Sanskrit Studies as a Foundation for Computational Linguistics

Nicholas Ostler
Linguacubun Ltd / Foundation for Endangered Languages
Bailbrook Lane, Bath BA1 7AA, England
+44-1225-852865 fax -859258
nostler@chibcha.demon.co.uk

LESAL Workshop, Mumbai 23-25 April 2001
Indo-UK Workshop on Language Engineering for South Asian Languages

I. A Sacred Language?

Sanskrit is principally known outside India as the sacred language of Hinduism. However, one effect of this sacred status has been the long-term development of linguistic science in India, on a rigorous empirical basis. In fact, the attitude to Sanskrit as sacred has been the solid foundation and justification for its position as the leading focus of Indian studies of language for three millennia.

These studies have ranged over the full gamut of the scientific study of language, and have for the most part been preserved up to the present day. This paper offers an overview of aspects of these studies, in order to indicate their relevance to current computational approaches to language processing, and in particular corpus linguistics. Sanskrit linguistics is in an excellent position to make immediate use of most modern techniques in language processing, since it is already provided with most of the infrastructural tools which are currently seen as desirable.

Although these studies have been thorough and systematic, they did not detract from the Indian sense of language as a mystical presence. The earliest text of Indian literature, the *R•gveda*, is a collection of hymns, and one is devoted specifically to the nature of language. It emphasizes its role in contacts between friends, but also the esoteric nature of progress in language analysis. This would seem to offer an auspicious point of departure for new ventures between Indian and UK linguists.

R•gveda X.71.

br•haspate prathamam• va¤co agre	When, O Lord of the World, the Wise established
yatprairata namadheyam• dadha¤na¤h•	Name-giving, the first principle of language,
yades•a¤m• çres•t•ham• yadariprama¤sî¤t	That which was excellent in them, that which was pure,
pren•a¤ tades•a¤m• nihitam• guha¤vih•	Hidden deep within, through love was brought to light.
	|1|
saktumiva titauna¤ punanto	When the Wise created language with the mind
yatra dhî¤ra¤ manasa¤ va¤camakrata	As if winnowing barley with a sieve,
atra¤ sakha¤yah• sakhya¤ni ja¤nate	Friends acknowledged the quality of friendship;
bhadra¤is•a¤m• laks•mî¤rnihita¤dhi va¤ci	Upon their speech was impressed the mark of grace.
	|2|…
uta tvah• paçyanna dadarça va¤cam•	Many aman who sees does not see the Word
uta tvah• çr•nvanna çr•n•otyena¤m•	And many a man who hears does not hear it.

uto tvasma¤i tanvam• 'vi sasre ja¤ya¤	Yet for another it reveals itself like
iva patya uçati suva¤sa¤h•	A radiant bride yielding to her husband.

|4|…

II. Characterizing Sanskrit

A. The Evident Differences

It is evident that the context in which ancient Indian grammatical analysis was undertaken was very different from the modern situation of computational linguistics.

1. Anti-business?

The predominant modern fashion is to emphasize the potential benefits of automatic language processing as part of an increasing automated society, where information technology is an asset to wealth creation, and is undertaken quite deliberately with material benefits in mind. Ancient Indian societies were no strangers to explicit theories of wealth creation and economics, a specialism known as *arthaça¤stra*. But linguistic analysis was never focused in this direction. Rather, it was seen as ancillary to correct use of language for spiritual purposes. And as the *R•gveda* shows from time to time, the motives of business and the professions were viewed with some scepticism as a guide to higher understanding.

R•gveda IX.112	
na¤na¤nam• va¤ u no dhiyo	Our thoughts wander in all directions;
vi vrata¤ni jana¤na¤m•	many are the ways of men.
taks•ta¤ ris•t•am• rutam• bhis•ag	The cartwright hopes for accident, the physician injury
brahma¤ sunvantamicchati	The Brahman a rich patron.
indra¤yendo pari srava	For the sake of Sprit, O Mind, let go these thoughts! |1|

2. Anti-writing

Another aspect of the context of ancient Indian linguistic studies which will be alien to modern attitudes is the pronounced doubts about the value of book-learning. Reliance on language in its written form was seen as crippling, and not giving true control over linguistic content. Hence this proverb:

pustakastha¤ tu ya¤ vidya¤ parahastagatam dhanam	Knowledge in a book [is like] money in another's hand. (*Ojha, Bha¤ratî¤ya Pra¤cî¤na Lipi Ma¤la¤,14, n.6, attrib.to Ca¤nakya-nî¤ti*)

In this ancient India was like many cultures as widely divided as Gaul in the first century BC (attested as a Druid view in Julius Caesar's *De Bello Gallico* 6.14), and modern

Guatemala (where Mayans remark [as noted by Martin Prechtel, p.c.] that outsiders note things down not to much to remember them, as not to have to remember them).

