AG);
 - (15) materials from Kusunda and Bangani languages (MW and colleagues),
 - (16) Comments on selected etyma from Mayrhofer 1986.
- Other data will be added as time permits. (Suggestions are welcome.)

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.tufts.ac.jp/sarva/, and the active windows can be entered directly at www.aa.tufts.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 *aṭṭ— 2 'obstruct, stop'. [Cf. *aṭṭha—2, *aḍ—, *aḍḍ—: prob. ← Drav. see *aḍ—] S. aṭaru 'obstinate'; L. aṭṭ m. 'sediment in a well, silt'; Or. aṭāi 'enclosure for cattle'; G. āṭṭ m. 'pounding' (?), aṭvāman f. 'obstacle'; M. aṭ f. 'obstruction', āṭ f. 'obstinacy', aṭṇē 'to thicken, inspissate'. (Turner, *CDIAL*, p. 10)

while others are words which may have existed in OIA but are unattested, such as

25 ***akṣakīla**— ... m. ‘lynch—pin’ lex. [ákṣa—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

- Burrow, T., and M. B. Emeneau. 1984. *Dravidian etymological dictionary*. Oxford: Clarendon Press
- Chitnis, Vijaya. 1964. *The Khandeshi dialect*. Pune: Deccan College doctoral dissertation
- De Silva, M. W. S. 1972. *Vedda language of Ceylon: texts and lexicon*. München: R. Kitzinger
- Emeneau, M. B. 1997. “Linguistics and botany in the Nilgiris”. in P. Hockings (ed.), *Blue Mountains revisited: cultural studies on the Nilgiri Hills*. New Delhi: Oxford University Press
- . 1962. Review of Pinnow 1959. *JAOS* 82.10.107-9
- Fuller, Dorian Q. 2003. “An 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.
- . 2002. “Fifty years of archaeobotanical studies in India: laying a solid foundation”, pp. 247-363 in S. Settar & R. Korisettar (eds.), *Indian archaeology in retrospect, Vol. III: Archaeology and interactive disciplines*. Delhi: Manohar
- Krishnamurti, Bhadriraju. 2003. *The Dravidian languages*. Cambridge University Press
- Kuiper, F. B. J. 1948. Proto-Munda words in Sanskrit. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij
- . 1962. *Nahali: a comparative study*. Amsterdam: N.V. Noord-Hollandsche Uitgevers Maatschappij
- . 1991. *Aryans in the Rigveda*. Amsterdam-Atlanta: Rodopi
- Kulkarni, K. P. 1964. **मराठी व्युत्पत्ति कोश** [*Marathi etymological dictionary*] (second edition). G. L. Thokal, Thokal Bhavan, Lakshmi Road, Pune 2
- Kulkarni, S. B. 1969. *Descriptive analysis of Kāṭkarī dialect*. Pune: Deccan College doctoral dissertation

² a A i I u U R RR e ai 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)

- Lubotsky, A. 2001. "The Indo-Iranian substratum" in Carpelan, Parpola, & Koskikallio (eds.), *Early contacts between Uralic and Indo-European*. Helsinki: Suomalais-Ugrilainen Seura 301-17
- Masica, C. 1979. "Aryan 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
- Mayrhofer, M. 1986. *Etymologisches Wörterbuch des Altindoarischen*. Heidelberg: Carl Winter, Universitätsverlag
- Mundlay, Asha. 1996. "Nihali lexicon", *Mother Tongue* 2.17-40.
- Pfeiffer, Martin. 1972. *Elements of Kurux historical phonology*. Leiden
- Pinnow, Heinz-Jurgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Otto Harassowitz
- Southworth, F. C. 2005a. *Linguistic archaeology of South Asia*. London: Curzon Press.
- 2005b. "Prehistoric implications of the Dravidian element in the NIA lexicon with special reference to Marathi", *IJDL* 34.1:17-28.
- , 1979.
- Turner, Sir R. L. 1966. *A comparative dictionary of the Indo-Aryan languages*. London: Oxford University Press.
- Witzel, Michael. 1999a. "Substrate languages in Old Indo-Aryan", *EJVS* 5.1.1-67
- 1999b. "Early sources for South Asian substrate languages", *Mother Tongue* (extra number) 1-70
- 1999c. "Aryan and non-Aryan names in Vedic India", in Bronkhorst & Deshpande (eds.), *Aryan and non-Aryan in South Asia*. Cambridge (MA): Harvard Oriental Series, Opera Minora vol. 3.337-404
- Zvelebil, Kamil V. 1990. *Dravidian linguistics: an introduction*. Pondicherry: Pondicherry Institute of Linguistics and Culture

Abbreviations:

AG = Arlo Griffiths	FCS = Franklin C. Southworth
BhK = Bh. Krishnamurti	MW = Michael Witzel
CD = Central Dravidian (v. BhK2003)	ND = North Dravidian (v. BhK2003)
CN = attested in CD & ND	SD = South Dravidian (v. BhK2003)
CDIAL = Turner 1966	S1 = South Dravidian I (v. BhK2003)
DEDR = Burrow & Emeneau 1984	S2 = South Dravidian II (v. BhK2003)