Lesser-Known Languages of South Asia

W
DE
G

Trends in Linguistics
Studies and Monographs 175

Editors

Walter Bisang
Hans Henrich Hock
Werner Winter
(main editor for this volume)

Mouton de Gruyter
Berlin · New York

Lesser-Known Languages of South Asia

Status and Policies, Case Studies
and Applications of Information Technology

edited by

Anju Saxena
Lars Borin

Mouton de Gruyter
Berlin · New York

♾ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

Preface

This volume grew out of our work on exploring the possibilities of using technology – specifically the modern *information and communication technologies* (ICT), and particularly *language technology* – in support of language documentation, language learning, and language maintenance, especially as these apply in the context of languages and cultures of South Asia, with particular emphasis on lesser-known South Asian languages.

Six of the papers presented here (those by Allwood, Borin, Grinevald, Nathan and Csató, Noonan, Singh) are extensively revised versions of papers read at a panel on "Globalization, technological advances and lesser-known languages in South Asia" organized by Anju Saxena in connection with the *18th European Conference on Modern South Asian Studies* at Lund University, Sweden, 2004, while the remaining 11 contributions have been solicited specifically for this volume.

The work on this volume has been funded in part by the Swedish Research Council (*Vetenskapsrådet*).

We would like to thank the series editor, Professor Werner Winter, for his encouragement and support during the preparation of this volume. We would also like to thank Birgit Sievert at Mouton de Gruyter for her advice at all stages of our long and sometimes crooked path from the first manuscript to the finished book, and John Wilkinson for his help in preparing the camera-ready copy.

Finally, a small point of orthography: When we write about the perhaps most salient aspects of modern ICT in this volume, we have decided to follow *Wired News* (among others) in not capitalizing the words *internet* and *web* (Tony Long: It's Just the 'internet' Now, *Wired News* Aug. 16, 2004. <http://www.wired.com/news/culture/0,64596-0.html>, accessed June 13, 2006).

<div align="right">

Anju Saxena
Lars Borin

</div>

Contents

Information and communication technologies and languages of South Asia

Introduction

Anju Saxena

> [I]t is alright to be Native, to speak the Native language, and
> to use Native tools and implements in play and work. After
> all, our technology was made by our ancestors to edify our
> Native worldviews. Please, what ever you do, do NOT give
> to the youngsters the idea that modern technology has an
> answer for everything. It does not. Use it merely as a tool,
> and use it minimally and judiciously. Remind the students,
> that technological tools are intensive in the use of natural
> resources and energy. To accept technology blindly is to
> negate the painful works to revitalize our Native languages
> and cultures. (Kawagley 2003: ix–x)

1. Going, going, gone: Vanishing languages and cultures[1]

The increasing globalization in the twentieth century, with a small group of
nations dominating the scene, has had an adverse effect on the maintenance of
social and cultural traditions of many communities. The pull factor (good
employment opportunities, standard of living, etc.) and the push factor (larger
and better trained and equipped armies, more modern weapons, etc.) have
conspired to make some groups socio-economically dominant, and as a conse-
quence promoted the cultures and languages of these groups over those of
other, non-dominant groups (Crystal 2000; Nettle and Romaine 2000), to such
an extent that the existence of a large number of smaller languages is threat-
ened.

   According to one estimate (Krauss 1996), 3000 of today's 6000 languages
will disappear in this century, if no special measures are taken. Issues relating
to language death, endangerment and threat to language diversity have come
to the foreground of linguistic discussion (Krauss 1992, 1996; Hale 1992a)
and efforts to revitalize endangered languages and to halt or prevent language
death have been the themes of several conferences (including a UN confer-
ence; see Bradley and Bradley 2002a).

   The term *language shift* refers to a situation where the use of a language is
replaced by the use of another (usually a socio-economically or numerically
dominant language). The end product of language shift is complete replace-
ment, or language death, but it is normally a gradual process, where a shift in

progress can affect a language in terms of the number of its speakers, the functional domains in which it is used and the degree of competence in the language (Dressler and Wodak-Leodolter 1977; Dorian 1989; Brenzinger 1992; Craig 1992; Grinevald 1997, 1998; Grenoble and Whaley 1998; Nettle and Romaine 2000; Bradley and Bradley 2002b). Linguists have noted the existence of language death[2] and language shift for quite some time (e.g., Swadesh 1948; Weinreich 1953: 106–110). However, since the 1960s increasing attention has been paid to language shift by linguists, who have been interested in studying the linguistic structure of the languages involved in language shift situations, where adjustments at all levels (phonological, lexical, grammatical) have been observed. In this connection, linguists have also been interested in examining if linguistic systems of dying languages ("obsolescent languages") show patterns which are just the opposite of creolization or first language acquisition (e.g., Dorian 1981; Dressler and Wodak-Leodolter 1977; Mithun 1989; Romaine 1989; Schmidt 1985; Trudgill 1978).

Factors such as migration, industrialization, urbanization, globalization, religion, government policies (e.g. the choice of the medium of instruction in schools, laws relating to language policies) and changing patterns of economy have been pointed out as potentially contributing to language shift and language death.

Social changes brought about by factors such as these may influence an individual or a speech community to revise his or its perceptions of his own self or of his language and/or their perception of the language of the other group or of the world. This may lead individuals or speech communities to change their pattern of language choice. Language shift is, in many cases, closely tied to ethnicity. In language shift situations, the language shift tends to take place when speakers want to leave behind a stigmatized ethnic identity and adopt a positive ethnic identity of some other group as a possible means for upward social mobility. A shift in the language choice patterns then becomes a means – a tool – for upward mobility (Dorian 1981).

Thus, one important factor in language shift – perhaps the most important factor – is arguably that of speaker (community) attitudes, which in turn are rooted in economical or political realities.[3] It is worth noting at this juncture that attitudes reveal themselves not so much in words as in actions, since the two often seemingly contradict each other. See the discussion of prior ideological clarification as the essential beginning for any program dealing with language and cultural preservation in Dauenhauer and Dauenhauer (1998: 62–66). Winter (1993) presents the relevant cases of (1) a Hualapai language revival activist and schoolteacher, who, while actively working for the use of Hualapai in school, nevertheless spoke only English to her children at home,[4]

and (2) a Bantawa couple who worked actively to promote Bantawa in various ways, but communicated with each other and with their children in Nepali and English.[5] Winter comments (1993: 311):

> What is to be observed in both cases is a conflict between wanting to do something for the language and wanting to improve the chances of the children to succeed in the macrosociety of which they are, and always will be, part. The linguist observing this state of affairs may feel regret at what is happening here; but if it is a fact that maintaining a small language at the expense of a major or national one means severely reducing prospects of an economically satisfactory life for one's children, does one have a right to blame the parents?

In the terminology of Freilich (1991), this represents an attempt to use *smart* means to achieve *proper* goals. By these terms, Freilich aims to capture the oft-observed tension in all kinds of human communities between on the one hand that which culture, in the form of tradition, requires of us – this is what is "proper" – and on the other hand "smart" actions – which break the letter of proper rules – and which are brought about by the pragmatics of a continually changing social environment in which we have to survive. This is a generally useful distinction, although this is not as clear in the case of language as in the other manifestations of culture discussed by Freilich, mainly because the only effective way of achieving the "proper" goal of preserving the language seems to be by actually using it. As a general strategy it goes some way toward explaining how a situation such as that cited above might come about, however. The speaker simply may not be aware that language constitutes a special case. This distinction still seems useful, since, arguably, there are ways in which smart means can be used to achieve proper goals in this sense, the creative use of new technologies possibly being one such (see section 4 below).

Language death is not a new phenomenon. Languages have disappeared all through recorded history. Classic examples are Gothic, Sumerian and Hittite, to mention a few, and in the past five hundred years we have lost half of the known languages of the world (Sasse 1992). But what makes this issue especially grievous in modern times is the changing world scene. Factors such as internationalism and globalization, a modern supraregional economy and media of mass communication have intensified the situation where a small group of politically and economically dominant communities and their languages manifest too great a power on a large number of small communities. Hale (1992a: 1) elaborates the differences between the earlier language death phenomenon and the situation we are facing today:

> [L]anguage loss in the modern period is of a different character, in its extent and in its implications. It is part of a much larger process of LOSS OF CULTURAL AND INTELLECTUAL DIVERSITY in which politically dominant languages and cultures simply overwhelm indigenous local languages and cultures, placing them in a condition which can only be described as embattled. The process is not unrelated to the simultaneous loss of diversity in the zoological and botanical worlds. [emphasis in the original]

Language death arguably affects even the prerequisites for maintaining biodiversity (Skutnabb-Kangas 2000; Nettle and Romaine 2000). According to Skutnabb-Kangas, language diversity is disappearing at a faster rate than biodiversity. Her prognosis for year 2100 is (expressed as percentage of diversity lost): Biodiversity 2% but linguistic diversity 50% (optimistic forecast) and biodiversity 20% but linguistic diversity 90–95% (pessimistic forecast), highlighting the urgency of the matter.