Even though the language had undergone a full phonological analysis by the fifth century BC, which was even incorporated into the official order of letters in the alphabet, reliance on written texts for important (especially spiritually important) documents. Hence another saying:

vedavikrayin•açca¤iva veda¤na¤m•ca¤iva du¤s•akah• veda¤na¤m• lekhakaçça¤iva teva¤i nirayaga¤minah•

The sellers of the Vedas,
the misreaders of the Vedas,
the writers of the Vedas,
all go on the path to hell.
(*Maha¤bha¤rata* - quoted by B.S. Kesavan: *The Book in India*, Nat. Book Trust 1992: 3)

By contrast the ideal was the rote learning of all the principal texts, through judicious use of mnemonic techniques. This learning then made possible true engagement with all aspects of the texts, including the composition of new texts and commentaries, which might indeed benefit from being written down.

B. Curious Precursors

It is curious that in this very different milieu, which looked askance at economic application of studies, and for which writing was kept in as ancillary role as possible, all the appurtenances of modern corpus linguistics nevertheless came into being.

1. The Sanskrit Corpus

The major corpus of Sanskrit linguistics is the Vedas themselves, above all the *R•gveda*, a compilation of 1028 anonymous hymns, preserved from second millennium BC. Although most of the hymns are dedicated to particular deities, it is not a not purely liturgical collection, and may be considered to some extent a balanced corpus, containing some apparently secular items, for example RV X.34 The Lament of the Gambler, enslaved to *babhravah•*, "the browns", i.e the particular nuts then used as dice, and quoting his desperate pleas to his creditors:

ra¤ja¤ cid ebhyo nama it kr•n•••oti...
tasmai kr•n••omi, 'na dhana¤ run•adhmi' daça¤ham
pra¤ci¤s, 'tad r•tam vada¤mi'.

even a king bows before them…
I open my palms to him: '
I am not holding out on you—it's the truth, I tell you.'"

The language is also rather unrestrained

R•gveda X.34
açvo vol•ha¤ sukham• ratham•
hasana¤m upamantrin•ah•
çepo roman•vanta¤u bheda¤u
va¤rin man•du¤ka icchati

The horse would draw a swift carriage;
The entertainer a good laugh.
The penis seeks a hairy slit
And the frog a stagnant pond.

The *R•gveda*, however, is just the beginning of the Vedas. It is succeeded by further compilations, the *Sa¤maveda* and *Yajurveda* (which are more purely liturgical in their content); the Atharvaveda (which seems less elevated in its theorology, and contains many spells, which seem to bespeak a more animist faith). Accompanying hese are the *Bra¤hman•as* (which are a commentary in prose on the Vedas);and then the *A¤ran•yaka*s,("forest books") and *Upan•is•ad*s, which are also in prose and develop a more mystical philosophy.

2. Sanskrit and Writing

We have already noted the exceedingly sceptical attitude of Indian tradition to written sources. However, the topic of written sources for Indian languages is an exceedingly rich one, not least because there was florid development of different regional alphabets in over two millennia of recorded Indian writing. As a result, there are now 22 modern scripts derived from the Brahmi script which was first attested in the emperor Açoka's edicts of mid 3rd cent. BC. (There is one other which seems to stems from another tradition, probably an independent development from Aramaic: this is *Kharos•t•hi,* the 'she-mule' script used in Afghanistan and the northerly mountains in the centuries around the beginning of the first millennium AD.)

Not all of these are used for Indian languages, but there are are recognizable regional styles, probably owing their distinct characters to the kind of writing implements that were available: 6 of them are used for modern *A¤ryan* languages in the north of India and Nepal, as well as Tibetan further north in the Himalayas; 5 are used in south India for Dravidian languages, including one for the isolated Aryan language *Sin•ha¤la* in Çri Lanka; 4 in Indochina; 5 more in Indonesia, as well as the Tagalog and Buginese scripts used in the Philippines. (A useful collation of a substantial part of the scripts can be found at http://www.geocities.com/Athens/Academy/9594/brahmi.html .)

