

Setting Priorities for South Asian Languages

Draft report prepared by David Gilmartin, Dwight Stephens and John Caldwell

North Carolina State University

February 11, 2004

I. Aims. As part of the eLCTL initiative, led by Michigan State University, and funded by the Department of Education, we have been asked to prioritize the languages of South Asia with respect to their teaching in the United States. This is part of a larger project to encourage national planning and coordination in the teaching of Less Commonly Taught Languages in the United States. We have focused our assessment on the languages of South Asia, in which, for this purpose, we have included India, Pakistan, Bangladesh, Afghanistan, Nepal, Sri Lanka, Bhutan, and the Maldives. We have included in our assessment some languages which are widely spoken both within South Asia and outside it as well (such as Persian, Uzbek, Tibetan), and we have prioritized these with respect to their importance to South Asia as a region.

II. Criteria for ranking. We have proposed to use three basic criteria for ranking.

A. First, and most significant, is the number of speakers of each language.

Rough figures for numbers of speakers of each language are included on the attached list. We had hoped to be able to use the latest figures from the 2001 Indian census for language speakers in India, but those figures, unfortunately, have not yet been released (though they are expected shortly). We have therefore used figures from the 1991 census, modified by more recent estimates. In some cases these figures are *very* rough estimates of the numbers of speakers. Figures for Pakistan, which has not had a recent census, are particularly problematic. Since dates of estimates differ, these numbers are not strictly comparable.

Generally, we have put all languages that have more than 5 million speakers in the highest priority category. Languages with more than 1 million speakers we have put in the second priority category. Languages with 200,000 speakers or more are in the third category (for India, mostly based on a reporting of 200,000 or more speakers under the 1991 census). However, these categorizations have in some cases been modified as a result of other criteria.

B. Second, we have assessed the importance of languages based on their official status and on their political importance.

In some cases, officially recognized status is straightforward, but in other cases it is not. Status as a national language carries weight. This has been used, for example, to raise Dzonghka (the national language of Bhutan) and Divehi (the national language of the Maldives) from category 3 (where numbers of speakers would have put them) to category 2.

There are problems, however, with using official status in India as a straightforward guide to the importance of a language. The Indian constitution provides for the recognition and listing of official (or “scheduled”) languages (as listed in the eighth schedule of the constitution). In some cases official recognition reflects specific political accommodations made by the Government of India with regional ethno-

linguistic movements. Recognition also hinges sometimes on what is considered a “language” and what is considered to be a “dialect” of another language. The 1991 Indian census tried to finesse this problem by distinguishing between “languages” (both “scheduled” and “non-scheduled”) and “mother tongues,” which were grouped under the headings of larger languages. Political controversies surrounding these issues, however, remain intense and widespread. These controversies are particularly heated with respect to the status of the various regional forms of Hindi. But they are hardly confined to Hindi-related issues. In Pakistan, for example, the relationship of Saraiki to Panjabi and Sindhi is a subject of considerable political controversy.

Initially, fourteen languages were scheduled as “official” under the Indian constitution. We have included all of these in the first category. Subsequently, Sindhi was added, and, in the 1990s, Konkani, Manipuri, and Nepali. Most recently (in December 2003), the government has added four more languages, Bodo, Dogri, Maithili, and Santali, to the list, bringing the total to 22. It is likely that in the future more will be added, which makes it difficult to use this as a simple guide to prioritizing languages. We have included Sindhi, Nepali, and Santali in the first category, but due to their more limited numbers of speakers, have left most of the rest in category 2.

The question of Maithili raises clearly the problem of how to treat the languages that are sometimes treated as separate languages and at other times as regional variants of Hindi. Based on its recent official status and numbers of speakers, Maithili certainly deserves to be in category 1. But it is difficult, in many ways, to distinguish the claims of Maithili from other eastern Hindi languages, such as Bhojpuri or Magahi, which also have large numbers of speakers. Interestingly, none of these were recorded as separate languages under the 1991 census, but were all classified under Hindi simply as “mother tongues.” An easy answer to this problem is therefore elusive, but we have decided to lump these together under the loose label, Bihari, as is sometimes done, and to include this in category 1, rather than listing each of these as separate languages. In some ways, this parallels the situation with Rajasthani (also in category 1), a cover term used for a number of western Hindi variants spoken in or near Rajasthan. The languages (or “mother tongues”) comprising Bihari and Rajasthani are thus (mostly) listed in the appendix (“List of South Asian Languages”), but they are not treated as separate entities in the priority language list.