Despite the depressing facts about the degree of language loss, there are some positive signs. The 1990s brought language endangerment to the forefront of the linguistic and political arenas, and some first steps were taken in order to turn the tide. This includes efforts by some communities involving local, national and international organizations and institutions.[6] The Hualapai Bilingual/Bicultural Education Program (Peach Springs, Arizona) is basically a local program which has been instrumental and effective in developing regional and national movements influencing Native American languages and their communities, e.g., the initiation of the American Indian Languages Development Institute and the Native American Languages Act (McCarty and Watahomigie 1999; see also presentations of various projects in the series of books published by Northern Arizona University: Cantoni 1996; Reyhner 1997; Reyhner et al. 1999, 2000, 2003; Burnaby and Reyhner 2002). Revitalization efforts are going on in smaller as well as larger communities (e.g. Mayan: England 1992, 1998; Rama: [Grinevald] Craig 1992; Grinevald 1998, 2005a; Hawai'ian: Warschauer and Donaghy 1997; Wilson 1999). There have been publications such as Dorian 1989, Grenoble and Whaley 1998, Crystal 2000, Nettle and Romaine 2000, Hinton and Hale 2001, Bradley and Bradley 2002b, *UNESCO's Red book on endangered languages*,[7] as well as conferences (e.g. the Endangered Languages Symposium organized by the Linguistic Society of America 1991), and the establishment of funding programs, such as the Hans Rausing Endangered Languages Project (HRELP) at London University's School of Oriental and African Studies.

This situation – a general loss of linguistic and cultural diversity and occasional efforts to counter the trend, i.e. using modern information and communication technologies – is prevailing everywhere in the world, including South

Asia. However, the linguistic situation in South Asia has been a bit out of focus in recent literature on language shift and language endangerment (Payne 1999; Ostler and Rudes 2000), with some notable exceptions (e.g. Abbi 1997; Saxena 2004). The aim of the present volume is to discuss the status of the lesser-known languages in South Asia and to discuss how modern technology can be a tool in documenting these languages and in spreading awareness about them. Issues that arise while applying technology developed using primarily Western literate languages to these for the most part oral languages will also be taken up here. The volume contains articles on the linguistic situation of South Asia, both general overviews portraying individual South Asian countries (Rahman, Singh, Turin), and case studies of particular South Asian language communities and/or sociolinguistic situations (Abbi, Kohistani and Schmidt, Zeisler). A number of articles raise issues of the impact of modern information and communication technology on lesser-known languages in general, and on South Asian language communities in particular (Annamalai, Bradley, Noonan, Renganathan and Schiffman), whereas others describe linguistic and cultural documentation work being carried out for South Asian languages (Hardie et al., Michailovsky), and some of the ethical issues raised in connection with linguistic fieldwork and language documentation (Grinevald). Finally, some of the contributions illustrate how cutting-edge information and communication technologies can be brought to bear on the problems of lesser-known language documentation and maintenance (Allwood, Borin, Nathan and Csató, Trosterud).

A note on terminology

Many terms are used in the literature to refer to the languages that are the focus of this volume. *Minority languages*, *indigenous languages*, and *endangered languages* are the terms most often met with in the linguistics literature, and in Indian literature the term *tribal languages* appears.[8] Elsewhere, e.g. in the language technology literature, one encounters terms such as *lesser used languages* (a term used officially in the European Union), *less prevalent languages*, *small(er) languages*, *low-density languages*, *vernacular languages*, *dialects*, *lesser-known languages*, and *less frequently taught languages*. See also Grinevald's article in this volume. In the more computer science-oriented presentations of work on language technology, one is sometimes confronted with the revealing "pseudo-term" *non-English languages*. This confusing multiplicity of terms is due at least to the different backgrounds of the scholars working in this area and also to the weights accorded the different criteria for classifying languages as distinct from, e.g., English.

In addition to this, many, perhaps all, of these terms are loaded in some way or another with pejorative connotations, or are ideologically charged in some way. Thus, it is no easy task to choose a general term for an introductory chapter such as this. However, we have decided to opt for *lesser-known languages*, as it is a relatively untainted term.

2. Language and linguistic diversity in South Asia

One sees some interesting patterns in different parts of the world concerning the direction of language shift. In the Americas and Australia the shift has mainly been to the languages of the colonial rulers (Spanish, Portugese, French and English) whereas in some other regions such as in Africa there is often a shift towards a non-colonial language (e.g. Amharic in Ethiopia, Bambara in Mali and Swahili in Zaire/Congo). In South Asia, some locally dominant languages (Hindi, Urdu, Nepali to mention a few, beside English, the colonial language) are gaining ground at the expense of the lesser-known languages.

The Indian subcontinent has a long history of linguistic diversity and multilingualism, spanning more than three millenia. Languages spoken in this region belong to at least four major language families: Indo-European (mostly Indo-Aryan), Dravidian, Tibeto-Burman and Austro-Asiatic. Societal multilingualism is an established tradition in South Asia, where not all of languages which are spoken in one community are employed in all spheres of activity (Pandit 1972). Despite this stable multilingualism, language death is not uncommon in the South Asian context.

As is typical of most of South Asia, speakers of lesser-known languages in India – the largest and most populous country in the region (population 1080 million in 2005) – are already or are increasingly becoming bilingual. Concerning the 114 languages mentioned in the Indian census, the rate of bilingualism recorded in the past four censuses indicates that bilingualism has doubled in 30 years from 9.7% in 1961 to 19.44% in 1991 (Bhattacharya 2002). While speakers of lesser-known languages learn the language(s) of the dominant group, the reverse is usually not the case. Whereas many adult Kinnauri[9] speakers, for example, speak Kinnauri as their mother tongue, and many elders living in the region are strongly monolingual, children and young adults are in very large numbers active bilinguals, with a preference for Hindi or the regional Indic variety. Many young people migrate outside this area for education and employment purposes, where the lingua franca is not their mother tongue. Such social situations have important linguistic consequences for these languages. Indigenous languages with no written tradition and with no or very little political and/or

economic power at the local and national level fall by the wayside *en route* to modernity either completely, or are given up in particular contexts.

While some of the languages (such as Hindi, Tamil and Bangla) have a long written literary tradition and there has been much work done on these languages, very little is known about many, perhaps most, languages of this region. A case in point is a language such as Great Andamanese, with only a handful of speakers and almost no documentation (for details, see Anvita Abbi's article "Vanishing voices: A typological sketch of Great Andamanese" in this volume). Similarly, there is a great disparity in the number of speakers. While the then 18 scheduled languages[10] constitute 96.29% of the total population, the remaining 96 non-scheduled languages (including 0.07% who speak "other languages", defined as those languages which have less than 10 000 speakers) are spoken by 3.17% of the population according to the 1991 census. (Bhattacharya 2002: 58)

It is impossible to say anything concrete about the extent of language endangerment in India. Information on language at the national level has been collected since 1881 as part of the Indian census, held every 10 years, but census reports provide almost no concrete information about languages with less than 10 000 speakers (in other words, about endangered languages). Further, motivations for the distinction between language and dialect are not always clear. The census figures are based on self-reporting by language users, so that if a particular language is provided as the mother tongue in the census returns by an individual or by a group, this may at times be a reflection of loyalty more than an indication of actual language proficiency (Southworth 1978). Furthermore, how and which languages are taken into consideration is a complicated matter. The 10 400 mother tongue names returned in the 1991 census of India are reduced to 113 languages, plus one other mother tongues category for all languages with less than 10 000 speakers (cf. the 386 languages listed for India in the *Ethnologue*).[11] However, many names are discarded in the process, on grounds that are not always clear (Annamalai 2003). Udaya Narayana Singh's contribution to this volume, "Status of lesser-known languages in India", presents an overall picture of lesser-known languages of India and their current status, focusing on the constitutional provisions that exist in India with respect to minor and minority languages, language policy issues and problems with implementation of official decisions. Population growth, economic growth, urbanization, literacy and education are mentioned as factors which slow down the process of implementing laws and policies furthering the use of lesser-known languages in India today.

Considering the fact that even today around 80% of the population in India still live in rural areas, this may lead one to believe that this multitude of lan-

guages are well and thriving. This, however, is far from always the case, largely because of extralinguistic factors, such as the medium of instruction in schools, social mobility, administrative language, modern media such as television, etc.