There are two paradoxes which are inseparable from any consideration of Sanskrit's own written tradition.

One is that, unlike all other Indian (and South-East Asian) languages, only Sanskrit is indifferent to the script used. Although there is a apredominant tradition for Sanskrit to be represented in nowadays in the Devanagari script, used for Hindi and a number of other north Indian languages, all the above scripts are on occasion used to write Sanskrit. This makes prayers etc. in Sanskrit accessible to users of various vernaculars. But it also serves the linguist as a convenient demonstration (if any were needed) that all the alphabets are phonologically isomorphic, having the same basic structure and organization that was formulated for them by Indian grammarians over two millennia ago.

The second paradox is that Sanskrit's written tradition begins much later than that of its 'daughter' languages, which are usually called the Prakrits, and from which the modern Aryan languages are descended. The first writing in the Indian subcontinent which has survived to the present is the corpus edicts issued by the emperor Açoka, who ruled the

vast majority of the subcontinent from Patna, (in the area then known as Magadha), in the mid 3rd century BC. These are written in Magadhi Prakrit. It is not until two hundred years later that the first writing is Sanskrit is found, further west, in Ayodhya and Mathura. There is already a clear division of function between Sanskrit and Prakrit visible in these inscriptions which contain both: Sanskrit is used for the verse, Prakrit for the prose dedications.

Sanskrit however came to predominate, indeed to become the exclusive language of inscriptions. This tradition begins 250 years later, in 150 AD with the rock inscriptions of *Juna¤gad•h* on the eastern coast, in Gujarat.

3. Programming in Sanskrit

Another aspect of modern computational linguistics that has a precursor in the Sanskrit tradition is the programming of linguistic analysis by synthesis in the form of explicit formal rules. This is not a metaphor, or anachronistic interpetation of Sanskrit grammar, but a straightforward description of the working of *su¤tras* in the system of *Pa¤n•ini's As•t•a¤dhya¤yî¤* (Panini is believed to have lived in the 5th century BC, *probably in th academic community of Taks•açila¤,* near modern Rawalpindi in the extreme north-east of the subcontinent.). This grammar, when combined with the *Dha¤tupa¤t˘•ha* (or list of roots) contains explicit rules adequate to characterize the full phonology and inflexional morphology of Sanskrit, and its application in sentence grammar (e.g. including the case and participant-role of all the noun-phrases dependent on the main verb). To show how this kind of programming could work in the absence of a computer, consider the application of a single *su¤tra:*

iko yan• aci

(*As•t•a¤dhya¤yî¤* VI.1.77)

The three words that constitute the *su¤tra* are not words of Sanskrit itself, but of a technical metalanguage that refers tersely to other *su¤tras* of the grammar. It is as if they are consonant-stem nouns, with the regular ending for genitive (-as), nominative (zero) and locative (-i). (There is a slight complication, in that both a voiced segment, a final -as is realized phonetically as -o. This is a regular principle of Sanskrit sandhi.) The *su¤tra* could *therefore be analysed functionally as*

$$[ik]_{GEN}\,[yan•]_{NOM}\,[ac]_{LOC}$$

In the context of a *su¤tra*, these cases have special interpretation, referring respectively to the input, the output and the right-hand context of a phonological rule. The *su¤tra is therefore to be understood as :*

$$[ik] \rightarrow [yan•] \,/\, \_[ac]$$

But what is the reference of the strange words themselves? They are to be understood as applications of another set of *su¤tras* (*known as the* Çiva-*su¤tras*), which plays the role of a system for defining natural classes of Sanskrit phones. This begins:

a i u N¤; r• l• K; e o N¤; a¤i a¤u a¤ C; h y w r T•; l N• ...