Some other languages with relatively large numbers of speakers, but which are often (and in politically-contested ways) treated as regional variants of larger languages, have been listed as separate languages in the language priority list, but have been given a lower priority than would be warranted on the basis of numbers alone. These languages include Saraiki and Hindko in Pakistan (Panjabi), Haryanvi, Awadhi, and Chhattisgarhi in India (Hindi), and Sylheti and Chittagonian in Bangladesh (Bengali).

C. Third, we have considered the historical and academic interest of languages. Languages with few speakers but with large literatures and great historical importance in South Asia (such as Sanskrit, Pali, and Prakrit) have thus been raised in priority. Sanskrit and Pali are in category 1, and Prakrit is in category 2. Some languages which have lower numbers of speakers, but high linguistic interest, such as Burushaski, have been added to category 3.

We have recognized the importance of issues relating to language teaching in the US for prioritizing languages, such as numbers of heritage speakers, or importance to American foreign policy interests, but these factors have not in fact shifted in any case priorities established on the basis of the three criteria.

III.. Prioritized List of Languages

Group One

Assamese
Baluchi
Bengali (Bangla)
Bihari (including Bhojpuri, Magahi, Maithili)
Gujarati
Hindi
Kannada
Kashmiri
Malayalam
Marathi
Nepali
Oriya
Pali
Panjabi
Pashto
Persian (including Dari, Tajik)
Rajasthani (including Marwari, Mewari, Harauti, Bagri, Gujari, Malvi, etc.)
Sanskrit
Santali
Sindhi
Sinhala
Tamil
Telugu
Tibetan
Urdu

Group Two

Awadhi
Bhili
Bodo
Brahui
Chhattisgarhi
Chittagonian
Divehi (Maldivian)
Dogri

Dzonghka
Garhwali
Gondi
Haryanvi
Hazaragi
Hindko
Ho
Khandesi
Konkani
Kumauni
Kurukh/Oraon
Lamani/Lambadi
Manipuri (Meithei)
Mundari (Munda)
Prakrit
Saraiki
Sylheti
Tulu
Uzbek

Group Three

Aimaq
Bagheli
Balti
Bundeli
Burushaski
Garo
Halabi
Kachchhi
Kharia
Khasi
Khowar
Kohistani
Kok Borok (Tripuri)
Korku
Koya
Kui
Kuvi (Khond/Kondh)
Lushai (Mizo)
Mikir/Karbi
Miri/Mishing
Newari
Savara
Shina
Tamang

Tharu
Turkmen

IV. Teaching of LCTLs in the US. We have surveyed the teaching of South Asian languages in the US, both at National Resource Centers (at the top of Appendix II) and at other universities (at the bottom of Appendix II). This survey is based primarily on University on-line course catalogs as of Spring 2004. The table also incorporates data from an informal survey done by Herman van Olphen and the South Asian Languages Teachers Association (SALTA) in Fall 2003. As enrollment figures are difficult to obtain for these courses, the presence of a course in an online course catalog does not guarantee that there was sufficient enrollment for these courses to be actually offered. See the notes to the table for explanation of the numbers and abbreviations used.