However, there are also some success stories, such as that of Santhali in the newly formed 28th state of Jharkhand. Languages belonging to three language families are spoken in this region: Indo-Aryan (e.g. Sadari, Hindi, Bengali), Dravidian (Kurux, Malto) and Austro-Asiatic (e.g. Mundari, Ho, Santhali). Hindi and English are dominant languages used widely in public spheres, while the lesser-known languages are primarily used for in-group communication. They are not used in government offices, state legislation, business or legal matters. In Jharkhand, one sees two opposing trends: On the one hand, there are signs of promoting English (e.g. the Jharkhand government's proposal to introduce English as a subject from first grade onwards in schools) and on the other hand, one sees attempts to strengthen and to make more visible some lesser-known languages – in particular Santhali – of this region. Kurux, Mundari, Ho and Santhali have been introduced as the medium of instruction in primary schools, as well as provided as optional subjects in secondary schools. These languages are also offered at Ranchi University at graduate and post-graduate levels. There are also other similar efforts to make these languages more visible (e.g. there are 3 newspapers and 20 magazines published in Santhali by private organizations). There have also been efforts by various organizations to include Santhali written in the Olchiki script[12] as the official language of Jharkhand in the VIII Schedule of the Indian Constitution. (Mohan 2002: 230–240)

The linguistic situation in Nepal is also one of great language diversity – the *Ethnologue* lists 120 languages in Nepal, spoken by a population of 28 million (in 2005). As in so many other places, we are faced here with a situation where lip service is paid to language diversity, but where in reality the situation is one where there is one dominant language, viz. Nepali, an Indo-Aryan language. According to Winter (1993), there are signs of decreasing language diversity also in Nepal, earlier so that larger languages would encroach on smaller ones in the same region, but increasingly – following in the wake of national centralization and a growing supraregional economy – so that Nepali tends to take over as the language of all walks of life. However, beginning in the 1990s, there seems to be a growing awareness in Nepal about the situation of lesser-known languages, and more enlightened language policies are being formulated. Progress is slow, however.

"Minority language policies and politics in Nepal" is the theme of Mark Turin's article in this volume. Turin points out that at the policy level, there

are some positive developments, where there is a shift from the "one nation, one language" policy in the 1950s to "an acknowledgement of the multi-ethnic and multi-lingual nature of the country post-1990. But these policies have not been put to practice to the extent one wishes for. Recognizing the important role script and literacy play (including the fact that the state provides more resources to written languages), Turin describes trends noticeable in Nepal today. While some organizations advocate either Devanagari or Tibetan scripts, there are communities who probably in their effort to establish their own identity are trying to devise their own new scripts.

The linguistic scene in Pakistan is similar to the Indian and Nepalese situation described above. Pakistan, like India and Nepal, is a multilingual state with 6 major and about 59 minor languages, spoken by a population of 162 million (in 2005). Tariq Rahman in his article "Language policy, multilingualism and language vitality in Pakistan" provides an overview of the linguistic scene in Pakistan with special reference to the unequal status of English and Urdu on the one hand, and lesser-known languages on the other. He further discusses some factors which have contributed to this unequal status. Government policies, according to him, is one significant factor. He attributes to English and Urdu the status of the languages of power in Pakistan. English is considered a symbol of power, sophistication and prestige, whereas small minority languages have a negative image associated with them. This trend is leading to language death in some cases and marginalization in others. Rahman advocates the promoting of additive multilingualism as a means to improve the status of these marginalized languages.

A concrete case study of the changing linguistic scene in Pakistan is presented in Razwal Kohistani and Ruth Laila Schmidt's article "Shina in contemporary Pakistan" in this volume. The focus of their article is on Shina, an Indo-Aryan language of the Dardic subgroup spoken in the Karakorams and the western Himalayas. They describe the increasing marginalization of the use of Shina, attributing it to factors such as modern education, advancements in the media, and communication. They point out that Urdu and English – which are dominant languages in the whole of Pakistan – as well as Pashto (the dominant language of the region) are gaining ground at the expense of Shina. In its urban center, Gilgit, Shina "has suffered a loss of prestige" and in the rural areas Shina is used (at least at present) both in public and private domains, but they fear that this relegation of the use of Shina to rural areas and private domains is preventing Shina from developing a standard language and literature. There are, however, some forces working against this development in the region (e.g. intellectuals who work in favor of Shina and Islamic missionaries who target the grassroots of the population).

3. Loss of linguistic diversity and the need for language documentation

There is a growing awareness about the negative consequences of language death and the concomitant loss of linguistic diversity. From the linguist's point of view, language diversity is essential for linguistic theory building and for a scientific study of mind and language (e.g. Hale 1992b, 1998), for which it is imperative that we have access to data from languages representing rich and diverse linguistic structures, underscoring the need for documentation and preservation of languages.

A language is a reflection of the community that speaks it. It embodies the philosophy and the world-view of its people. In communities which lack a writing system, this knowledge is handed down orally from one generation to the next. When a language dies, we lose not only the linguistic knowledge of that community, but also the knowledge about its culture:

> The most important relationship between language and culture that gets to the heart of what is lost when you lose a language is that most of the culture is in the language and is expressed in the language. Take it away from the culture, and you take away its greetings, its curses, its praises, its laws, its literature, its songs, its riddles, its proverbs, its cures, its wisdom, its prayers. The culture could not be expressed and handed on in any other way. What would be left? When you are talking about the language, most of what you are talking about is the culture. That is, you are losing all those things that essentially are the way of life, the way of thought, the way of valuing, and the human reality that you are talking about. (Fishman 1996: 81)

The loss of artistic and intellectual resources accompanying the loss of language has been addressed in the literature by a number of linguists. Mithun (1998), for example, presents some linguistic features of Central Pomo and Mohawk to illustrate how some specific ways that these languages conceptualize the world will be lost, if the languages are lost. On a similar note, Woodbury (1998) presents some cases to illustrate that the loss of a language implies the inability to express particular concepts. Cup'ik Eskimo has a series of affective suffixes with translations as 'poor dear N; poor dear (subject) does V'; 'darned N; darned (subject) does V'; 'funky N; funky (subject) does V'; and 'shabby old N; shabby old (subject) does V' (Woodbury 1998: 240). In English there are no affixes expressing these meaning(s), so one is forced to use lexical items, if anything. Woodbury conducted an experiment where a speaker first told the story in Cup'ik and then narrated the same story in English and finally provided a sentence-by-sentence translation. The results of this experiment showed that in the sentence-by-sentence translation there were no words expressing the meanings expressed by affective suffixes and in

the free English narration there were only a few items expressing the meaning of the affective suffix, suggesting that the interpretation of these affective suffixes can, at the most, be captured only poorly in the English translation.

Similarly, the disappearance of a language may also imply loss of culture-specific information. Like many other smaller communities, the Mohawk people believe that they do not cease to be Native Americans if they do not speak their language. Jocks (1998) demonstrates convincingly how if a community does not have a rich knowledge of its cultural tradition manifested in its language, that community may become a caricature of itself, as it were. Traditional ceremonies, for example, may not only become formalized rituals: there is also a risk that translations of traditional ceremonies, for instance, may implicitly bring with them conceptions that outsiders have of these indigenous communities. He illustrates his point by pointing out differences in the conceptualization of knowledge in English and Mohawk: In English, knowledge is something which one can POSSESS, whereas in Mohawk, knowledge is an ACTIVITY (something one does and which must be maintained).[13]

One such unique linguistic/cultural configuration is described by Anvita Abbi in her paper "Vanishing voices: A typological sketch of Great Andamanese", in the form of a case study of Great Andamanese. Previous studies suggest that Great Andamanese could represent the remaining linguistic link to pre-Neolithic Southeast Asia. Great Andamanese has 13 speakers, highlighting the urgent need to document and describe this language. In this paper Abbi presents the results of her pilot study, outlining the phonological, morphological and syntactic features of Great Andamanese. Abbi's article is illustrative of many languages which are in danger of extinction because of changing socio-cultural patterns – languages which we know almost nothing or very little about.

Against this background, it is perhaps not surprising that much emphasis in recent times has been put on the need for documentation of lesser-known languages, especially endangered languages. Earlier, and to some extent still today, we often see linguists referring to their chief activity as description of languages. These are related, but not identical, activities. Conceptually, documentation precedes description:

> LANGUAGE DOCUMENTATION provides a record of the linguistic practices of a speech community, such as a collection of recorded and transcribed texts. LANGUAGE DESCRIPTION, on the other hand, provides a systematic account of the observed practices in terms of linguistic generalizations and abstractions, such as in a grammar or analytical lexicon. (Bird and Simons 2003: 557)

Logically, then, documentation in this sense can be the basis for description, but not vice versa. The products of documentation – including linguistic descriptions – it is increasingly realized, can be used to support languages that are still used, whereas mere description without documentation cannot be used to revitalize languages where there are no or almost no living linguistic practices left. There is at least the hope, however, that description and documentation together – "preservation for the record" in the words of Allwood (this volume) – could be used to accomplish this.