There is no distinction of upper or lower case in Sanskrit, nor any semi-colons. But the use of this Roman typographical convenience is simply to show explicitly what a student of Paninian grammar learns by example, namely that the phones here written in upper-case are functioning as control characters. Any term consisting of one of the lower-case letters L followed by one of the control characters M denotes the sequence of phones starting with L and ending just before M. So for example, "*aC*" denotes the set of vowels, "*haT•*" the set of semi-vowels excluding l. *It can be seen then that the su¤tra* being analysed is nothing less than a concise statement of the rule:

$$\langle i,u,r\bullet,l\bullet\rangle \longrightarrow \langle y,w,r,l\rangle \text{ before } \{ a,i,u,r\bullet,l\bullet,e,o,\ a\text{¤}i,\ a\text{¤}u \}$$

Terse, indeed, but it should be remembered that this level of concision is only possible because a number of controlling principles can be taken for granted — e.g. the interpretation implicit in the brackets: the first four phones map respectively onto second four phones, but this occurs before any of the nine phones in the environment. Part of the task of the tradition of commentary which followed on from Panini was to make explicit the precise nature of the *paribha¤s•a¤* (auxiliary principles) on which the correct interpretation of the *su¤tras* rests.

4. Lexicon and Thesaurus

Besides a full grammar, the Sanskrit tradition can also boast a strong dictionary, or rather a thesaurus, since traditional lexica were organized semantically, rather than phonetically. Perhaps the best example of this is the versified thesaurus called *Amarakoça* (by *Amarasim•ha*), believed to go back to the classical period of the sixth century AD. It is a work organized by synonyms, arranged in three books by subjects, and ends with an appendix that gathers up homonyms, indeclinables and genders. The whole work is intended, like Panini's *su¤tras*, for rote memorization. However, Sanskrit thesauri (*kos•as*) are limited in containing only nouns and indeclinables. For verb roots, one needs to refer to lists like the *Dha¤tupa¤t˘•ha* already mentioned.

Besides these fruits of the indigenous tradition, there are extended bilingual dictionaries with Western languages, notably German (e.g. the monumental Böhtlingk & Roth from the nineteenth century) and English (e.g. V.S. *Apte's Practical Sanskrit-English Dictionary*, re-edited in three volumes by P.K. Gode and C.G. Karve (1957-9). The first bilingual dictionary appears to have been the Persian-Sanskrit *Pa¤rasî¤¤praka¤ça* ("Light on Persian") in the late 16th century.

5. Sociolinguistics

India has always presented special interest as a linguistic area with significant relations among languages in contemporary use. The precise status of Sanskrit has long benn a

matter of controversy, vis-à-vis both the more archaic language of the Vedas, and the vernacular languages, often closely related to Sanskrit, called Prakrits (*pra¤kr•ta* - natural).

Although the earliest inscriptions, from the 4th century BC as already mentioned, are in Magadhi Prakrit, later writings such as the in early Mathura inscriptions (1st cent. BC) show a more complex picture; the inscription on the well at Mora, for example, shows prose in Prakrit, but verse in Sanskrit. Later, in the 6th century AD, it becomes a convention of Sanskrit drama to observe a certain diglossia or more exactly polyglossia: noble males speak in Sanskrit; ladies speak in Çauraseni (the Mathura Prakrit), but sing in *Maha¤ra¤s•t•rî¤;* meanwhile, low characters are scripted in *Ma¤gadhî¤* (from the lower Ganges area).

Interestingly, this last is likely to have been the dialect which Gautama Buddha, though about one millennium earlier. Although from the Buddha's time Buddhism was keen to adopt a vernancular language in which to preach, in practice it soon came to adopted *Pa¤li* (meaning 'canonical') as its standard. This was a mixed Prakrit. Later on, as the faith developed, it increasingly adopted a grander style of language, morphologically closer to classical Sanskrit, which is known as Buddhist Hybrid Sanskrit.

Although Sanskrit, therefore, has always been in contact with other languages more or less closely related to it, it cannot be said that the indigenous tradition has adapted itself to this more varied subject matter. These different dialectal form are well documented, but they have not been subjected Paninian-style analysis.

6. Phonetics

Phonetics has always been emphasized as part of the Sanskrit linguistic tradition: this priority was after all likely to be one implication of the low esteem for the written form already noted. This had various implications on the content and style of Sanskrit culture and literature.

First of all, there was an elaborated articulatory theory of the various Sanskrit phones (resumed at length in W.S. Allen's *Phonetics in Ancient India,* Oxford 1953). This was widely enough known among non-linguists to be given a theological correlate: so Chandogya Upanishad 2.22.3, while offering a few technical hints on their articulation, states that the vowels, sibilants and *sparça* (stops and semivowels) are the very self of **Indra, Praja¤pati** and Death respectively.