The basic findings of the survey can be summarized as follows:

- Of the languages prioritized in Group One, 17 out of 25 are taught through the Intermediate level or higher at US universities. The eleven National Resource Centers for South Asia and the South Asia Language Resource Center offer the widest portfolio of the Group One languages.
- Hindi is now offered regularly in at least 35 universities across the United States, often in combination with Urdu.
- Persian is offered regularly in at least 15 universities.
- Of the 25 Group One languages, only Assamese, Baluchi, Bihari, Kashmiri, Oriya, Rajasthani, Santali, and Sindhi are not offered regularly anywhere.
- Many “off-campus” opportunities exist to study South Asian languages, including semester, year, and summer Study Abroad programs in South Asia. Many of these programs are sponsored by US universities and credit is issued through the sponsoring institution, however they have not been listed here.

**Appendix I:
List of South Asian Languages (based loosely on a minimum of 200,000 speakers)**

Aimaq as Persian	Ir	480,000+ in Afghanistan (1993); sometimes categorized
Assamese	In	15+ million (1997)
Awadhi	In	20 million (1999), but recorded as only 480,000+ under 1991 census; also classified as Hindi.
Bagheli	In	390,000+ (1997); also classified as Hindi
Bagri	In	2+ million; 1.8+ million India, 200,000+ Pakistan (1997). Included as Rajasthani.
Balti	TB	300,000+, mostly in Pakistan
Baluchi	Ir	5+ million (2001 Garry & Rubino)
Bengali (Bangla)	In	200+ million total, incl. 100+ million in Bangladesh (1994) and 80+ million in India
Bhili	In	1.3+ million Bhili proper (1994) (1991 Indian census: 5,572,000, but this apparently includes many smaller languages and dialects)
Bhojpuri	In	24.5+ million; 1+ million in Nepal (1997). Included with Maithili and Magahi as Bihari
Bihari	In	60+ million; includes primarily Bhojpuri, Maithili, Magahi; also Angika, Sadri, and others with lesser numbers as well
Bodo	TB	1.2+ million (1991 census); 500,000+ (1997)
Brahui	D	2+ million in Pakistan, 200,000+ in Afghanistan (1998)
Bundeli	In	600,000+ (1997); 1.6+ million (1991 census); also classified as Hindi.
Burushaski	Isolate	60,000+ in Pakistan (1981)
Chittagonian	In	14+ million (1998) Bangladesh; also classified as Bangla
Chhattisgarhi	In	11.5+ million (1997); also classified as Hindi
Divehi (Maldivian)	In	200,000+, national language of the Maldives.
Dogri (Kangri)	In	2.2+ million (1997)
Dzongkha	TB	160,000+, national language of Bhutan.
Garhwali	In	2+ million (1997); also classified as Hindi
Garo	TB	500,000+, 150,000+ in Bangladesh (1997)
Gondi	D	2.6+ million (1997)
Gujarati	In	45.4+ million (1997)
Gujari	In	1+ million; 600,000 to 700,000 in India (1996), 300,000 or more in Pakistan (1992). Included as Rajasthani.
Halabi	In	700,000+ (1994)
Haryanvi	In	13+ million in Haryana (1992); also classified as Hindi
Hazaragi	Ir	1.4+ million in Afghanistan (1989), 200,000+ in Pakistan; sometimes categorized as Persian.
Harauti	In	500,000+ (1997). Included as Rajasthani.
Ho	M	1+ million (1997)