Unlike traditional linguistic descriptions, then, where the secondary products of the primary linguistic materials – grammars, dictionaries, presentation of theoretical analyses of various linguistic phenomena, etc. – were in focus and the linguistic data itself was not seen as primarily interesting,[14] in language documentation the focus is on primary linguistic material in a representative spectrum of genres, with an emphasis on naturally occurring discourse in different speech situations. Another objective is to include not only linguistic material but also material which provides some insights into the cultural aspects of these societies. This means that language documentation in fact has much in common with modern corpus linguistics (see Borin's contribution in this volume).

It should follow from Bird and Simons's characterization, quoted above, that literacy automatically implies documentation of a language, provided that writing and its products can be considered part of "the linguistic practices of a speech community". In this sense, then, language documentation has been going on for a very long time, at least in some cases, on clay tablets, on papyrus scrolls, on runestones, on bark, on wood, on leather, on paper, on bricks, on cloth, etc. In the same way, language description has a long history, but a peculiar one in the case of languages other than the classical languages or the new national languages of Europe (in the case of which it is perhaps better to speak of "language prescription"). The flip side of this particular coin is, again, that most of the lesser-known languages everywhere – South Asia being no exception in this regard – are non-written languages.

Woodbury (2003) attributes this recent interest in linguistic documentation to three elements, namely, an increasing awareness about diversity among and within languages not as a kind of aberration, but as an intrinsic definitional feature of language, of the threat of language endangerment, and of technological advances opening new possibilities for documenting linguistic data. He also points to a growing realization among linguists that primary linguistic data have never been properly theorized, but remain largely epiphenomenal to the generalizations expressed in grammatical discourse, in a way which in other sciences would be considered quite naive (Woodbury 2003: 40).

Austin (2003) emphasizes the need to talk about guidelines, e.g. ethics, relations with the language community, our responsibility to the community, to researchers and to the discipline (Grinevald 2005b). In her article "Worrying about ethics and wondering about 'informed consent': Fieldwork from an Americanist perspective", Colette Grinevald discusses a number of ethical (and at their core eminently practical) issues that arise in connection with linguistic fieldwork, driving home the point that linguistic fieldwork, in particular fieldwork on languages facing extinction, will have to deal with "a complexity of pressures which academia and financing foundations may have very little sense of as yet", because "fieldwork projects are not laboratory experiments", and forging a long-term working relationship with the language community "is one of the most challenging of the multiple responsibilities that fall on the fieldworkers, who are academics usually raised and trained far away from the realities of the field".

The need to discuss ethical guidelines has arisen from a realization that linguists and other fieldworkers are just as likely to be caught in the trap of ethnocentrism as anybody else, but that awareness-raising is one way of avoiding this. At the same time, indigenous language communities have realized and increasingly begun to question the way their cultures and languages are portrayed by outsiders:

> With the growing interest in things Indian in the United States and around the world, Native American culture has become a highly saleable commodity … While this commercialization of Indian culture might seem to make good business sense to the Anglo-American majority, many native people experience it as an expropriation of their heritage by the dominant society. This taking is understood to involve the alienation, popularization and corruption of native traditions and imagery through their unauthorized reproduction and commercial exploitation by non-Indians. There is widespread consensus among native spokespeople that such 'cultural appropriation' is as potentially damaging to the survival of native ways of life as the expropriation of Indian lands in the nineteenth century, or the assimilationist strategies pursued by the Indian Schools. (Howes 1996: 138)

4. The role of technology in language preservation and loss

Modern technology – here I include both the somewhat older broadcast (analog) mass media technologies radio and televison, and the newer (digital) so-called information and communication technologies (ICT), i.e., computers, the internet, cell phones, interactive digital cable television, etc. – have been depicted as both foe and friend with respect to non-mainstream cultures and lesser-known languages. The former view is reflected in Krauss's (1992) characterization of television as "cultural nerve gas". Many researchers and

other observers perceive that modern mass media pose a threat to diversity, forcing everything that comes in their way into the same cultural and linguistic straitjacket. E. Annamalai in his article on "The impact of technology on language diversity and multilingualism" describes how changes in the socio-cultural structure of a community (including the introduction of new technology) has a strong impact on its language, drawing on his work on the Andamanese language (Annamalai and Gnanasundaram 2001).

On the other hand, especially the most recent information and communication technologies (ICT) are often seen as holding great promise for the documentation, protection and promotion of language diversity, creating unprecedented opportunities for small language communities (e.g. Bredin 1996; Cazden 2003).

In order to discuss the role of technology *vis-à-vis* lesser-known languages, it will be appropriate to keep separate certain different aspects of modern information and communication technology, viz. its *form* (what *could* faithfully be conveyed by it), its *content* (what *is actually* conveyed by means of this technology), and its *uses* (more generally how technology can be potentially beneficial to small languages and cultures). But first of all, of course, one needs access to computers and the skills to use them, which is generally less likely to be the case in lesser-known language communities (McHenry 2002), illustrating another aspect of what has been called the "digital divide".

The form of ICT is relevant at least in two respects. Firstly, we are still very far from the hypothetical ideal state where texts in any (literary) language can be input, stored, processed, and presented on equal terms with all other languages in word processors, on the web, in email, in chat rooms, etc. This has to do with developments in the areas of input methods (e.g. for scripts with large character inventories), character coding and rendering, and software for natural language processing. It has to be emphasized at this point that the issue of development or non-development in these areas is not primarily a technical issue (although there is also a technical dimension to it), but has everything to do with policy and a will to have things be a particular way.

David Bradley's article "Lisu orthographies and email" in this volume reports the case of the use of Lisu on the internet and a revision this media has necessitated in the Lisu writing system. Lisu is a Tibeto-Burman language spoken in India, Burma, Thailand and China. It has a Latin-based orthography. The writing system uses upper case letters, upright and inverted. A revised version of this orthography has been devised for the internet and Bradley reports that its use is gradually spreading.

Renganathan and Schiffman's article "The impact of technological advances on Tamil language use and planning" highlights the complex and

intertwined nature of the status of the concerned languages and its effect on the application of technological advances in South Asia. The focus of their article is on Tamil (though not a lesser-known language, one which faces competition from English in some public domains). Tamil, like other languages in South Asia, faces a challenging situation. These languages have to struggle for their survival and use in fields such as science and technology, where English has been (and still is) the dominant language. Language activists push for the use of these languages in all domains, including university education. This is in sharp contrast to the prevailing situation in higher education institutions which promote English, e.g. by using English as the medium of instruction, by (explicitly or implicitly) encouraging academic publications in English (rather than in Tamil, for instance). This obviously hampers or slows down the application of recent technological advances to Tamil. Despite this, some efforts have been made in the fields of Tamil computing and language technology. Some examples concern the creation and use of technical vocabularies in Tamil and the development of localized software.

The second important aspect of the form of ICT in this context is the circumstance that the technology is still predominantly geared toward the written language. Thus, only literary communities can make full use of it (whereas many lesser-known language communities are exclusively or primarily oral; see, e.g., Bernard 1996; Buszard-Welcher 2001).

It is frequently remarked in the literature that literacy is a prerequisite for the long-term survival of a language in the modern world. Bernard (1996) and Borin (this volume) make a useful distinction between two quite different usages of written language, noting that many languages of the world have been written, often by linguists but sometimes even by native speakers, without developing a literary tradition (Bernard 1996: n.p.). Only if a language is literary, rather than merely written, will it stand a chance in the long term, it is claimed. Michael Noonan in his article on "The rise of ethnic consciousness and the politicization of language in west-central Nepal" argues that standardization is a necessary component of a literary language, meaning that a standard orthography be devised, that a uniform spelling of words be introduced, that a canonical form be selected from among variants used by speakers, etc. Noonan observes that ethnic consciousness is a relatively recent phenomenon in west-central Nepal. Despite some official rhetoric on this matter, not much is done in reality to preserve and promote other languages than Nepali in education and other domains. He recounts the case of two cousins in Nepal. Both are fluent speakers of Chantyal, a Tibeto-Burman language, and who regularly exchange emails in Nepali – in which both are also fluent – written in Latin script transcription on a keyboard configured for English. Somehow the

notion of instead writing the emails in Chantyal never occurred to the two correspondents, presumably because, in Noonan's words, "people ordinarily write the languages they were taught to write in school". Some lesser-known languages in Nepal, including Chantyal, have written forms of which their users are aware, but just like earlier, they are still hardly used in the school system (see also Bernard 1996). Noonan here calls attention to a chicken-and-egg quandary involving the relationship between the availability of primary education in a language and a standardization of that language, concluding that this is ultimately a political matter, but that the will to realize that this is so, let alone act on this realization, still seems to be lacking in Nepal.