Secondly, the articulatory theory and its attendant classification of all the phones was influential enough early enough to impose itself on the conventional order of the Indian alphabets. If one orders the sequencing principles:

1. vowel before consonant;
2. simple before complex;
3. place of articulation from back to front;
4. voiceless before voiced;

5. surd before nasal;
6. *sparça* before fricatives

and applies them to the letters of the alphabet (all of which are used quite phonetically), the standard order is generated (except perhaps for the order of the fricatives at the end):

a a¤ i î¤¤ u u¤ r• r¤• l• e ai o au k kh g gh ~ c ch j jh µ t• t•h d• d•h n• t th d dh n p ph b bh m y r l v ç s• s h

Thirdly, the developed theory of speech sounds, which as we have noted were assigned theological correlates formed an articulated background for the practice of the mantra, which literally means 'pondering'. The term was used for Vedic hymns used as incantations, and later for certain sacred formulae whose incessant repetition was held to have important mystical effects. The best known are probably OM, the cosmic syllable, and the "Six Syllables" (*s•ad•aks•ara¤)* of Tantric Buddhism, *Om man•i padme hu¤m* —"Hail the jewel is in the lotus."

Fourthly, this emphasis on the power of spoken formulae was applied systematically in Indian pedagogy, in the theory of the *su¤tra, or* 'thread'. Whereas Western didactic texts have been formulated axiomatically (after Euclid), or more often as didactic verse, the preferred approach in the Sanskrit tradition has been to encapsulate treatises as a series of memorable aphorisms, usually phrased as verse couplets or çlokas. An example of this is the following, which might be called a "meta-*su¤tra*", since while being a *su¤tra* itself, it states the essential qualities of a *su¤tra*:

svalpa¤ks•aram asandigdham• sa¤ravad viçvatomukham
astobham anavadyam• ca su¤tram• su¤travido viduh•

brief, unambiguous, pithy, universal, non-superfluous and faultless the sutra known to the sutra-sages.

Finally, a literary effect of the heightened consciousness among Sanskrit-users of the actual speech which mediated expression was the luxuriant growth in their literature of what was called *çles•a* 'word-play' (literally 'adhesion'). Sanskrit, having borrowed freely in prehistoric times from the unrelated Dravidian languages of India, always had a vast vocabulary. Among other things, this resulted in a profusion of homonymy, one lexeme often appearing to have a variety of unrelated meanings. Works such as the Amarakoça already mentioned helped writers to take advantage of this in their poetry. But a special characteristic of Sanskrit is a complicated system of word liaison, known as *sandhi*. This means that word-boundaries are often effaced, and a single stream of syllables, as pronounced or even written, becomes susceptible of multiple interpreta-tions. The combined result of these two properties of Sanskrit is an opportunity for punning on an almost inconceivable scale. This opportunity was amply taken up in literary composition. The ultimate in this was acieved by the poet *Kavira¤ja* ("poet-king") who in his *Ra¤ghavapa¤n•d•avî¤ya* (12th cent. AD), set himself the task of re-telling simultaneously the stories of both the great epics of India, the *Ra¤ma¤yan•a* and *Maha¤bha¤rata*, in ambiguous (and highly ornate) verses. In a way, this can be seen a release of meaning from its expression in words, for it is difficult to conceive how the work could have been understood, in either of its senses, without active and detailed pre-knowledge by the audience of the tales being told. Author and audience share the stories, but are focused exclusively on the verbal details of their expression. This in practice forces not only the use of ambiguous terms, but

an analogy to be set up between the narrative flow of the two epics. So, to quote one couplet (VI.8):

paracakram• parikra¤mann açokagahanam• gatah•:
ks•ana¤d iva kr•ta¤rtho 'bhu¤n maheyî¤darçanena sah•.

Going round the enemy's **kingdom** / *forces,* he came to **a thicket of Açoka trees** / *the reverse of grief*: in an instant as it were, his task was accomplished, by his sight of **the daughter of the earth** / *the cows.*

Here the first of the variant translations (in bold) of phrases applies to Hanuman seeking for Sita, and the second (italicized) to Arjuna on a cattle-rustling expedition behind enemy lines. But to maintain a coherent narrative, most of the phrases still have an unambiguous translation.