Hindi	In	337+ million (1991 census), 364+ million (1997), 170,000+ in Nepal. Includes Hindustani, Haryanvi, Braj Bhasha, Awadhi, and Bundeli.
Hindko as Panjabi	In	3+ million in Pakistan (1993); also sometimes classified
Kachchhi	In	800,000+ (1997); also sometimes classified as Sindhi
Kannada	D	35+ million (1997)
Kashmiri	In	4.3+ million India; 100,000+ Pakistan (1997)
Khandesi	In	1.5+ million (1997)
Kharia	M	250,000+ (1994)
Khasi	M	800,000+ India, and 85,000+ Bangladesh (1997)
Khovar	In	220,000+ Pakistan (1992)
Kohistani	In	220,000+ Pakistan (1993)
Kok Borok (Tripuri)	TB	690,000+ (1997)
Konkani	In	Estimates range from 1.5 million (2001, Garry & Rubino) to 4+ million (1999, World Almanac)
Korku	M	400,000+ (1997)
Koya	D	330,000+ (1997); related to Gondi-Kui group
Kui	D	700,000+ (1997)
Kumauni	In	2.3+ million (1998); also classified as Hindi
Kurukh/Oraon	D	2+ million in India, includes 1.8+ million Oraon (1997)
Kuvi (Khond/Kondh)	D	300,000+ (1990); related to Gondi-Kui group
Lamani/Lambadi	In	2.8+ million (1994); also classified as Hindi
Lushai (Mizo)	TB	500,000+ (1997)
Magahi (Magadhi) as Bihari	In	11+ million (1997). Included with Magahi and Bhojpuri
Maithili	In	Estimates range from 7.7+ million (1991 census) to 24+ million (1981). Included with Magahi and Bhojpuri as Bihari
Malayalam	D	35.3+ million (1997)
Malvi	In	1+ million (1997). Included as Rajasthani
Manipuri (Meithei)	TB	1+ million (1997)
Marathi	In	68+ million (1997)
Marwari	In	12.9+ million in India (1997) and another 200,000+ in Pakistan (1998). Included as Rajasthani
Miri/Mishing	TB	400,000+ (1998)
Mewari	In	1.2+ million (1997). Included as Rajasthani
Mikir/Karbi	TB	400,000+ (1997)
Mundari (Munda)	M	2+ million (1997)
Nepali	In	9+ million in Nepal (1993), 58.3% of the population (1985); Estimates up to 16 million (2001, Garry & Rubino), including 10 million in Nepal and 6 million in India.
Newari	TB	600,000+ in Nepal (1991 census)
Oriya	In	31+ million (1997)
Pali	In	
Panjabi	In	67+ million total. 40+ million in Pakistan; 27+ million in India (1997)

Pashto	Ir	17.5+ million; 8+ million in Afghanistan (Southern Pashto) and 9.5+ million Pakistan (Northern Pashto) (1993)
Persian	Ir	Eastern Persian: 7+ million (includes Dari and Tajik; estimates may also include Aimaq and Hazaragi); mostly in Afghanistan
Prakrit	In	
Rajasthani	In	18.5+ million (1997) (includes Marwari, Mewari, Bagri, Gujari, Harauti, Malvi, and others)
Sanskrit	In	
Santali	M	6+ million total, of which 5.9+ million in India (1997) 150,000+ in Bangladesh (1993), and 30,000+ in Nepal (1991 Census)
Saraiki	In	15-30+ million in Pakistan (1998); also sometimes classified as Panjabi
Shina	In	300,000+ in Pakistan (1981 Census)
Sindhi	In	20+ million total; 16.9+ million in Pakistan (1993), 2.8+ million in India (1997)
Sinhala	In	13+ million in Sri Lanka (Official language) (1993)
Sylheti	In	5+ million in Bangladesh; also classified as Bangla
Tamil	D	66+ million; 61+ million in India (1997); 3+ million in Sri Lanka (1993)
Tamang	TB	600,000 to 700,000+ in Nepal (1991 census)
Telugu	D	69.6+ million (1997)
Tharu	In	900,000+ in Nepal
Tibetan	TB	Overall, 4-6 million (2001 Garry & Rubino)
Tulu	D	1.9+ million in India (1997).
Turkmen	T	500,000+ in Afghanistan (1995)
Urdu	In	60+ million; 10.7+ million first language speakers in Pakistan (1993); Estimates range up to 104-127 million (2001, Garry & Rubino)
Uzbek	T	Southern Uzbek: 1.4+ million in Afghanistan (1991)

Language Family Key

In	Indic (<Indo-Iranian, <Indo-European)
Ir	Iranian (<Indo-Iranian, <Indo-European)
D	Dravidian
M	Munda (<Mon Khmer, <Austro-Asiatic)
TB	Tibeto-Burman (<Sino-Tibetan)
T	Turkic (<Altaic)

Note: Population figures are from the 1991 Indian census (or the censuses, as noted, of other countries). Subsequent dated estimates are, unless otherwise noted, taken from Ethnologue. Garry & Rubino = Facts About the World's Languages, 2001.