Indeed, present-day language technology as we meet it in the form of spelling and grammar checking software relies on the existence of a standardized orthography. Ultimately, standardization means that some of the diversity in the language is eliminated. This issue is obviously not that straightforward. Bettina Zeisler in her article "Why Ladakhi must not be written – Being part of the Great Tradition: Another kind of global thinking" presents the illuminating case of a local mainly spoken language which faces competition not only from the officially dominant language, but also from within its own group. Ladakhi is spoken in the north of the Indian state of Jammu and Kashmir. It is not only under strong pressure from the official state language (Kashmiri), but also from the elitist attitudes of Ladakhi Buddhist scholars, who advocate literacy and literature only in Classical Tibetan – which many feel ought to be used for all writing, but which in practice only a few individuals master – and who work against promoting literacy and literature in Ladakhi. According to Zeisler, the classical orthography and grammar which represent some ninth century varieties – about as close to Ladakhi as Latin is to modern Spanish – are not suitable for writing Ladakhi, but at the same time, there are strong protests against using Ladakhi for literacy and literature by those who want to maintain the high status of Classical Tibetan.

Even though many linguists feel that it is self-evident that language standardization is all for the good, there are also dissenting opinions. Bernard (1996) feels that the same kind of market mechanisms that (over a period of several centuries) resulted in the regularization of the orthographies of languages like English should also be allowed to work for new literacies, whereas Östman (2001) questions the impartiality and universal validity of the principles commonly used to argue for language standardization: "If the Hualapai feel that there is nothing wrong with writing the Hualapai word for 'water' in a number of different ways, then that feeling and decision should be respected." (Östman 2001: 52; see also Foley 2003).

Turning now to the content of ICT, we find that it has two facets which are particularly pertinent in the context of lesser-known – and in particular endangered – languages. Firstly, there is the general circumstance that content here, as in media in general, is predominantly in "mainstream" languages, conveying the values, norms and attitudes of their cultures. About two thirds of the content of the web is in English, although less than half the online population are native speakers of English. Thus, ICT – together with the older mass media – is not culturally neutral as to its content, but instead provides immersion in majority languages and cultures to an unprecedented extent, and also at times provides inappropriate models for the use of the same technology for lesser-known languages (Cazden 2003). Secondly, although especially the internet is a democratic medium in the sense that lesser-known language communities may cut out the middleman and use this medium to spread information about themselves, to exchange information and to organize themselves in their own terms, this also comes with concomitantly greater risks of misappropriation: On the web, anyone can claim to represent a particular community (Warschauer 1998; McHenry 2002), and there is no reason to believe that this will happen less frequently in the new digital world than in the old non-digital one (cf. Howes 1996).

Keeping these potential stumbling blocks in mind, lesser-known language communities and researchers have endeavored to put modern technology to creative and culturally appropriate uses for their languages (Bredin 1996; Nettle and Romaine 2000, chapter 8; Buszard-Welcher 2001; McHenry 2002; Cazden 2003).

There have been top-down (i.e., by government agencies) as well as bottom-up efforts (i.e., the by speech communities themselves) in promoting lesser-known languages. ICT can play an important role in maintaining and promoting linguistic diversity, for instance, in documenting lesser-known languages and cultures and also in making information available to both speakers of these languages and outsiders. The web makes it easier to spread awareness about lesser-known languages and their communities. It also provides more flexible and easier means of communication within and outside the community (thus increased opportunities for the active use of languages). Certain communities in the Americas gathered for the first time by internet to organize themselves. An all-Hawai'ian language computer environment (with on-screen menus, messages, etc. only in Hawai'ian) and the Leokī chatroom has allowed a geographically dispersed community of Hawai'ian medium school classes to keep in touch electronically using the language (Warschauer and Donaghy 1997; Warschauer 1998). Finally, the availability of this modern, cool technology in a language confers prestige

to that language, raising its status in the eyes of its users and others. In this vein, David Nathan and Éva Csató in their article in this volume, "Multimedia: A community-oriented Information and Communication Technology", emphasize the importance of turning field research results into products which immediately support communities speaking endangered languages in their efforts to maintain their linguistic and cultural heritage. They describe three different genres of ICT products for documentation of community language heritage and language learning designed by Nathan and Csató for and together with the endangered language communities and delivered to these communities.

Language documentation as described here has been made possible more than anything else by modern information and communication technology. This technology has brought about a digital revolution in the way that primary linguistic data can be recorded, stored, annotated, retrieved and correlated (including high-quality sound and video recordings; see Hinton 2001). Further, it provides the means to present older written material and analog recordings in more modern media as well, thereby making their information accessible in new ways. For instance, traditional paper dictionaries can be scanned and stored in lexical databases (Corris et al. 2002), enabling access in the reverse direction (target language to source language), or the production of a reverse direction word list on paper (Miyashita and Moll 1999).

We are, at present, witnessing some positive efforts in documenting lesser-known languages in South Asia, using information and communication technology. In their article "Corpus-building for South Asian languages", Hardie et al. describe their work in the EMILLE project on building a South Asian language corpus. The goal of the project – which was largely achieved – was to create a combination of corpora (monolingual written, monolingual spoken and multilingual parallel written, with English as the source language) of a number of South Asian languages representing the Indo-Aryan and Dravidian language families. The completed EMILLE corpora consist of about 92.5 million words of written corpora in 13 languages, 2.6 million words of spoken corpora in 5 languages, and 1.2 million words of parallel corpora in 6 languages, making a respectable total of 96.3 million words. The article illustrates some of the difficulties which tend to beset work on languages which deviate from the Western European "norm" in various respects:

– poor availability of electronic texts, both in amount and variety
– a plethora of text and character encodings
– different linguistic tradition with regard to normativity vs. "pure description" (especially relevant for the spoken language corpora)

– lack of language technology resources and tools for corpus analysis and annotation

Boyd Michailovsky in his article "Digitized resources for languages of Nepal" presents an overview of available IT resources for languages of Nepal. This includes tools for the coding and rendering of Nepalese languages and scripts, spoken and written corpora with special focus on annotated speech recordings and dictionaries and wordlists. (LACITO in France has initiated an archive which contains texts of lesser-known languages (including some languages of Nepal, e.g. Hayu and Limbu). This archive contains transcriptions with time-aligned sound recordings, linked to glosses and translations, all available on the web. For portability, the archive is designed using standard formats and is accessible via standard web browsers.

The Chintang and Puma Documentation Project, carried out by Universität Leipzig, Germany, together with Tribhuvan University, Nepal, aims to provide a rich linguistic and ethnographic documentation of two highly endangered but almost totally undocumented languages in eastern Nepal, Chintang and Puma. Documentation includes language practices in context, together with transcripts with rich linguistic and ethnographic annotations. The project also includes a detailed study of language acquisition (for Chintang) over a period of approximately two years, the purpose of which is to gain insights on the micro-process of language endangerement, the role of bilingualism and trilingualism in this process, and the social and psychological mechanisms that lead to language death.

A particular kind of ICT which ought to be particularly relevant in this connection is *language technology*. Lars Borin in his article "Supporting lesser-known languages: The promise of language technology" in this volume presents a short introduction to language technology. Recently, there has been a good deal of concern about the creation of language technology resources for other languages than English and a few others, and especially for lesser-known languages. Proposed methods for the automatic acquisition of linguistic knowledge by computer potentially allow for the rapid creation of such resources with minimal human work, which if realized would be very useful. However, current such methods – like language technology in general – have arguably been shaped by the typological and other traits of the most explored language, namely English, which is in many respects an atypical language from a linguistic point of view. There is a need to test and refine these methods on a number of structurally diverse languages, making South Asia a good testing ground, in order for us to get a better understanding of the generality or language-specificness of these methods. Such experiments could be coordi-

nated with general documentation efforts going on in South Asia, resulting in the embryo of language technology resources for some lesser-known South Asian languages, as well as general methods for turning language documentation into linguistic description in the most economical way. This point is also emphasized by Jens Allwood in his article "Language survival kits", where he reiterates some cogent arguments in favor of efforts to preserve linguistic diversity. He further points to some ways in which modern technology and especially language technology can be brought to bear on this problem, namely primarily by supplying the basic tools making up the "language survival kits" outlined in the article.

A potentially useful way of looking at this issue is proposed by Trond Trosterud in his article on "Grammatically based language technology for lesser-known languages" in this volume, where he points out that the development of (at least certain kinds of) language technology applications can be seen as equivalent to doing basic linguistic descriptive work. In this way, the results of this work will be both a detailed formal linguistic description of some aspect of the language – morphology and some syntax in Trosterud's examples – and the beginning of basic language technology tools for the language.