It is clear then that the peculiar Sanskrit consciousness of the phonic medium of language served to complicate the scope for analysis of language in use, as well as to make the language, in principle, subject to more explicit and concrete analysis.

C. Sanskrit and Other Languages

1. South-east Asia in First Millennium AD

Besides the Sanskrit tradition of language analysis as reviewed above, there is another aspect which will be an important background to future computational studies of the Sanskrit language. Sanskrit is a language that has spread far beyond India itself, and the perceived effects of this spread paint an encouraging picture of what could be the place of Sanskrit analysis in the future of computational language analysis in India itself.

There is copious evidence that Sanskrit proliferated, along with Indian cultural norms, in South-east Asia and the islands of Indonesia, especially during the first millennium and a half AD. This area was known as *Suvarn•abhu¤mi* ('Gold Land') and *Suvarn•advî¤pa* ('Gold Island'), although actual evidence for gold mining is fairly slight. Nevertheless, names of rulers are typically Sanskritic (e.g. more than thirty Cambodian kings whose names end in "varman" from *Jayava¤rman* who died in 514 AD to Çrî¤ndra¤jayava¤rman 1307-27, andthe Majapahit kings of Indonesia from *Ra¤jasa* in 1222-7 to *Suhita¤* 1429-47. Many of the place-names, including the most important, are derived from Sanskrit, or are Sanskritized versions of local names: so Java, Sumatra, Cambodia, Malaya, and the Irawaddy are derived from *Yava-dvî¤pa* "barley-island"; *samudra* "ocean"; *Kamboja* — a kingdom in the Khyber Pass area; *Malaya*, actually from a Dravidian word *malai* "a hill", in south India near Malabar; and *Ira¤vatî¤* "having drinking water", the old name of Ravi river in Panjab.

The inscriptions from this period are either all in Sanskrit, often in literary verse which evidences familiarity with classics, or (in Java) bilingual with Javanese. (This is all amply attested in R C Majumdar *Study of Sanskrit in South-East Asia*. Calcutta: Sanskrit College, 1974.) Furthermore, at the level of popular culture, Javanese puppet theatre known as *wa¤ya¤ng* is based for its characters and plot lines on the *Maha¤bha¤rata*.

This is a bizarre state of affairs, taking into account widespread Hindu prohibitions on crossing the sea, and against consorting with *mlecchas* 'barbarians'. Nevertheless, it is clear that, as a result of from some combination of trade, adventuring or missionary activity (both by Hinduism and Buddhism), this cultural and linguistic spread did occur. Nevertheless, it is very different from many other spreads overseas of developed cultures to less sophisticated and well-organized cultures, not least the concurrent spread of Chinese influence in the reverse direction, from China in the north into Vietnam. As G. Coedès writes (*The Indianized States of Southeast Asia*, Univ. Hawaii: 1964, pp. 34-5):

> [by contrast with China''s influence in the north] Indian penetration or infiltration seems to have almost always been peaceful; ... Far from being destroyed by the conquerors, the native peoples of Southeast Asia found in Indian society, transplanted and modified, a framework in which their own society could be integrated and developed... It is this that explains the differentiation, and in a certain measure, the originality of the Khmer, Cham, and Javanese civilizations, in spite of their common Indian origin.

2. The vehicle of Buddhism (along with *Pa¤li*)

Besides this influence on South-east Asia which seems to have been a transit that involved most aspects of the culture, the new religion of Buddhism, in various forms, powered other spreads of the Sanskrit language which were much more focused on its sacred use. In this way, it came to spread not only to South-east Asia but also, very early, to Çrî¤ Lan•ka (in the fourth century BC), and in a reverse direction to Kashmir and Afghanistan. A little later, in the 1st century AD, missionaries took it (via the Karakorum and the Pamir mountains) to China. Thence it was ultimately transmitted along with the Chinese approach to Buddhism, to Korea and Japan, its most easterly homes, where it arrived at the end of the 6th century. Surprisingly, it was only a thousand years later than the mission to China that it moved due north of its original home in Magadh and southern Nepal into Tibet, ca 1200. Then the final area to be exposed to Buddhism (and hence sacred Sanskrit) on a large scale was Mongolia, its northernmost home, proselytized from China in the middle of the second millennium.