Appendix II
Languages Taught in US Institutions of Higher Education

Institution	Hindi	Urdu	Persian	Bengali	Tamil	Sanskrit	Pali	Panjabi	Telugu	Malayalam	Tibetan	Gujarati	Nepali	Marathi	Sinhala
SASLI	2	2		1	1				1		2	1	2	1	
NC State	3, S2	3c	3	2	1										
UNC Chapel Hill	3	3c	2	2	2	2									
Duke	3														
NC Central	1		1												
Berkeley	3	3, S	2+		2+	2+		2							
Chicago	4	4	2+	4	4	4	4		2		2			2+	
Columbia	3	3	3	2	2	3		2			2				
Cornell	3	Script		3	?	2	1				?		3		3
Syracuse	3														
Michigan	3	3			3	2		1	2						
Pennsylvania	3	3	2+	2	3	2		2		2		2			
Texas	2+	2+	2+		3	3				2					
Virginia	3	3c	2+			1+					3				
Wash. U, Seattle	3	Script		1		3									
Wisconsin	3+	3+	3		2	2+	1		3+		2+		?		
Institution	Hindi	Urdu	Persian	Bengali	Tamil	Sanskrit	Pali	Panjabi	Telugu	Malayalam	Tibetan	Gujarati	Nepali	Marathi	Sinhala
Alabama	1	1				1									
Arizona			3												
Brown	3					3									
Cal State, Sacto.								1							
Emory	3		3			2									
Florida State						1+									
Georgia	2					2									
Harvard	3c	3				3+	2				3		1		
Hawaii	2					2									
Illinois	3					1+									
Indiana	2					2					2				
Iowa	3					2									
Johns Hopkins	2		1												

Loyola, Chicago	2														
Minnesota	3+	2				2+									2+
Monterey Inst.			1												
No Illinois															
NYU	3	2	2	2	2			2							
Ohio State			3+												
Princeton			3												
Rice	2														
Rutgers	1		3												
San Jose State								2							
So Methodist	2														
SUNY – Stonybrook	1					1									
UCLA	2					?									
UCSB	2		1			2+	1	S1			2				
UCSC	2	Script													
U Mass Amherst	1														
Wash. U, St. Louis	2		3												
William Paterson	1														
Yale	3														
Institution	Hindi	Urdu	Persian	Bengali	Tamil	Sanskrit	Pali	Panjabi	Telugu	Malayalam	Tibetan	Gujarati	Nepali	Marathi	Sinhala

Notes:

1. Numbers indicate how many years of instruction are offered regularly
2. A “c” after the number in Urdu means the course is taught in combination with Hindi
3. A “+” after the number means there are literature or other courses offered after the indicated level, but they are not specifically labeled as “intermediate,” “advanced,” etc.
4. An “S” in a language column indicates that this language is offered during a study abroad program in India. For example the designation “3, S2” indicates that the university offers three levels of the language during the normal semester, and 2 levels of the language during a study abroad program.
5. Several universities (e.g. Yale, Stanford, Arizona) have large Self-Instructional Language Programs (SILP), under which rubric South Asian languages are offered. Since these courses are not taught regularly, or by regular full time trained faculty, we have not included them above.
6. Some universities (e.g. Cornell, U Wash Seattle) offer a course introducing students to the Urdu script, but no courses in Urdu language per se. Presumably

- the students who take such courses a) already know how to speak and simply want to learn to read (i.e. heritage students), or b) are learning Hindi concurrently.
7. The US Dept. of State's School of Language Studies in the Foreign Service Institute offers courses in most of the South Asian languages, however these courses are open only US government employees and course information is not listed on the Internet.
 8. A Question mark indicates that the University's website indicates that the course exists, but the course does not appear in the current course listings. Such courses may be offered only sporadically, or may have been discontinued.
 9. This table was compiled primarily by John Caldwell, NC Center for South Asia Studies (caldwell@unity.ncsu.edu).