5. Towards a pooling of knowledge

One important aim of this volume is to make available in one place articles belonging to areas of research that so far do not interact to any significant extent, namely those dealing with traditional South Asian descriptive linguistics and sociolinguistics, with documentary linguistics, intellectual and cultural property and fieldwork ethics, and with language technology. Researchers working in the areas of documentary linguistics and language technology have slowly become aware of each other in the last few years, and of how work in the other area could be potentially useful in furthering their own aims (see Borin's and Trosterud's articles in this volume). Similarly, the insights of documentary linguistics are slowly making their way into traditional descriptive linguistics and sociolinguistics, largely because of documentation funding initiatives such as those described above. However, the potential for synergy among these areas of research is almost limitless. In juxtaposing this assortment of seemingly quite disparate articles here, we wish to provide the reader, not so much with a do-it-yourself recipe for applying modern technology to the problem of language shift in South Asia today, but rather with some basic knowledge about the problems involved and some directions from which solutions could be forthcoming, a toolbox rather than a blueprint, if you

like. Hopefully these articles will give you both a glimpse of the shape of things to come, and enough information so that you can contribute to the shaping of that future.

Notes

1. I would like to thank Colette Grinevald for her input. The preparation of this volume was partly funded by a Swedish Research Council/SIDA–Swedish Research Links project and a conference grant by the Swedish Research Council.
2. Various other terms (e.g. *language murder*, *language suicide* and *language extinction*) have been used in this context. Nettle and Romaine (2000) eschew the use of the term "language suicide" as in most cases there are external factors forcing speakers to shift as their only means of survival. There is no consensus as to what is meant by language death (when a language should be considered dead). A commonly held view is that if a language does not have any active speakers, the language is considered dead/extinct. McLendon (1980: 147–148) provides a strikingly apt simile for the process of language shift:

    … like a social gathering where some people leave early without affecting the interactions of the rest of the participants much or even being noticed. But after a certain time, more and more people leave … At some point the few remaining participants realize that the majority of the participants at this event are gone and it must be defined as over even though some participants are left. Just as suddenly the few surviving speakers of a language discover they no longer have sufficient occasions which permit the use of the language because so few other individuals speak it and for a variety of reasons, such as lack of contact because of distance, or lack of compatibility or downright dislike, they rarely talk with the few individuals who are still able to speak. They do not turn mute, however. Rather they turn to the contacting language in an ever-expanding number of speech situations, and the 'dying' language ceases to be spoken not from lack of speakers but from lack of use.

3. But at the same time, it is important to keep in mind that attitudes of a speech community are not completely determined by these external factors; there are numerous observations showing that even under comparable external conditions, two speech communities may react diametrally differently (Dorian 1998).
4. Hualapai (Hwalbáy, Hwalbá:y, Walapai; see Östman 2000: 48–49) is an indigenous North American language, a Yuman language spoken in Arizona.
5. Bantawa is a Tibeto-Burman language spoken in Nepal. Nepali is the official state language of Nepal.
6. Officially, national and local authorities usually support the rights of so-called minority groups (including the use of one's own language), but in practice, such official views – sometimes even taking the form of laws or other regulatory documents – are not often implemented, partly because of limited resources and partly because of lack of genuine interest, or simply because at heart decision makers subscribe to an assimilationist ideology (Skutnabb-Kangas 1990; Dorian 1998;

Kawagley 2003), what some researchers (e.g. Spolsky 2004) have termed "ideological monolingualism".

7. This book lists endangered languages according to region. Some information is available at: <http://www.helsinki.fi/~tasalmin/nasia_report.html#Smansi>.

8. The term *tribal* is primarily used in the Indian context to refer to those languages which are listed as "tribal languages" in the Constitution of India (Article 342). The use of the term tribal in this sense is purely an administrative term – devoid of any linguistic motivation or basis. A community has been labelled as tribal in the Constitution of India because of a number of factors, factors such as historical, socio-economic and cultural (and language may be included as a subdomain of culture), but no linguistic motivation has been provided for treating or not treating a language as a tribal language.

9. Kinnauri is a Tibeto-Burman language spoken in the eastern part of the Indian state of Himachal Pradesh, and also in the neighboring region in China.

10. They are Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, Tamil, Telugu, Urdu. Today the scheduled languages number 22. See Udaya Narayana Singh's article in this volume for further details.

11. *Ethnologue* figures are cited from the web version <http://www.ethnologue.com> which reflects the 14th edition (published in 2000) of the printed *Ethnologue* at the time of writing of this introduction.

12. There are, at present, five different scripts to write Santhali. In Bihar, it is written in Devanagari, in West Bengal it is written in Bengali, in Orissa it is written in Oriya, Christians write in Roman and it is also written in Olchiki, a native Santhali script.

13. In the words of Östen Dahl (p.c.), when it comes to arguing for language preservation, most linguists seem to turn Whorfian. And indeed it seems that linguistic relativity *à la* Sapir and Whorf – "Facts are unlike to speakers whose language background provides for unlike formulation of them" (Whorf 1956: 235) – must be invoked in order for the kinds of arguments just cited here to hold water.

14. Truth be told, this is still often the case; it may even be generally considered detrimental to an academic career in linguistics to indulge in too much primary data collection, i.e., linguistic documentation (Grenoble and Whaley 2002; Grinevald 2001).

References

Abbi, Anvita (ed.)
    1997   *Languages of Tribal and Indigenous Peoples of India: The Ethnic Space.* Delhi: Motilal Banarsidass.
Annamalai, E.
    2003   The opportunity and challenge of language documentation in India. In *Language Documentation and Description*, vol. 1, Peter K. Austin (ed.), 159–167. London: Hans Rausing Endangered Languages Project.
Annamalai, E., and V. Gnanasundaram
    2001   Andamanese: Biological challenge for language reversal. In *Can Threat-*

ened Languages be Saved?, Joshua A. Fishman (ed.), 309–322. Cleve-don: Multilingual Matters.

Austin, Peter K.
  2003   Introduction. In *Language Documentation and Description*, vol. 1, Peter K. Austin (ed.), 6–14. London: Hans Rausing Endangered Languages Project.

Bernard, H. Russell
  1996   Language preservation and publishing. In *Indigenous Literacies in the Americas: Language Planning from the Bottom up*, Nancy H. Hornberger (ed.), 139–156. Berlin: Mouton de Gruyter. (References here are to the electronic version: <http://nersp.nerdc.ufl.edu/~ufruss/commodit.html>).

Bhattacharya, S. S.
  2002   Languages in India: Their status and function. In *Linguistic Landscaping in India with Particular Reference to the New States*, N. H. Itagi, and Shailendra Kumar Singh (eds.), 54–97. Mysore: CIIL.

Bird, Steven, and Gary Simons
  2003   Seven dimensions of portability for language documentation and description. *Language* 79 (3): 557–582.

Bradley, David, and Maya Bradley
  2002a   Conclusion: Resources for language maintenance. In *Language Endangerment and Language Maintenance*, David Bradley, and Maya Bradley (eds.), 348–353. London: RoutledgeCurzon.

Bradley, David, and Maya Bradley (eds.)
  2002b   *Language Endangerment and Language Maintenance*. London: RoutledgeCurzon.

Bredin, Marian
  1996   Transforming images: Communication technologies and cultural identity in Nishnawbe-Aski. In *Cross-Cultural Consumption: Global Markets, Local Realities*, David Howes (ed.), 161–177, London/New York: Routledge.

Brenzinger, Matthias (ed.)
  1992   *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*. Berlin: Mouton de Gruyter.

Burnaby, Barbara, and Jon Reyhner (eds.)
  2002   *Indigenous Languages across the Community*. Flagstaff: Northern Arizona University. Online edition:
           <http://jan.ucc.nau.edu/~jar/books.html>.

Buszard-Welcher, L.
  2001   Can the web help save my language? In *The Green Book of Language Revitalization in Practice*, Leanne Hinton, and Ken Hale (eds.), 331–345. San Diego: Academic Press.

Cantoni, Gina (ed.)
  1996   *Stabilizing Indigenous Languages*. Flagstaff: Northern Arizona University. Online edition: <http://jan.ucc.nau.edu/~jar/books.html>.

Cazden, Courtney B.
  2003   Sustaining indigenous languages in cyberspace. In *Nurturing Native Languages*, Jon Reyhner, Octavia V. Trujillo, Roberto Luis Carrasco, and Louise Lockaid (eds.), 53–57. Flagstaff: Northern Arizona University.