This kind of Sanskrit was very different from the classic norm. In many cases, the language used was not Sanskrit at all, but its descendant *Pa¤li*; and where it was (e.g. as the Buddhist Hybrid Sanskrit referredto above), it was specialized for use in sermons, prayers and mantras. Gone were the literary allusions, complex verb inflexions and elaborate sandhi. In its place came Sanskrit in a much less involved style.

A good example would be the Heart Sutra, a common Buddhist text, in the form in which it was written (in Chinese characters) by Hsuan-tsang in a cave at Ta-hsing-shan-ssu near Lo-yang (on the Silk Road), where he passed on his long and stupendously famous individual pilgrimage from China to the Buddhist centres of India in the in 7th century AD. Its last lines read:

gate gate paragate parasam•gate gone, gone, totally gone, totally completely gone;

Sanskrit, then, has a far-flung history, and has been in contact with cultures conducted in other languages all over southern and eastern Asia. Nowhere has this linguistic contact led to loss or replacement of other linguistic traditions, though Sanskrit has always been central to new cultural developments wherever it has reached. This is an enviable record (when contrasted with the usually devastating impact of such other languages of large-scale proselytizing civilizations as Greek, Latin, Arabic, Spanish, French and English).

III. Practical Implications

To wrap up, we can consider the potential role of Sanskrit in the future electronic notation, analysis and transmission of languages world-wide.

It cannot be denied that Sanskrit, in its past and present sociolinguistic roles in India, has certain disadvatages in representing itself as a *lingua franca*, or even as a model for development to its close neighbours. There is a tradition of Brahman arrogance in India, inseparable from the caste system, with which Sanskrit is bound to be associated, as the vehicle of the civilization of which the Brahmans have been the ex officio guardians. In recent years, Sanskrit has also become associated with some strains of Hindu nationalism, emphasizing the language's north Indian roots rather than its pan-Asian ramifications. It is clear that this attempt to fence Sanskrit in, with some understanding of its traditional home culture, is very alien to the character of Sanskrit's actual career over the past three thousand years.

Furthermore, the style in which Sanskrit has been taught, at least since the time of Panini, has not been one to encourage original, and possibly radical, thought: rather it has emphasized mastery of the received wisdom, together with all the interlocking arguments of its vast corpus of rationale. In this, it is very similar in spirit to the merits of the European mediæval tradition of scholasticism, but without a Renaissance to fertilize it, and send it off in new directions.

There is also evident competition in India both from Hindi and from English, as de facto and de jure languages of pan-Indian communication in the modern world. Nevertheless, Hindi has not had the benefit of 2,500 years of linguistic analysis on which to found its computer development. And English, despite its feverish development over the past 250 years, can never offer the well-established cultural links with languages all over India that are inalienable from Sanskrit.

The overall implication of this paper is that Sanskrit has all the necessary components of a good foundation for a pan-Indian enterprise of computational linguistic analysis. It will be possible to let computers take some of the strain which Sanskrit pedagogy has traditionally laid on human brains, in working out the correct application of vast batteries of formalized rules, and organizing the vast lexicon into coherent sets of synonyms, antonyms and hyponyms, as well as recognizing patterns of semantic relationships perhaps not yet

dreamed of. It should also be possible, where computer reason fails, for the programmer and the pundit to share their expertise to mutual, and indeed general, advantage. For the first time, perhaps, it will be possible for Sanskritists to offer their expertise as a model for the development of the linguistic analysis of their neighbouring languages in India, and ideally much further afield.

It is good to end with some heartening words from the Vedas:

R•gveda X.71.

yastitya¤ja sacividam• sakha¤yam•	He who forsakes a companion in knowledge
na tasya va¤cya¤pi bha¤go asti	Has no way left open of sharing the word.
yadî¤m• çrn•otyalakam• çrn•oti	Indeed whatever he hears, he hears in vain;
nahi praveda sukr•tasya pantha¤m•	He knows nothing of the path of right action.
	|6| …
sarve nandati yaçasa¤gatena	All the friends rejoice for their glorious friend
sabha¤sa¤hena sakhya¤ sakha¤yah•	At the end of his journey, reaching fulfilment,
kilbis•aspr•tpitus•an•irhyes•a¤m	For he brings nourishment, and removes their guilt,
aram• hito bhavati va¤jana¤ya	And he is prepared to act courageously.
	|10| …