Corris, Miriam, Christopher Manning, Susan Poetsch, and Jane Simpson
  2002   Dictionaries and endangered languages. In *Language Endangerment and Language Maintenance*, David Bradley, and Maya Bradley (eds.), 329–347. London: RoutledgeCurzon.
Craig [= Grinevald], Colette
  1992   A constitutional response to language endangerment: The case of Nicaragua. *Language* 68 (1): 17–23.
Crystal, David
  2000   *Language Death*. Cambridge: Cambridge University Press.
Dauenhauer, Nora Marks, and Richard Dauenhauer
  1998   Technical, emotional, and ideological issues in reversing language shift: Examples from Southeast Alaska. In *Endangered Languages: Language Loss and Community Response*, Lenore A, Grenoble, and Lindsay J. Whaley (eds.), 57–98. Cambridge: Cambridge University Press.
Dorian, Nancy C.
  1981   *Language Death: The Life Cycle of a Scottish Gaelic Dialect*. Philadelphia: University of Pennsylvania Press.
  1998   Western language ideologies and small-language prospects. In *Endangered Languages: Language Loss and Community Response*, Lenore A. Grenoble, and Lindsay J. Whaley (eds.), 3–21. Cambridge: Cambridge University Press.
Dorian, Nancy C. (ed.)
  1989   *Investigating Obsolescence: Studies in Language Contraction and Death*. Cambridge: Cambridge University Press.
Dressler, Wolfgang, and Ruth Wodak-Leodolter (eds.)
  1977   *Language Death* (= IJSL 12). The Hague: Mouton.
England, Nora C.
  1992   Doing Mayan linguistics in Guatemala. *Language* 68 (1): 29–35.
  1998   Mayan efforts toward language preservation. In *Endangered Languages: Language Loss and Community Response*, Lenore A. Grenoble, and Lindsay J. Whaley (eds.), 99–116. Cambridge: Cambridge University Press.
Fishman, Joshua A.
  1996   What do you lose when you lose your language?. In *Stabilizing Indigenous languages*, Gina Cantoni (ed.), 80–91. Flagstaff: Northern Arizona University.
Foley, William A.
  2003   Genre, register and language documentation in literate and preliterate communities. In *Language Documentation and Description*, vol. 1, Peter K. Austin (ed.), 85–98. London: Hans Rausing Endangered Languages Project.
Freilich, Morris
  1991   Smart rules and proper rules: A journey through deviance. In *Deviance: Anthropological Perspectives*, Morris Freilich, Douglas Raybeck, and Joel Savishinsky (eds.), 27–50. New York: Bergin & Garvey.
Grenoble, Lenore A., and Lindsay J. Whaley
  2002   What does Yaghan have to do with digital technology? *Linguistic Discovery* 1 (2). Online journal:
            <http://linguistic-discovery.dartmouth.edu/>.

Grenoble, Lenore A., and Lindsay J. Whaley (eds.)
    1998   *Endangered Languages: Language Loss and Community Response*. Cambridge: Cambridge University Press.

Grinevald, Colette
    1997   Language contact and language degeneration**. *Handbook of Sociolinguistics*, Florian Coulmas (ed.), 257–270. Oxford: Blackwell.
    1998   Language endangerment in South America: A programmatic approach**. In *Endangered Languages: Language Loss and Community Response*, Lenore A. Grenoble, and Lindsey J. Whaley (eds.), 124–160. Cambridge: Cambridge University Press.
    2001   Encounters at the brink: Linguistic fieldwork among speakers of endangered languages. In *Lectures on Endangered Languages*, vol. 2, O. Sakiyama (ed.), 285–313. Kyoto, Japon, ELPR Publication series C002.
    2005a   Why Rama and not Rama Cay Creole? In *Language Documentation and Description*, vol. 3, Peter K. Austin (ed), 196–224. London: Hans Rausing Endangered Languages Project.
    2005b   Globalization and language endangerment: poison and antidote. The HRELP Annual Public Lecture (Hans Rausing Endangered Languages Project at SOAS), London, 11 February 2005.

Hale, Ken
    1992a   On endangered languages and the safeguarding of diversity. *Language* 68 (1): 1–3.
    1992b   Language endangerment and the human value of linguistic diversity. *Language* 68 (1): 35–42.
    1998   On endangered languages and the importance of linguistic diversity. In *Endangered Languages: Language Loss and Community Response*, Lenore A. Grenoble, and Lindsay J. Whaley (eds.), 192–216. Cambridge: Cambridge University Press.

Hinton, Leanne
    2001   Audio-video documentation. In *The Green Book of Language Revitalization in Practice*, Leanne A. Hinton, and Ken Hale (eds.), 316–329. San Diego: Academic Press.

Hinton, Leanne, and Ken Hale (eds.)
    2001   *The Green Book of Language Revitalization in Practice*. San Diego: Academic Press.

Howes, David
    1996   Cultural appropriation and resistance in the American Southwest: Decommodifying 'Indianness'. In *Cross-Cultural Consumption: Global Markets, Local Realities*, David Howes (ed.), 138–160, London/New York: Routledge.

Jocks, Christofer
    1998   Living words and cartoon translations: Longhouse 'texts' and the limitations of English. In *Endangered Languages: Language Loss and Community Response*, Lenore A. Grenoble, and Lindsay J. Whaley (eds.), 217–233. Cambridge: Cambridge University Press.

Kawagley, Angayuqaq Oscar
    2003   Nurturing native languages. In *Nurturing native languages*, Jon Reyhner, Octavia V. Trujillo, Roberto Luis Carrasco, and Louise Lockard (eds.), vii–x. Flagstaff: Northern Arizona University.

Krauss, Michael
    1992   The world's languages in crisis. *Language* 68 (1): 4–10.
    1996   Status of native American language endangerment. In *Stabilizing Indige-
               nous Languages*, Gina Cantoni (ed.), 16–21. Flagstaff: Northern Arizona
               University.
McCarty, Teresa L., and Lucille J. Watahomigie
    1999   Reclaiming indigenous languages. *Preservation on the Reservation (and
               Beyond)* [= *Common Ground*, Fall 1999]. Online edition:
               <http://www.cr.nps.gov/aad/cg/fa_1999/Language.htm>.
McHenry, Tracey
    2002   Words as big as the screen: Native American languages and the internet.
               Language Learning & Technology 6 (2) [Nicholas Ostler, and Jon Reyh-
               ner (eds.), *Special Issue on Technology and Indigenous Languages*]:
               102–115. Online journal: <http://llt.msu.edu/>.
McLendon, Sally
    1980   How languages die: A social history of unstable bilingualism among the
               Eastern Pomo. In *American Indian and Indoeuropean Studies: Papers in
               Honor of Madison S. Beeler*, Kathryn Klar, Margaret Langdon, and
               Shirley Silver (eds.), 137–150. The Hague: Mouton.
Mithun, Marianne
    1989   The incipient obsolescence of polysynthesis: Cayuga in Ontario and
               Oklahoma. In *Investigating Obsolescence: Studies in Language Contrac-
               tion and Death*, Nancy C. Dorian (ed.), 243–257. Cambridge: Cambridge
               University Press.
    1998   The significance of diversity in language endangerment and preservation.
               In *Endangered Languages: Language Loss and Community Response*,
               Lenore A. Grenoble, and Lindsay J. Whaley (eds.), 163–191. Cambridge:
               Cambridge University Press.
Miyashita, Mizuki, and Laura A. Moll
    1999   Enhancing language material availability using computers. In *Revitaliz-
               ing Indigenous Languages*, Jon Reyhner, Joseph Martin, Louise Lockard,
               and W. Sakiestewa Gilbert (eds.), 113–116. Flagstaff: Northern Arizona
               University.
Mohan, Shailendra
    2002   Linguistic landscape and social identity: A case of Jharkhand. In *Linguis-
               tic Landscaping in India with Particular Reference to the New States*,
               N. H. Itagi, and Shailendra Kumar Singh (eds.), 230–240. Mysore: CIIL
               and Mahatama Gandhi International Hindi University.
Nettle, Daniel, and Suzanne Romaine
    2000   *Vanishing Voices: The Extinction of the World's Languages*. Oxford:
               Oxford University Press.
Ostler, Nicholas, and Blair Rudes (eds.)
    2000   *Endangered Languages and Literacy*. Bath, UK: The Foundation for
               Endangered languages.
Östman, Jan-Ola
    2000   Ethics and appropriation – with special reference to Hwalbáy. In *Issues of
               Minority Peoples*, Frances Karttunen, and Jan-Ola Östman (eds.), 37–60.
               Publications No. 31, Department of General Linguistics, University of
               Helsinki.

Pandit, P. B.
   1972   *India as a Sociolinguistic Area*. Poona: University of Poona.
Payne, Doris
   1999   Review of Grenoble and Whaley 1998. *Journal of Linguistics* 35 (3): 618–624.
Reyhner, Jon (ed.)
   1997   *Teaching Indigenous Languages*. Flagstaff: Northern Arizona University.
Reyhner, Jon, Gina Cantoni, Robert N. St. Clair, and Evangeline Parsons Yazzie (eds.)
   2000   *Revitalizing Indigenous Languages*. Flagstaff: Northern Arizona University. Online edition: <http://jan.ucc.nau.edu/~jar/books.html>.
Reyhner, Jon, Joseph Martin, Louise Lockard, and W. Sakiestewa Gilbert (eds.)
   1999   *Learn in Beauty: Indigenous Education for a New Century*. Flagstaff: Northern Arizona University. Online edition: <http://jan.ucc.nau.edu/~jar/books.html>.
Reyhner, Jon, Octavia V. Trujillo, Roberto Luis Carrasco, and Louise Lockard (eds.)
   2003   *Nurturing Native Languages*. Flagstaff: Northern Arizona University. Online edition: <http://jan.ucc.nau.edu/~jar/books.html>.
Romaine, Suzanne
   1989   Pidgins, creoles, immigrant, and dying languages. In *Investigating Obsolescence: Studies in Language Contraction and Death*, Nancy C. Dorian (ed.), 369–383. Cambridge: Cambridge University Press.
Sasse, Hans-Jurgen
   1992   Theory of language death. In *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*, Matthias Brenzinger (ed.), 7–30. Berlin: Mouton de Gryuter.
Saxena, Anju (ed.).
   2004   *Himalayan Languages. Past and Present*. Berlin: Mouton de Gruyter.
Schmidt, Annette
   1985   *Young People's Dyirbal: An Example of Language Death from Australia*. Cambridge: Cambridge University Press.
Skutnabb-Kangas, Tove
   1990   Legitimating or delegitimating new forms of racism – the role of researchers. *Journal of Multilingual and Multicultural Development* 11 (1–2): 77–100.
   2000   *Linguistic Genocide in Education – or Worldwide Diversity and Human Rights?* Mahwah, NJ & London, UK: Lawrence Erlbaum Associates.
Southworth, Franklin C.
   1978   On the need for qualitative data to supplement language statistics: Some proposals based on the Indian census. *Indian Linguistics* 39: 136–154.
Spolsky, Bernard
   2004   *Language Policy*. Cambridge: Cambridge University Press.
Swadesh, Morris
   1948   Sociologic notes on obsolescent languages. *International Journal of American Linguistics* 14 (4): 226–235.
Trudgill, Peter
   1978   Creolization in reverse: Reduction and simplification in the Albanian dialects of Greece. *Transactions of the Philological Society 1976–7*, 32–50. Oxford: Basil Blackwell.

Warschauer, Mark
    1998   Technology and indigenous language revitalization: Analyzing the expe-
                rience of Hawai'i. *Canadian Modern Language Review* 55(1). (Refer-
                ences here are to the electronic version:
                <http://www.utpjournals.com/jour.ihtml?lp=
                product/cmlr/551/551-Warschauer.html> or
                <http://www.gse.uci.edu/markw/revitalization.html>).
Warschauer, Mark, and Keola Donaghy
    1997   Leokï: A powerful voice of Hawaiian language revitalization. *Computer
                Assisted Language Learning* 10 (4): 349–361.
Weinreich, Uriel
    1953   *Languages in Contact*. [Publications of the Linguistic Circle of New
                York, No. 1]. Page references to reprinted edition 1963. The Hague:
                Mouton.
Whorf, Benjamin Lee
    1956   Languages and logic. In *Language, Thought, and Reality: Selected Writ-
                ings of Benjamin Lee Whorf*, John B. Carroll (ed.), 233–245. Cambridge,
                Massachusetts: MIT Press.
Wilson, William H.
    1999   Return of a language: Hawaiian makes a remarkable comeback. *Preser-
                vation on the Reservation (and Beyond)* [= *Common Ground*, Fall 1999].
                Online edition: <http://www.cr.nps.gov/aad/cg/fa_1999/Language.htm>.
Winter, Werner
    1993   Some conditions for the survival of small languages. In *Language Con-
                flict and Language Planning*, Ernst Håkon Jahr (ed.), 299–314. Berlin:
                Mouton de Gruyter.
Woodbury, Anthony
    1998   Documenting rhetorical, aesthetic, and expressive loss in language shift.
                In *Endangered Languages: Language Loss and Community Response*,
                Lenore A. Grenoble, and Lindsay J. Whaley (eds), 234–258. Cambridge:
                Cambridge University Press.
    2003   Defining documentary linguistics. In *Language Documentation and
                Description*, vol. 1, Peter K. Austin (ed.), 35–51. London: Hans Rausing
                Endangered Languages Project.

Language situation and
language policies in South Asia

Status of lesser-known languages in India

Udaya Narayana Singh

1. Introduction

India accounts for 2.4% of the world's land surface with a total land-area of 2 973 190 square kilometres,[1] but it is obviously a densely populated area with 16% of the world's population living here (Heitzman and Worden 1996). Consequently, it has always been a home for a large number of languages. For instance, Census 1961 reports a total of 1652 "mother tongues", out of which 184 had more than 10 000 speakers (R. A. Singh 1969). The figures have changed in later census reports.[2] The encyclopaedic *People of India* series of the Anthropological Survey of India (K. S. Singh 1992), identified 75 "major languages" out of a total of 325 languages used in Indian households. *Ethnologue* (Gordon 2005), too, reports India as home for 398 languages, including 387 living and 11 extinct languages. Since as early as in the 1990s, India was reported to have at least 32 languages with a large population base of one million plus speakers. In fact, all seven countries of South Asia put together are considered as the third most linguistically populous area (Nettle 1999), after Papua New Guinea in Asia and the African region of Ivory Coast to Tanzania; South Asia is comparable only with Mexico in the new world (Grimes 1993).[3]

It is estimated that there are about 700–1000 languages spoken in the South Asian region, belonging to at least four major language families – Indo-European (most of which belong to one sub-branch, Indo-Aryan), Tibeto-Burman, Dravidian, and Austro-Asiatic. Multilingualism is not a new phenomenon in the Indian context. Even Sir George Grierson's (1903–1923) twelve-volume *Linguistic Survey of India* – material for which was collected in the last decade of the 19th century, had identified 179 languages and 544 dialects. One of the early Census reports also showed 188 languages and 49 dialects (Census 1921). But, despite this, recent social changes such as technological advances, urbanization and globalization are rapidly changing the linguistic tapestry of India – upsetting, in some ways, the linguistic equilibrium.

Since Independence the Indian government has made pronouncements in favour of linguistic diversity and promotion of less privileged groups (including languages) by means of introducing language policies and laws, but partly because of social factors, such as large population growth, low literacy level,

disparity between rich and poor, between urban and rural areas, effects of these government policies have not been as visible as one would have liked to.

The aim of this paper is to present an overview of the linguistic situation in India today, beginning with some background information on the linguistic demography of India. The focus in the next section will be on government efforts to promote and maintain linguistic diversity and document as well as support less privileged groups (including languages), including a discussion on the constitutional provisions for smaller language communities. This will be followed in section three by a discussion of some factors complicating the implementation of these language policies. Section four will focus on some noticeable trends visible today relating to lesser-known languages. These observations are based on a comparative study of census reports of the last several decades.

Although the focus here is on the linguistic scene in India, some pointers concerning South Asia will be made because the socio-linguistic scene in India is similar to other South Asian countries on several fronts.

2. Languages in India

Languages spoken in the South Asian region belong to at least four major language families: Indo-European (most of which, 74.24%, belong to its sub-branch Indo-Aryan), Dravidian (with 23.86% speakers), Austro-Asiatic (1.16%), and Sino-Tibetan (0.62%) as pointed out by Baldridge (1996, 2002; see also Gordon 2005).

The biggest chunk of languages and mother tongues belong to the Indo-Aryan sub-family of Indo-European languages. The immediate predecessor of Indo-Aryan is Indo-Iranian, the oldest specimens of which are available in the Zend-Avesta. Among the modern Indo-Aryan languages, Hindi and Bangla are the most well-known languages. Western Hindi is a Midland Indo-Aryan language, spoken in the Gangetic plain and in the region immediately to its north and south. Around it, on three sides, are Panjabi, Gujarati, Rajasthani. Eastern Hindi is spoken in Oudh and to its south. In the outer layer, we get languages such as Kashmiri, Lahnda, Sindhi, Gujarati, Marathi, in the northern and the western region, and Oriya, Maithili, Bengali and Assamese in the east.

The word *Dravidian* was first used by Robert A. Caldwell (1856), who introduced the Sanskrit word *Dravida* to designate the speech community. Among Dravidian languages, besides the four internationally known languages (Tamil, Telugu, Kannada and Malayalam), there are 26 languages by the current count, of which 25 are spoken in India and one (Brahui) is spoken